Developmental Science DOI: 10.1111/j.1467-7687.2007.00621.x

PAPER

Decision-making in healthy children, adolescents and adults explained by the use of increasingly complex proportional reasoning rules

Hilde M. Huizenga,[1] Eveline A. Crone[2] and Brenda J. Jansen[1]

1. *Department of Psychology, University of Amsterdam, The Netherlands*
2. *Department of Psychology, Leiden University, The Netherlands*

Abstract

In the standard Iowa Gambling Task (IGT), participants have to choose repeatedly from four options. Each option is characterized by a constant gain, and by the frequency and amount of a probabilistic loss. Crone and van der Molen (2004) reported that school-aged children and even adolescents show marked deficits in IGT performance. In this study, we have re-analyzed the data with a multivariate normal mixture analysis to show that these developmental changes can be explained by a shift from unidimensional to multidimensional proportional reasoning (Siegler, 1981; Jansen & van der Maas, 2002). More specifically, the results show a gradual shift with increasing age from (a) guessing with a slight tendency to consider frequency of loss to (b) focusing on frequency of loss, to (c) considering both frequency and amount of probabilistic loss. In the latter case, participants only considered options with low-frequency loss and then chose the option with the lowest amount of loss. Performance improved in a reversed task, in which punishment was placed up front and gain was delivered unexpectedly. In this reversed task, young children are guessing with already a slight tendency to consider both the frequency and amount of gain; this strategy becomes more pronounced with age. We argue that these findings have important implications for the interpretation of IGT performance, as well as for methods to analyze this performance.

Introduction

The Iowa Gambling Task (IGT) is an important paradigm designed to mimic real-life decision-making (Bechara, Damasio, Damasio & Anderson, 1994; see also Dunn, Dalgleish & Lawrence, 2006). In this task, participants are repeatedly asked to pick one card from four decks of cards. Two decks, A and B, result in consistent high gains but also in unpredictable high losses; the outcome is therefore unfavorable in the long run. The two remaining decks, C and D, yield low constant gains and low unpredictable losses and are favorable in the long run (see Table 1).

Bechara *et al.* (1994) demonstrated that healthy adults opt for the advantageous decks C and D, whereas patients with ventromedial prefrontal lesions choose the disadvantageous decks A and B. This pattern of results was termed ‘myopia for the future’, that is, a focus on immediate outcomes and no consideration of future consequences. In order to rule out the possibility that results were due to a general insensitivity to punishment, patients and matched controls were also tested in a reversed version of the task (Bechara, Tranel & Damasio, 2000). In this reversed task, in which the decks are characterized by their constant loss and unpredictable gains, normal controls again chose advantageous decks and ventromedial patients again chose disadvantageous decks. This finding corroborated the earlier interpretation of myopia for the future.

Recent developmental studies have shown that there are also age-related changes in performance on age-appropriate versions of standard and reversed gambling tasks (e.g. Kerr & Zelazo, 2004; Hooper, Luciana, Conklin & Yarger, 2004; Garon & Moore, 2004; Hongwanishkul, Happaney, Lee & Zelazo, 2005). With increasing age, participants sample more from advantageous than from disadvantageous decks. These results were interpreted to suggest that children also focus on immediate outcomes and therefore may also have ‘myopia for the future’ (Crone & Van der Molen, 2004). This conclusion is consistent

Address for correspondence: Hilde M. Huizenga, Department of Developmental Psychology, University of Amsterdam, Roeterstraat 15, 1018 WB Amsterdam, The Netherlands; e-mail: h.m.huizenga@uva.nl

Table 1 *Properties of A, B, C and D in the standard and reversed IGT (derived from Crone & Van der Molen, 2004). Gain: constant gain; % Loss: percentage of trials in which loss is delivered; Med loss: median of losses; Net: net result over 10 trials; Net Loss: net loss over 10 trials*

	Standard task					Reversed task				
	Gain	% Loss	Med. loss	Net	Net loss	Loss	% Gain	Med. gain	Net	Net gain
A	4	50%	−10	−10	−50	−4	50%	10	10	50
B	4	10%	−50	−10	−50	−4	10%	50	10	50
C	2	50%	−2	10	−10	−2	50%	2	−10	10
D	2	10%	−10	10	−10	−2	10%	10	−10	10

with the hypothesis that the prefrontal cortex matures relatively late (e.g. Casey, Tottenham, Liston & Durston, 2005).

The IGT is a complex task (see Table 1). Decks A and B are characterized by high constant gains, where deck A has a high loss in 50% of the trials and B a very high loss in 10% of the trials. Decks C and D are characterized by low constant gains. Deck C has a low loss in 50% of the trials and D a high loss in 10% of the trials. The participant needs to infer these properties of the decks, that is, the amount of constant gain, and the frequency and amount of probabilistic loss. Moreover, the participant has to combine these three properties in order to determine the net result. Bechara *et al.* (1994) hypothesized that this task is too complex and that participants therefore rely on a more intuitive decision process, based on somatic markers that signal net favorable outcomes (the somatic marker hypothesis; Damasio, 1994).

This interpretation was recently called into question by Dunn and colleagues (2006), who showed, by a formal analysis of IGT properties, that the standard IGT can be performed adequately if the dimension of constant reward is neglected. That is, only the amount and frequency of loss have to be multiplied to derive the correct answer (see Table 1, compare the Net column and the Net Loss column). In that case, the IGT can be framed as a proportional reasoning task.

Proportional reasoning is a classic theme in the study of cognitive development. This type of reasoning is studied with, for example, balance scale (Siegler, 1981) and probability tasks (e.g. Falk & Wilkening, 1998). These tasks share the fundamental characteristic that an item is characterized by two dimensions, one dominant and one subordinate. A correct response requires multiplication or division of these dimensions and comparison of the results among items. For example, in a balance scale task, each side of the scale is characterized by the number of weights and the distance of these weights from the fulcrum. The participant should multiply the number of weights with their distance and then compare the results from each side of the balance scale.

Children progress through a series of suboptimal stages before they use the correct multiplication/proportional rule (Siegler, 1981). Early in development, children are merely guessing, a strategy coined ‘Rule 0’. Subsequently, children consider the dominant dimension in their answer (‘Rule 1’). In a balance scale task, these children only focus on the number of weights on each side of the fulcrum. In the next stage, children focus on the dominant dimension, but if the dominant dimensions of two stimuli are equal, then they consider the subordinate dimension as well (‘Rule 2’). In the balance scale task, Rule 2 involves considering the distances at which the weights are placed, but only if the number of weights is equal. Subsequently, some children adopt idiosyncratic ‘muddle-through’ strategies, such as adding the dominant and subordinate dimension or guessing when the dominant and subordinate dimensions hint in different directions (‘Rule 3’). Finally, the correct strategy is adopted in which the dominant and subordinate dimensions are multiplied (‘Rule 4’).

The primary aim of the present study is to investigate whether the developmental results of Crone and Van der Molen (2004) can be explained by the notion that children use increasingly complex proportional reasoning rules. We hypothesize that *frequency* is the dominant dimension in the IGT. Prior studies have indicated that there is a low frequency preference in the choice pattern of children and adults in the standard Iowa Gambling Task (Overman, Frassrand, Ansel, Trawalter, Bies & Redmond, 2004; Crone & Van der Molen, 2004; Hooper *et al.*, 2004; Yechiam, Stout, Busemeyer, Rock & Finn, 2005). In addition, in the reversed version of the IGT, participants prefer choices associated with high-frequency reward (Crone & Van der Molen, 2004). Moreover, dominance of frequency over amount is a common theme in the decision-making literature (e.g. Tversky, Sattah & Slovic, 1988; Slovic, Finucane, Peters & MacGregor, 2004). We propose that the subordinate dimension is most likely the *amount* of infrequent punishment in the standard task and *amount* of infrequent reward in the reversed task. We expect that *constant* reward in the standard task

Table 2 *Predicted response patterns in the final 60 trials. The focus in the present paper is not on how participants infer the properties of response options A, B, C and D, but only on how participants combine these inferred properties. Therefore it is important to analyse the final 60 trials in which the choice behavior is not affected by an updating of the properties of response options anymore*

Rule	Description	Prediction standard	Prediction reversed
0	Guessing	A = B = C = D = 60/4	A = B = C = D = 60/4
1	Frequency	(B&D) > 60/4	(A&C) > 60/4
2	If frequency equal, then amount	D > 60/4	A > 60/4
4	Frequency*amount	(C&D) > 60/4	(A&B) > 60/4
	Myopia for the future	(A&B) > 60/4	(C&D) > 60/4
	Healthy adults, somatic markers	(C&D) > 60/4	(A&B) > 60/4

and *constant* punishment in the reversed task are neglected. This hypothesis is based on two observations. First, Dunn *et al.* (2006) showed that the task can be performed adequately if the dimension of constant reward/punishment is neglected (see Table 1). Second, constant reward/punishment is present in all trials and thus is likely to habituate, i.e. will attract less attention. Frequency as the dominant dimension, amount of infrequent punishment/reward as the subordinate dimension, and neglect of constant reward/punishment yield specific predicted response patterns that are summarized in Table 2.

In the developmental literature, strategic differences are often reported between subjects of the same age (e.g. Jansen & van der Maas, 2002). Similarly, Crone *et al.* (2004) and Hooper *et al.* (2004) reported that some age groups, in particular adolescents, showed considerable variation in IGT performance. The second aim of this study was therefore to test the hypothesis that this variation can be explained by the use of different proportional reasoning rules within age groups.

Method

We performed a re-analysis of the data reported by Crone and Van der Molen (2004). Here we present a summary of task and sample characteristics pertinent to the present paper, followed by a detailed description of our new analysis.

Participants

Four age groups were investigated: 61 young children (33 boys, 28 girls) between 6 and 9 years, 60 older children (27 boys, 33 girls) between 10 and 12 years, 59 adolescents (29 boys, 30 girls) between 13 and 15 years and 61 university students (12 males, 49 females) between 18 and 25 years. Children and adolescents were recruited by contacting schools, and students were recruited through flyers. All participants were reported to be healthy. All participants took a computerized version of the Raven Standard Progressive Matrices to provide an estimate of their inductive reasoning ability (Raven, Court & Raven, 1985). Raven scores were compared to Dutch age appropriate norms, resulting in age-corrected inductive reasoning scores varying between 0 and 100.

Task

All participants performed the 'hungry donkey task', a computerized version of the Iowa Gambling Task that can also be performed by children. The stimulus display consisted of four doors, A, B, C and D, and a donkey sitting in front of those doors. The participants were told to assist the hungry donkey to collect as many apples as possible by pressing one of the keys corresponding to the doors. Upon pressing one of the keys, the stimulus display was replaced by an outcome display showing the number of apples gained and the number of apples lost.[1] The participants performed both the standard and the reversed task, and each task consisted of 200 trials. Participants were randomly assigned to the standard first condition or the reversed first condition.

The gains and losses associated with doors A, B, C and D are given in Table 1. As an example, consider door A in the standard task. This door is associated with a gain of four apples on every trial. In addition, on 50% of the trials a loss is encountered, with a median loss of 10 apples. The net loss over 10 trials is 10*(−0.5*10) = −50. This yields in 10 trials a net result of 10*(+4−0.5*10) = −10. In the reversed task, door A is characterized by a *loss* of four apples on every trial, and a gain on 50% of the trials, with a median gain of 10 apples. It is important to note that these detailed door properties were not presented to the participants; instead, they had to infer these properties themselves during the experiment.

[1] Participants were randomly assigned to one of three feedback conditions in which the detail of the outcome was varied. Feedback condition did not affect the results, and therefore was omitted as a factor in the analysis.

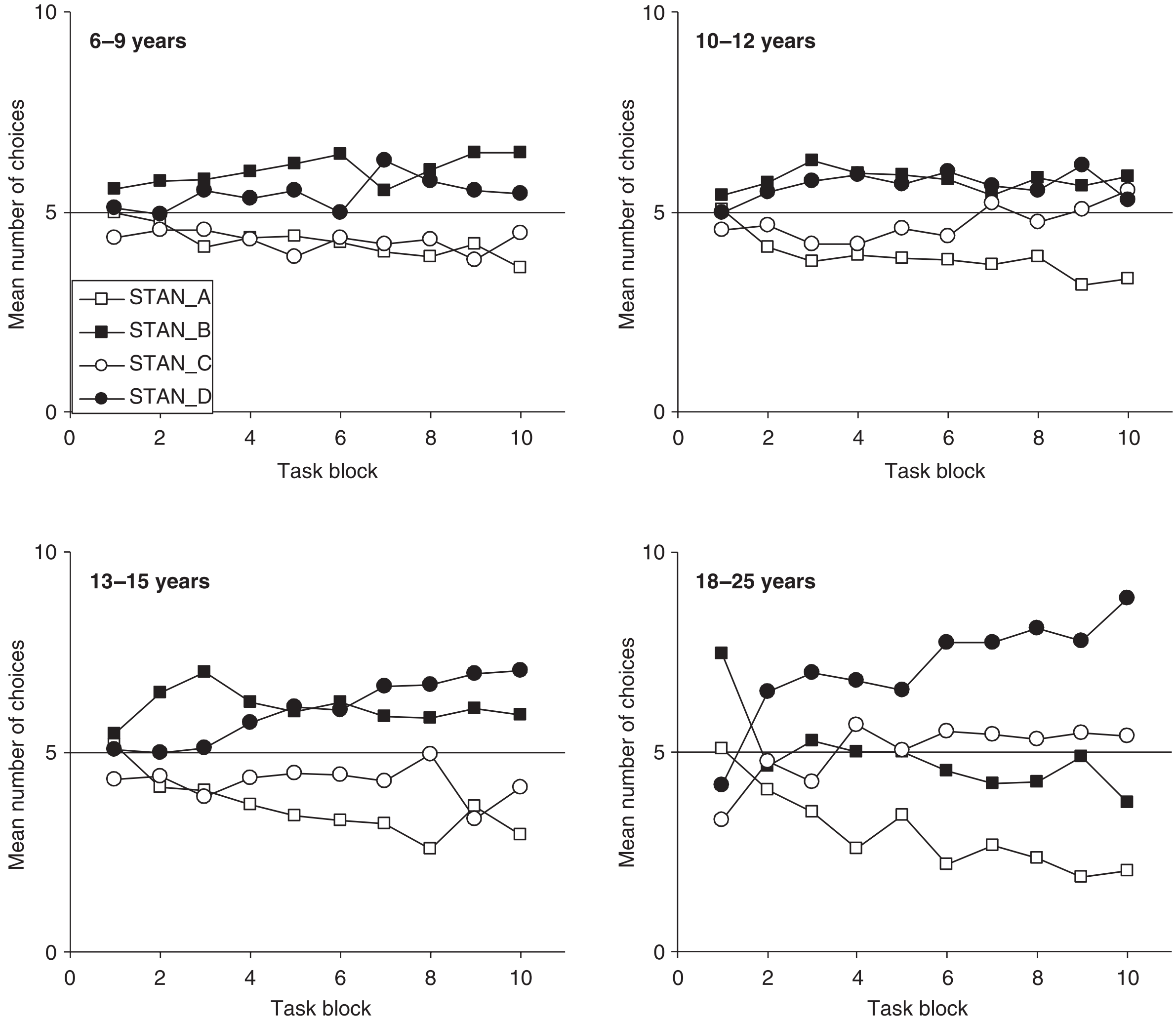

Figure 1 *Mean number of choices in each task block of the standard task. Each block consists of 20 trials.*

Analysis of door preferences by age groups

In this paper, we focus only on choice strategies after participants inferred door properties. Therefore we analyzed only the final part of the experiment in which choices stabilized. In Figures 1 and 2 it is shown that performance stabilized in the final 60 trials of the experiment; therefore we limit our analysis accordingly.

In order to test the hypotheses given in Table 2, the number of choices for each door must be analyzed separately. It does not suffice to adopt the standard approach in the IGT literature, i.e. to analyze the mean difference between net favorable doors and net unfavorable doors (A & B vs. C & D). A within-age group repeated-measures MANOVA with deviation contrasts was used to test whether the number of choices for a particular door deviated significantly from the number that would be expected if participants were only guessing, that is 60/4 = 15. In addition, between-age group ANOVAs were performed to test whether door preferences changed with age.

Multivariate normal mixture analysis

Our second aim was to obtain further insight into the variability within age groups. More specifically, we aimed to determine whether different rules were used within age groups. Suppose, as a hypothetical example, that only two rules were applied in the standard task: guessing (Rule 0) and focus on frequency (Rule 1). In that case the multivariate distribution of the data will not be normal but will instead consist of a mixture of two multivariate normal distributions: one centered at the multivariate mean (A = 15, B = 15, C = 15, D = 15) and one centered at the multivariate mean (0, 30, 0 and

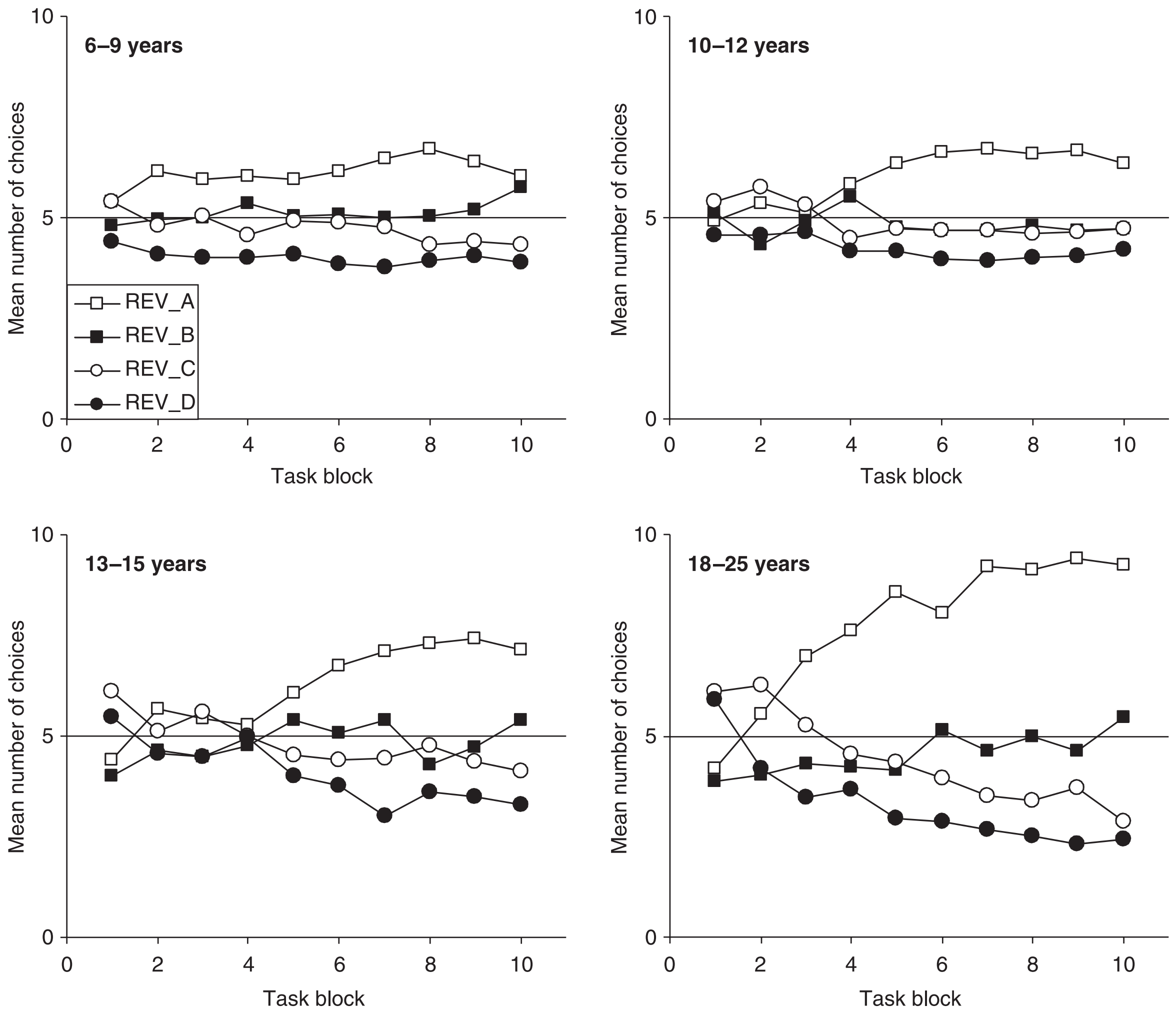

Figure 2 *Mean number of choices in each task block of the reversed task. Each block consists of 20 trials.*

30). If the observed data can indeed better be described by a mixture of two multivariate normal distributions instead of just one, then this indicates that the sample consists of two groups that apply different rules. This is exactly the purpose of a multivariate normal mixture analysis. Such an analysis is a statistically sound (Fraley & Raftery, 2002) and increasingly popular tool to determine whether participants can be divided into groups that respond differently (e.g. Fossati, Cittero, Garazioli, Borroni, Carretta, Maffei & Battaglia, 2005, for an application in psychiatry; Dolan, Jansen & van der Maas, 2004, for an application in developmental psychology). In the following, we will provide a detailed description of this procedure as applied to Iowa Gambling data.

In each trial, a participant has to choose between four options. Therefore, the distribution of choices in one trial follows a multinomial distribution. However, we do not analyze one trial, but the sum over 60 trials. Such a sum of multinomials follows a multivariate normal distribution (by the central limit theorem, e.g. Parzen, 1960, p. 430). It is therefore valid to perform a multivariate normal mixture analysis.

The mixture analysis is performed on all age groups simultaneously. The result of such a mixture analysis is a set of rule groups that describes the variation in strategies in the entire sample. This implies the assumption that strategies are fixed over age. This assumption is not problematic, however: if a different strategy is used in a specific age group, then this will turn up as an additional rule group in the mixture analysis.

The four variables, A, B, C and D, generate a redundant dataset: the total number of choices of the four doors is 60. Therefore the score on D can always be calculated if A, B and C are known. Such a redundant, or 'singular', dataset cannot be subjected to a mixture analysis. Therefore, we first performed a Principal Component Analysis on these four variables and obtained three principal components, which, per definition, completely describe the original data (e.g. Morrison, 1967). The scores on these three principal components, let's say P1, P2, and P3, were used as input variables for the mixture analysis (for a similar approach, see Fossati *et al.*, 2005).

In a mixture analysis on the variables P1, P2 and P3, solutions with one and two groups are first computed. The one-group solution is characterized by nine parameters: three parameters for the means of P1, P2 and P3, and six parameters for the unique elements in their covariance matrix. The two-group solution requires 2*9 +1 parameters: again nine parameters per group and one additional parameter to describe the weight of one distribution with respect to the other distribution. The two solutions are compared according to their Bayesian Information Criterion (BIC; Fraley & Raftery, 2003). If the one-group solution provides the best description, then the estimation procedure is stopped. If the two-group solution provides a better fit, then two- and three-group solutions are compared. This process is repeated until the best solution is found.

The parameters defining the multivariate distributions are estimated in an iterative way given a set of starting values on the means and covariances. There is a risk that this solution is not the optimal solution. In order to minimize this risk, the iterative procedure was restarted 5000 times with different starting values. These starting values were random, and were not chosen according to our hypotheses. Out of the 5000 solutions, the solution with the optimal BIC outcome was selected. Given this optimal solution, each participant was assigned to the most likely group. Based on this assignment, each group was characterized by its mean scores on the original variables A, B, C and D.

Results

Standard task

Analysis of door preferences by age groups

The average number of choices for doors A, B, C and D for the final 60 trials is presented in Figure 3. Our first analyses tested whether, within each age group, the average of each choice deviated significantly from 15. The results indicate that door B is preferred in the youngest age group ($F(1, 60) = 13.64$, $p < .001$), and that there is a slight but nonsignificant preference for D ($F(1, 60) = 2.66$, $p = .11$). Doors A and C were chosen less than 15 times ($p < .05$). Older children aged 10–12 opted for door B ($F(1, 59) = 6.30$, $p < .05$), and D ($F(1, 59) = 4.93$, $p < .05$). Door C was chosen about 15 times, and door A was chosen less than 15 times ($p < .05$). Adolescents prefer doors B ($F(1, 59) = 4.91$, $p < .05$) and D ($F(1, 59) = 15.38$, $p < .001$). Doors A and C were chosen less than 15 times ($p < .05$). Adults prefer door D ($F(1, 60) = 21.00$, $p < .001$); doors B and C were chosen about 15 times, and door A was chosen less than 15 times ($p < .001$).

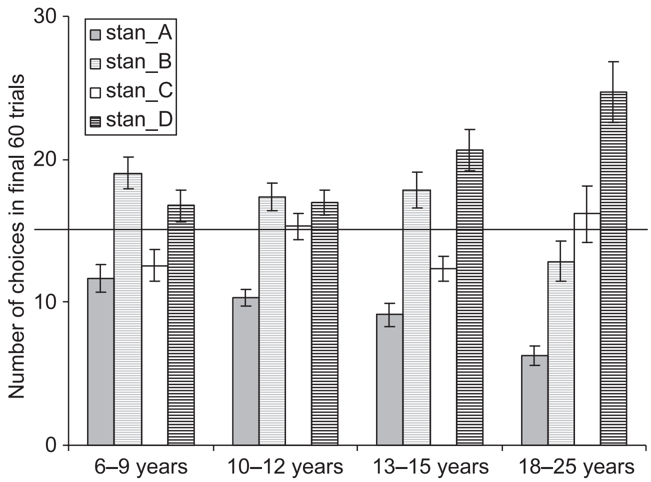

Figure 3 *Mean choices for doors A, B, C and D in the final 60 trials of the standard task. According to the 'myopia for the future' rule, children should prefer A and B, whereas adults should prefer C and D. According to Rule 1, participants should prefer B and D; according to Rule 2, participants should prefer D. Error bars represent standard errors of the mean.*

These results support a developmental trend of increasingly sophisticated rule use. Participants aged up to 15 years use Rule 1, demonstrating a focus on low frequency (doors B and D). Adult participants use Rule 2, first focusing on frequency and considering amount if frequency is equal (option D). The results of a between-age group analysis of B and D choices are in line with this suggestion. There was an effect of age on the number of choices for door B ($F(3, 238) = 5,13$, $p = .002$), due to a decreased preference of adults as compared to adolescents ($p = .004$). In addition, there was an effect on door D preference ($F(3, 238) = 6.65$, $p < .001$); adolescents tended to choose this option more often than older children ($p = .076$), and adults chose this option more often than adolescents ($p = .050$).

Multivariate normal mixture analysis

For a more detailed assessment of rule use, however, we determined whether these age groups consisted of groups that used different rules.

The mixture analysis yielded four latent groups, whose mean choices are given in Table 3. The first group was

Table 3 *Results of the mixture analysis in the standard task. For each latent group the average number of choices for A, B, C and D, and an interpretation of this average response pattern. Results are rounded and sums may therefore deviate from 60*

A	B	C	D	Interpretation
13	17	14	17	Rule 1a
8	20	15	17	Rule 1b
4	7	7	42	Rule 2
0	0	60	0	Rule 4 pref.

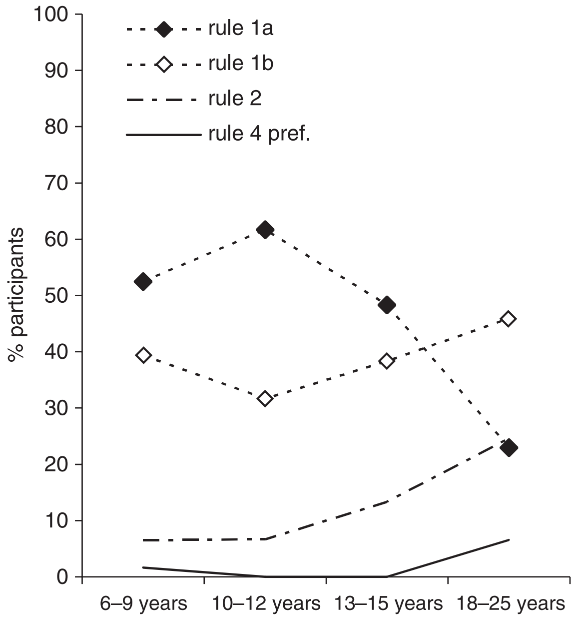

Figure 4 *Percentage of participants assigned to the four latent groups in the standard task.*

characterized by guessing with a slight preference for doors B and D and will thus be referred to as the Rule 1a group. The second group had a more pronounced focus on frequency (Rule 1b group). The third group preferred door D (Rule 2 group). The fourth, very small, group opted for door C. We interpret this as Rule 4 with a preference for one of the favorable options, namely C (Rule 4 pref. group). In Figure 4 it can be seen that until adolescence, children were classified in either the Rule 1a or Rule 1b groups. In adolescence, the Rule 1a group decreased in size, whereas the Rule 2 group increased in size. This developmental trend of a decrease in Rule 1a use and an increase in Rule 2 use continued into adulthood. Note that only a very small number of adults used Rule 4 pref. χ^2 tests of adjacent age groups indicated that younger and older children did not differ in the number of children that used a particular rule (χ^2 = ns). Similarly, older children and adolescents also did not differ from each other (χ^2 = ns). There were, however, differences between adolescents and adults ($\chi^2(3) = 11.85$, $p = .008$): the adult group contained a decreased number of Rule 1a users and an increased number of Rule 2 users.

The 'myopia for the future' rule predicts that children prefer the immediate high rewards offered by doors A and B and that adults prefer the net favorable outcomes of doors C and D. This expected pattern of preferences was not supported by the data: A was chosen at a level below chance (see Figure 3, Table 3) by all participants, including children. Moreover, a preference for B was always coupled with a preference for D (see Figure 3, Table 3), which is characterized by low immediate gain. The data do lend support to a proportional reasoning explanation, with a gradual shift from (a) guessing with a slight tendency to consider frequency of loss to (b) focusing on frequency of loss, to (c) considering both frequency and amount of probabilistic loss. The mixture analysis indicates that these rules are all present in the adolescent sample, which may give an explanation for the heterogeneity within this age group.

Reversed task

Analysis of door preferences by age groups

The average number of choices in the final 60 trials is presented in Figure 5. Door A is preferred by the youngest age group ($F(1, 60) = 12.93$, $p < .001$), B was chosen about 15 times and C and D were chosen less than 15 times ($p < .05$). Older children aged 10–12 also opted for

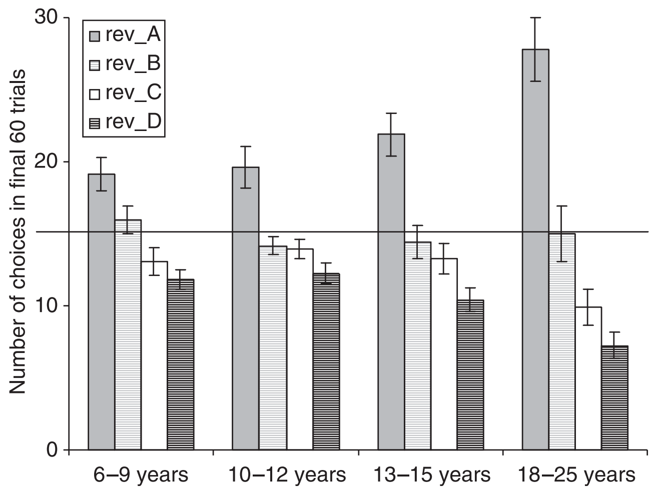

Figure 5 *Mean choices in the final 60 trials of the reversed task. According to the 'myopia for the future' rule, children should prefer choices C and D, whereas adults should prefer both A and B. According to Rule 1, participants should prefer A and C; according to Rule 2, participants should prefer A. Error bars represent standard errors of the mean.*

Table 4 *Results of the mixture analysis in the reversed task. For each latent group the average number of choices for A, B, C and D, and an interpretation of this average response pattern. Results are rounded and sums may therefore deviate from 60*

A	B	C	D	Interpretation
18	15	15	13	Rule 2a
34	16	6	4	Rule 2b
55	2	2	1	Rule 2c
13	19	15	12	Rule 4 pref.

A, ($F(1, 59) = 10.54$, $p < .01$), B and C did not deviate significantly from 15, and D was avoided ($p < .001$). Adolescents preferred A ($F(1, 59) = 21.22$, $p < .001$), B and C were chosen about 15 times, and D was avoided ($p < .001$). Adults preferred A ($F(1, 60) = 33.08$ $p < .001$), chose B about 15 times, and avoided C and D ($p < .001$).

These within-group results suggest that participants of all ages use Rule 2: focus on frequency and select the option with highest gains if frequency is equal. The between-group results on the high frequency reward options A and C indicate that this rule is especially pronounced in the mature sample. That is, there was an effect of age group on the number of choices for door A ($F(3, 237) = 6.040$, $p = .001$). Follow-up comparisons indicate that this is due to an increased preference of adults as compared to adolescents ($p = .011$). There was also an overall effect on preferences for door C ($F(3, 237) = 3.05$, $p = .029$). Adults showed a decreased preference as compared to adolescents ($p = .023$).

Multivariate normal mixture analysis

The mixture analysis revealed four rule groups (Table 4). Three of these latent groups, termed Groups 2a, 2b, and 2c, used Rule 2, with an increasing preference for option A. The fourth group had a slight preference for option B. This might be explained by a Rule 4 strategy with a preference for one of the two net favorable options, namely B. Therefore this final group was termed ‘Group 4 pref.’. Figure 6 indicates that participants in the two youngest age groups are mainly classified as Group 2a (guessers with a slight preference for A). In adolescence, an increasing number of participants was classified in Group 2b, 2c, or 4 pref., a developmental trend that continued into adulthood. Younger and older children did not differ in the number of participants that used a particular rule (χ^2 = ns). In contrast, older children and adolescents did differ in rule use ($\chi^2(3) = 9.35$, $p = .025$): the adolescent group contained a decreased number of Rule 2a users, and an increased number of participants that used the other rules. Adolescents and adults also

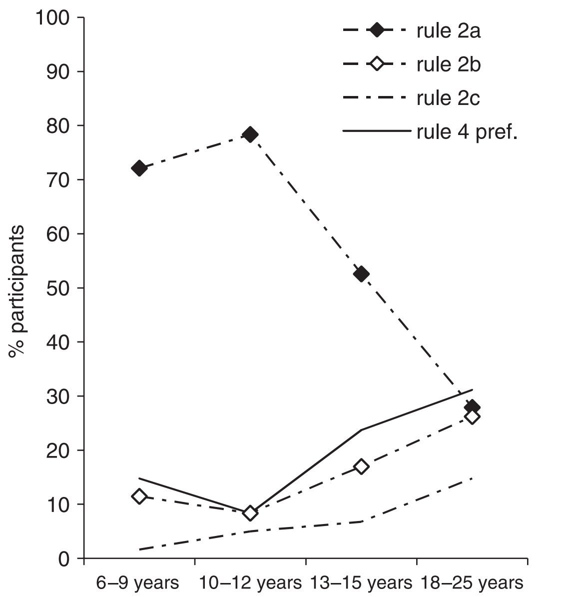

Figure 6 *Percentage of participants assigned to the four latent groups in the reversed task.*

differed in rule use ($\chi^2(3) = 8.12$, $p = .044$): the adult group contained a decreased number of Rule 2a users and an increased number of participants that used the other rules.

These results indicate that although Rule 2 is the dominant strategy at all ages, this strategy develops from a very slight preference for option A at younger ages to a very pronounced preference in maturity. The mixture results indicate that the variability in the adolescent sample might be explained by the presence of multiple decision rules within this age group.

Characteristics of Rule 2 users in the standard task

In the standard task, a with age increasing number of participants was classified as Rule 2 users. What characterizes these Rule 2 users? In the following exploratory analysis, we will show that in all age groups, except the adult group, these participants can be characterized by their significantly higher inductive reasoning skills.

In the young children group the mean standardized Raven scores for Rule 1a, Rule 1b and Rule 2 users were 56.72, 57.62, and 86.25, respectively; the difference between the Rule 1a and Rule 2 groups was significant ($t(14.11) = 5.13$, $p < .001$), as well as the difference between the Rule 1b and Rule 2 groups ($t(16.33) = 4.49$, $p < .001$). In the older children group, the mean standardized Raven scores for Rule 1a, Rule 1b and Rule 2 were 48.57, 51.05 and 78.75, respectively, with again significant differences between Rule 1a and Rule 2 ($t(16.89) = 4.95$, $p < .001$), and Rule 1b and Rule 2 ($t(18.86) = 3.77$, $p = .001$). In the adolescent group, the mean standardized Raven scores were 59.31,

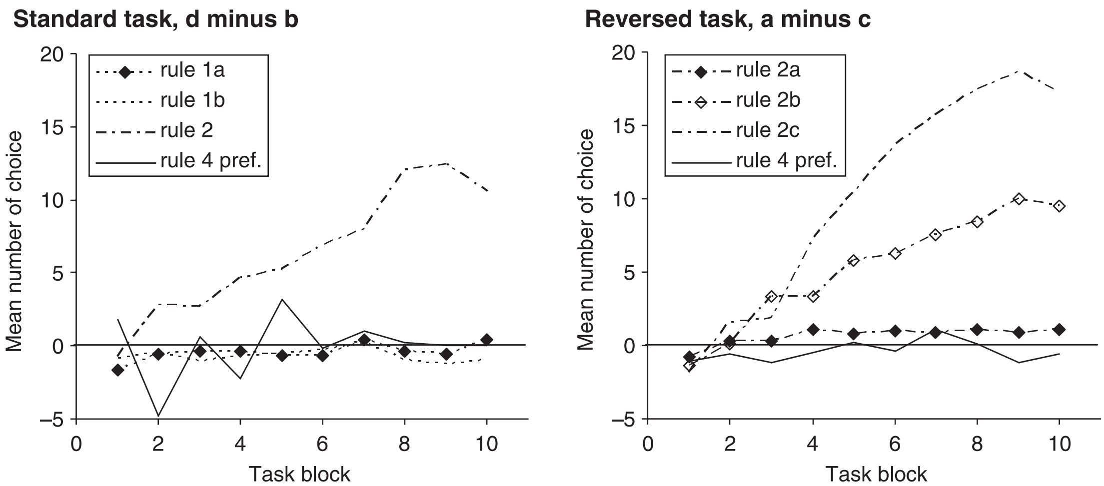

Figure 7 *Learning curves in standard and reversed tasks. In the left hand panel the learning curves of Rule 1a, Rule 1b, Rule 2 and Rule 4 pref. users in the standard task. Each curve represents the number of D minus B choices. Rule 1 users should not develop a preference of D over B, whereas Rule 2 users should develop such a preference. In the right hand panel the learning curves of Rule 2a, 2b, 2c and Rule 4 pref. users. The curves represent the number of A minus C choices.*

56.52 and 81.88, with significant differences between Rule 1a and Rule 2 ($t(32.50) = 3.70$, $p = .001$) and Rule 1b and Rule 2 ($t(28.99) = 3.73$, $p = .001$).

In not one of the age groups did we find sex differences on Rule 2 as opposed to Rule 1a use, or on Rule 2 as opposed to Rule 1b use. There is one exception: in the adolescent group, there was a sex effect on Rule 2 versus Rule 1b ($\chi^2(1) = 8.67$, $p = .003$): Rule 2 was used by seven males and one female, and Rule 1b was used by six males and 17 females.

In order to further investigate the Raven and sex effects in the adolescent sample we selected only those participants with a Raven score exceeding 70. We then performed a χ^2 test, which indicated a significant association between sex and rule use ($\chi^2(2) = 5.95$, $p = .05$). High scoring males used either Rule 1a or Rule 2, whereas high scoring females used Rule 1a or Rule 1b.

Learning curves

The primary aim of the present paper was to investigate the decision rules applied by participants after they inferred the properties of the four options. It is interesting to note, however, that the classification based on the mixture analysis of the final part of the experiment yielded quite distinct learning curves for the entire experiment. In Figure 7, left hand panel, it can be seen that participants classified as Rule 1a or Rule 1b in the standard task favored both low-frequency loss options equally during the entire experiment. The Rule 2 group started to develop a preference of D over B after about 40 trials, a preference which was gradually increasing during the experiment. In Figure 7, right hand panel, it can be seen that the Rule 2a group (reversed task) developed a slight preference of A over C after about 80 trials. The Rule 2b and 2c groups showed an earlier preference of A over C, after about 60 trials. This preference gradually increased during the experiment, especially in the Rule 2c group.

Discussion

In this study, proportional reasoning theory (Siegler, 1981; Falk & Wilkening, 1998) and the methodology of a multivariate normal mixture analysis (e.g. Dolan *et al.*, 2004) were used to examine the development of strategy use in the Iowa Gambling Task. In IGT research, the typical analysis method is to compare the sum of two disadvantageous choices with the sum of two advantageous choices, that is (A+B)-(C+D). Prior analyses using this standard IGT methodology have demonstrated that children have a diminished preference for advantageous doors, which was interpreted as myopia for the future (Crone & Van der Molen, 2004; Hooper *et al.*, 2004). The present analysis of the same data indicates that performance on the IGT is better described by the development of increasingly sophisticated proportional reasoning rules. The results show that in the standard task the dominant strategy in all age groups is to focus on frequency (Rule 1), and that this rule becomes more pronounced with age. However, there is also a subgroup, increasing with age, that first focuses on frequency and then considers the amount of probabilistic loss (Rule 2). It is interesting to note that, up to 15 years, the Rule 2 group is characterized by significantly higher inductive reasoning skills as assessed by Raven's Standard Progressive matrices. Note that Rule 2 behavior necessitates the skill to integrate two dimensions of a stimulus: frequency of loss and amount of loss. A similar integration of stimulus dimensions is required in the Raven's test. Further research is required to assess whether brain regions that are activated by this dimensional integration in the Raven's test (e.g. Christoff, Prabhakaran, Dorfman, Zhao, Kroger, Holyoak & Gabrieli, 2001) are also activated by the Rule 2 group in the IGT.

The present task, however, does not offer the possibility to investigate why some participants are able to use Rule 2, whereas others stick to Rule 1. Possible explanations are that Rule 1 users have a limited processing capacity and therefore are only able to focus on frequency and do not have the ability to consider the other characteristics of choice options. A second explanation is that Rule 1 users are able to consider another characteristic of the choice options, but are puzzled about which dimension to choose. In contrast, Rule 2 users do know which dimension to choose, they consider the amount of loss, because they either might (a) quickly habituate to constantly provided reward, (b) are very focused on errors or (c) are able to estimate and compare the net result of constant gain and infrequent loss of options B and D. Additional research is required to investigate these mechanisms in more detail.

Prior research has demonstrated that children aged 6–9 perform worse in the standard compared to the reversed task (Crone & Van der Molen, 2004). The present analysis indicates that this result can be explained by children's use of Rule 1 in the standard task and Rule 2 in the reversed task (although these rules were not very pronounced). One tentative explanation for this context dependency of rule use is that the subordinate dimension, amount of probabilistic loss/gain, is more salient in the reversed task (probabilistic gain) than in the standard task (probabilistic loss). That is, children are more affected by winning items (in the reversed task) than by losing items (in the standard task). If the subordinate dimension is more salient, then they may consider this dimension earlier. This is in line with observations of Jansen and van der Maas (2001) who showed, using a balance scale task, that children tend to switch from Rule 1 to Rule 2 if the salience of the subordinate dimension is increased, and tend to switch from Rule 2 to Rule 1 if this salience is decreased. Further research is required to determine whether this mechanism can also explain the discrepancy between standard and reversed IGT tasks.

The results indicate that it is informative to analyze each of the response options separately (see Yechiam *et al.*, 2005). Differences between specific decks A, B, C and D are generally not reported, despite their large informative value about the specifics of performance differences. For example, Shurman, Horan and Ntuechterlein (2005) showed that schizophrenic patients in the standard task favored infrequent punishment (B and D). This can, in the present terminology, be classified as Rule 1 – a focus on frequency. These patients may not have the capacity to consider the subordinate dimension of the amount of probabilistic loss. Second, Toplak, Jain and Tannock (2005) showed that in the standard task, adolescents with ADHD opted for the disadvantageous deck B. One tentative explanation for this finding might be that frequency is again the dominant dimension, but that constant reward, instead of probabilistic loss, serves as the subordinate dimension. Third, Yechiam and colleagues (2005, Figure 2) also report an advantage of low-frequency over high-frequency decks in the standard IGT paradigm, and, within low-frequency options, a slight preference for the option with low probabilistic loss. This is consistent with the present conclusion that adults performing the standard task use either a pronounced Rule 1 or a pronounced Rule 2. Finally, Overman (2004), Overman *et al.* (2004), and Overman, Graham, Redmond, Eubank, Boettcher, Samplawski and Walsh (2006) showed that in the standard task, females, as opposed to males, tended to focus on low-frequency loss while they ignored the amount of low-frequency loss. In our terminology these females would be classified as Rule 1 users. In the present study we found only marginal sex differences. In the adolescent sample, males with high inductive reasoning scores tended to use either Rule 1a or Rule 2, whereas females with high scores tended to use Rule 1a or 1b: These high-functioning females were thus not considering the amount of loss, whereas a subgroup of high-functioning males did consider the amount of loss. This sex difference might be associated with the fact that the right lateral orbitofrontal cortex, a region which is specifically sensitive to the *amount* of loss (O'Doherty, Kringelbach, Rolls, Hornak & Andrews, 2001; De Martino, Kumaran, Seymour & Dolan, 2006), is activated more by males than by females while they are performing the IGT (Bolla *et al.*, 2004). Our results on sex differences are, however, based on very small samples and therefore further research is required to investigate this effect.

Although Bechara *et al.* (1994) argued that healthy adults in the standard task favor options C and D to an equal extent, this pattern was not observed in the present results. An equal favorability for C and D would be categorized as the correct proportional reasoning Rule 4. The present results, however, show that adults do not equally favor C and D but rather opt for D, which was interpreted as Rule 2. Importantly, in the traditional IGT paradigm, Rule 4 is not required to perform optimally – Rule 2 suffices to generate optimal outcomes.

What, then, might be the origin of the discrepancy with the Bechara *et al.* findings? One explanation is that in the present study, all doors can be chosen infinitely, whereas Bechara *et al.* used decks of cards that are finite. Therefore, participants might have switched to proportional reasoning (more C choices) if deck D runs low. A second explanation is that in the task used by Bechara *et al.* (1994), the participant is allowed to make 100 choices, whereas in the current study each task contained 200 trials, allowing for the analysis of stabilized performance. Despite these differences, the traditional IGT paradigm does not offer the possibility to test whether participants who favor D are applying Rule 2 or Rule 4 with a preference for one of two possibilities (i.e. D). Rule 4 use can best be investigated in a modified paradigm, in which optimal outcomes can only be attained if Rule 4 is used.

The mixture results suggest that Rule 4, the proportional reasoning rule, is used by only a few adult participants. This implies that mathematical models of IGT performance (for an overview, see e.g. Yechiam & Busemeyer, 2005) should not only model decisions guided by expected values (Rule 4), but should also allow for the possibility that these decisions are guided by Rules 1 or 2. Moreover, the near absence of Rule 4 implies that somatic markers signaling net favorable options are unlikely to be found. This is not to say that somatic markers cannot be found for options that are favorable according to Rule 2 or to Rule 1. The latter possibility is in fact supported by Crone and Van der Molen (submitted), who showed that in children and adolescents the skin conductance response (SCR) differentiates between options with frequent as opposed to infrequent probabilistic loss, rather than advantageous versus disadvantageous choices. This may imply that the SCR signals options are favorable on the dominant dimension, in this case frequency (see, however, Tomb, Hauser, Deldin & Carmazza, 2002).

A mixture analysis can be very informative, offering the possibility of investigating the origin of variation within groups. The present results indicate that the large variation in the adolescent sample is due to the fact that adolescents use different rules to solve the gambling task. Such a mixture approach may also be suited to investigate strategic differences in clinical samples that show heterogeneous IGT results. The mixture approach can then result in homogeneous subgroups which subsequently can be compared on secondary measures like their skin conductance response or fMRI activation profiles. Note, however, that a mixture analysis on IGT data is not without limitations. Mixture analysis is highly dependent on starting values, and the algorithm should thus be restarted repeatedly to avoid suboptimal solutions. Moreover, a mixture analysis requires large datasets in order to obtain stable solutions. In addition, a mixture analysis on IGT data 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. However, we do have confidence in the present solution since the solution: (a) was obtained from arbitrary starting values, (b) was based on a large sample, (c) yielded very interpretable rule groups, (d) yielded very interpretable rule group differences on inductive reasoning scores, and (e) yielded distinct learning curves associated with each rule group.

In sum, the present analysis shows that it is fruitful to apply the proportional reasoning framework and its associated methodology of mixture analysis to the Iowa Gambling Task. Conversely, it might also be beneficial to apply the concepts and methods in the IGT literature to proportional reasoning research. In particular, it is well known that children, although they can perform a proportional reasoning task adequately, are not able to verbalize their strategy (e.g. Falk & Wilkening, 1998). If somatic markers indeed serve to select options that score favorably on the dominant dimension (Rule 1) (see Crone & Van der Molen, 2004), then somatic markers might also signal these favorable options in other proportional reasoning tasks, like the balance scale task.

Acknowledgements

The authors thank Han van der Maas, Maurits van der Molen, Eric-Jan Wagenmakers, Bertjan Doosje, Santani Teng and three anonymous reviewers for their very helpful comments.

References

Bechara, A., Damasio, A.R., Damasio, H., & Anderson, S.W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, **50** (1–3), 7–15.

Bechara, A., Tranel, D., & Damasio, H. (2000). Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain*, **123** (11), 2189–2202.

Bolla, K.I., Eldreth, D.A., Matochik, J.A, & Cadet, J.I. (2004). Sex-related differences in a gambling task and its neurological correlates. *Cerebral Cortex*, **14**, 1226–1232.

Casey, B.J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: what have we learned about cognitive development? *Trends in Cognitive Sciences*, **9** (3), 104–110.

Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J.K., Holyoak, K.J., & Gabrieli, J.D.E. (2001). Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage*, **14** (5), 1136–1149.

Crone, E.A., & Van der Molen, M.W. (2004). Developmental changes in real life decision making: performance on a gambling task previously shown to depend on the ventromedial prefrontal cortex. *Developmental Neuropsychology*, **25** (3), 251–279.

Crone, E.A., & Van der Molen, M.W. (submitted). Development of decision-making in school-aged children and adolescents: evidence from heart rate and skin conductance analysis.

Damasio, H. (1994). *Descartes' error*. New York: Grosset/Putnam.

De Martino, B., Kumaran, D., Seymour, B., & Dolan, R.J. (2006). Frames, biases, and rational decision-making in the human brain. *Science*, **313**, 684–687.

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

Dunn, B.D., Dalgleish, T., & Lawrence, A.D. (2006). The somatic marker hypothesis: a critical evaluation. *Neuroscience and Biobehavioral Reviews*, **30** (2), 239–271.

Falk, R., & Wilkening, F. (1998). Children's construction of fair chances: adjusting probabilities. *Developmental Psychology*, **34** (6), 1340–1357.

Fossati, A., Citterio, A., Grazioli, F., Borroni, S., Carretta, I., Maffei, C., & Battaglia, M. (2005). Taxonic structure of schizotypal personality disorder: a multiple-instrument, multi-sample study based on mixture models. *Psychiatry Research*, **137** (1–2), 71–85.

Fraley, C., & Raftery, A.E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, **97** (458), 611–631.

Fraley, C., & Raftery, A.E. (2003). Enhanced model-based clustering, density estimation, and discriminant analysis software: MCLUST. *Journal of Classification*, **20** (2), 263–286.

Garon, N., & Moore, C. (2004). Complex decision-making in early childhood. *Brain and Cognition*, **55** (1), 158–170.

Hongwanishkul, D., Happaney, K.R., Lee, W.S.C., & Zelazo, P.D. (2005). Assessment of hot and cool executive function in children: age-related changes and individual differences. *Developmental Neuropsychology*, **29** (2), 617–644.

Hooper, C.J., Luciana, M., Conklin, H.M., & Yarger, R.S. (2004). Adolescents' performance on the Iowa gambling task: implications for the development of decision making and ventromedial prefrontal cortex. *Developmental Psychology*, **40** (6), 1148–1158.

Jansen, B.R., & van der Maas, H.L.J. (2001). Evidence of a phase transition from Rule I to Rule II on the balance scale task. *Developmental Review*, **21** (4), 450–494.

Jansen, B.R., & 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.

Kerr, A., & Zelazo, P.D. (2004). Development of 'hot' executive function: the children's gambling task. *Brain and Cognition*, **55** (1), 148–157.

Morrison, D.F. (1967). *Multivariate statistical methods*. New York: McGraw-Hill.

O'Doherty, J., Kringelbach, M.L., Rolls, E.T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience*, **4** (1), 95–102.

Overman, W.H. (2004). Sex differences in early childhood, adolescence, and adulthood on cognitive tasks that rely on orbital prefrontal cortex. *Brain and Cognition*, **55**, 134–147.

Overman, W.H., Frassrand, K., Ansel, S., Trawalter, S., Bies, B., & Redmond, A. (2004). Performance on the IOWA card task by adolescents and adults. *Neuropsychologia*, **42**, 1838–1851.

Overman, W.H., Graham, L., Redmond, A., Eubank, R., Boettcher, L., Samplawski, O., & Walsh, K. (2006). Contemplation of moral dilemmas eliminates sex differences on the Iowa gambling task (2006). *Behavioral Neuroscience*, **120** (4), 817–825.

Parzen, E. (1960). *Modern probability theory and its applications*. New York: Wiley.

Raven, J.C., Court, J.H., & Raven, J. (1985). *Raven progressive matrices*. London: J.C. Raven.

Shurman, B., Horan, W.P., & Ntuechterlein, K.H. (2005). Schizophrenia patients demonstrate a distinctive pattern of decision-making impairment on the Iowa gambling task. *Schizophrenia Research*, **72** (2–3), 215–224.

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

Slovic, P., Finucane, M.L., Peters, E., & MacGregor, D.G. (2004). Risk as analysis and risk as feelings: some thoughts about affect, reason, risk and rationality. *Risk Analysis*, **24** (2), 311–322.

Tomb, I., Hauser, M., Deldin, P., & Carmazza, A. (2002). Do somatic markers mediate decisions on the gambling task? *Nature Neuroscience*, **5** (11), 1103–1104.

Toplak, M.E., Jain, U., & Tannock, R. (2005). Executive and motivational processes in adolescents with Attention-Deficit-Hyperactivity Disorder. *Behaviour and Brain Functions*, **1** (1), 8.

Tversky, A., Sattah, S., & Slovic, P. (1988). Contingent weighting in judgement and choice. *Psychological Review*, **95** (3), 371–384.

Yechiam, E., & Busemeyer, J.R. (2005). Comparison of basic assumptions embedded in learning models for experience-based decision making. *Psychonomic Bulletin and Review*, **12** (3), 387–402.

Yechiam, E., Stout, J.C., Busemeyer, J.R., Rock, S.L., & Finn, P.R. (2005). Individual differences in the response to forgone payoffs: an examination of high functioning drug abusers. *Journal of Behavioral Decision Making*, **18** (2), 97–110.

Received: 9 June 2006
Accepted: 17 October 2006