Journal of Behavioral Decision Making
J. Behav. Dec. Making, **23**: 69–82 (2010)
Published online 16 September 2009 in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/bdm.672

Developmental Insights into Experience-based Decision Making

TIM RAKOW* and SALEEMA B. RAHIM
University of Essex, Colchester, UK

ABSTRACT

In three experiments involving children and adults ($N = 324$), option payoffs for sure versus risky choices were either described or experienced via observation of 20 outcomes. Choices revealed a description-experience gap for payoffs with rare events, implying greater impact of small probabilities ($\leq .2$) for described than for experienced choices. The size of this effect was independent of participant age. Therefore, the role of cognitive limitations in the description-experience distinction remains unclear, as the age groups would have differed in cognitive capacity. Age-related differences in 'sampling style' in decisions from experience were observed. Pre-choice data acquisition changed markedly with age: From frequent alternation between options towards separate systematic exploration of options with increasing age. A fourth experiment, that manipulated sampling style, failed to demonstrate its link to other age-related features of choice (e.g. risk preferences). Our studies illustrate the value of developmental research for testing theoretical claims and revealing novel phenomena in decision research. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS risky choice; decisions from experience; description versus experience; information acquisition; sampling; sequence-order effects

INTRODUCTION

In experience-based decision making (EBDM), decision makers learn about option payoffs by observing outcomes—analogous to observing payouts from different gaming machines. For instance, a decision maker may observe that Option A has recently made seven payments as follows: \$10, \$10, \$10, \$10, \$0, \$10, \$10—whereas, Option B has been observed to pay \$9 on each of its last eight payouts. This procedure contrasts with description-based risky choice—the dominant paradigm in decision research of recent decades. Here participants are told the *a priori* payoffs, such as: 'Option A is a lottery with a 90% chance of winning \$10 and a 10% chance of winning nothing; whereas, Option B pays \$9 for certain'. In such described decisions, participants generally overweight small probabilities relative to a unit weighting function (Kahneman & Tversky, 1979; Prelec, 1998; Tversky & Kahneman, 1992). For instance, in the example above, Option B

* Correspondence to: Dr Tim Rakow, Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK.
E-mail: timrakow@essex.ac.uk

(the ‘sure thing’) is usually preferred. In contrast, small probabilities appear to be given less weight when option payoffs are experienced than when they are described. For instance, Weber, Shafir, and Blais (2004) found that 72% of participants picked Option A in the above choice problem when they learned about the option payoffs via cards dealt from two decks with the appropriate proportion of each payout. The dominant choice in this example is consistent with underweighting small probabilities (it is *as if* participants pay little regard to the possibility of receiving nothing).

This description–experience (D-E) gap (in patterns of choice) has been much discussed (Barron & Erev, 2003; Hertwig, Barron, Weber, & Erev, 2004; Rakow, Demes, & Newell, 2008), including elsewhere in this *Special Issue* (Erev et al., 2009; Hau, Pleskac, & Hertwig, 2009). There is little disagreement that reliance on small samples in EBDM contributes to the D-E gap. At the heart of this consensus is the observation that the binomial distribution is skewed when event probabilities are small, and markedly so for small samples. For instance, in the above example: The number of participants who encounter \$0 less often than the expected 10% of the time should exceed the number who encounter it more than 10% of the time. Where there is disagreement, it is over the relative contribution of actual sampling—a *statistical* phenomenon—and mental sampling (i.e. drawing sub-samples of the actual sample encountered from memory)—a *cognitive* phenomenon (cf. Hertwig et al., 2004, with Fox & Hadar, 2006; or see Camilleri & Newell, 2009, or Rakow et al., 2008). If decision makers use samples of observations that are held in working memory to evaluate the option payoffs, then individual differences in memory capacity should predict reliable differences in patterns of choice in EBDM because the skewness of the sample depends partly upon its size.

A developmental approach[1] offers a promising strategy for exploring the role of memory in EBDM and the D-E gap because there are clear differences in memory capacity between children and adults (Hitch & Towse, 1995). Thus, if the relevant sample for an experience-based choice is one that is drawn from memory (Hertwig et al., 2004; Kareev, Lieberman, & Lev, 1997) children should make their decisions based on smaller samples (because their working memory capacity is smaller). Consequently, the D-E gap (the discrepancy between described and experienced choices) should be greater for children than for adults, as the binomial distribution is more highly skewed for small samples. Furthermore, if recent observations are more likely to form part of such a memory-based sample (Hertwig et al., 2004), then the influence of recent observations should be greater for children than for adults (as recent outcomes should constitute a higher proportion of their smaller sample). Rakow et al. (2008) failed to find such a relationship between working memory capacity and the degree of recency in EBDM. However, their participants were adults who could choose what size of sample to draw for each option prior to making a one-shot decision. Consequently, participants often tailored the size of their examined sample to ‘fit’ their memory capacity: The size of the sample was correlated with a measure of working memory capacity, and was often inside the presumed limit of short-term storage. In the experiments reported here, participants making experience-based choices always made ten observations per option. If memory limitations are implicated in the D-E gap, this should be sufficient for these to emerge.

EXPERIMENTS 1–3: DESCRIBED AND EXPERIENCED CHOICE WITH DIFFERENT AGE GROUPS

Method

Three experiments were conducted, in which participants made either described choices or experienced choices for a series of choice problems. Each problem presented participants with a pair of boxes containing

[1]Our introduction to this paper is purposefully brief. Extensive reviews of decision research with children and adolescents can be found in Boyer (2006) or Reyna and Farley (2006). Levin and Hart (2003) provide a brief overview of developmental research on risky choice. Boyer (2007) reports a pair of studies on experience-based choice in children with some similarities to those reported here, and reviews several developmental studies of EBDM.

DOI: 10.1002/bdm

cubes that could vary in point value. The participant's task was to choose one box from which he/she drew one cube at random. This presentation format was adopted so that the task was appropriately concrete for younger participants. Participants were instructed that their goal was to win as many points as possible.

Experienced choices followed the decisions-from-sampling paradigm (see Erev et al., 2009—this issue; Hertwig et al., 2004; Rakow et al., 2008; Weber et al., 2004). Participants in the experience condition learned about both options by making 10 observations of the payout for each option (i.e. random draws *with replacement*), in any order that they chose. They then made their one-shot choice between the two options. Table 1 shows the choice problems, which in each case involved choice between a sure thing and a risky option with two possible payout values (participants were not informed of this problem structure). In the described condition, the payoffs for these same problems were presented using a frequency format (i.e. no explicit mention of probabilities or percentages). For instance, the risky option in Problem 1 of Experiment 1 was described as follows: 'In this box, nine cubes are yellow and worth ten points, and one cube is green, worth zero points'. This format was chosen because younger participants might not yet have been taught to use percentages or probabilities.

Each experiment involved two separate age groups, and participants within each group were randomly assigned to one of the conditions (description or experience). Details specific to each experiment are as follows.

Experiment 1

The participants were 52 Years 5 and 6 children (26 male) from a mixed-ability inner-London primary school (mean age of 9.6 years, range 9–11 years), and 49 university students (17 male, mean age of 21.6 years, $SD = 2.8$ years). The problems were presented using pairs of opaque boxes, in a single fixed order (1–4–3–2). When described, the option payoffs were shown on the front of each box, always indicating 10 cubes per box. The experimenter read the task instructions and tested each participant individually. Participants learned the outcome of their one-shot choice before moving on to the next problem.

Experiment 2

The participants were 75 Years 5 and 6 children (44 male) from another inner-London primary school (aged 9–12 years), and an adolescent group consisting of 77 Year 12 students (23 male) aged 16–17 years. A computer-based version of the task was used with the boxes shown on screen, with responses made using a 'mouse'. The boxes were described as containing 20 cubes (to accommodate outcome probabilities such as .15). Problem order was randomized for each participant, and participants only learned the outcome of their one-shot choices on completion of the final problem. Task instructions were presented on screen, but were also read aloud to the children (to ensure comprehension). The children were tested individually, and the adolescents were tested in small groups.[2]

Experiment 3

The participants were a young adolescent group of 34 Year 8 pupils (16 male) aged 12–13 years, and an adolescent group of 37 Year 12 students (2 male) aged 16–17 years. The task and procedure were identical to those described above for the adolescent participants in Experiment 2, with the addition that upon completion

[2]Additionally, the adolescent participants in Experiment 2 completed the other condition (i.e. description or experience) after they had completed their assigned condition. These additional data are not reported here to ensure like-for-like comparison between age groups. As participants were unaware of this second set of problems until the first set had been completed, this has no bearing on the between-subjects analysis presented here.

DOI: 10.1002/bdm

Table 1. Examination of the description–experience (D-E) gap: Percentage of participants choosing the risky option by condition (description or experience)

	Option			% choosing risky option			
Problem	Risky V_1, p; V_2	Sure V_s	Option favoured if rare event overweighted	Description (D)	Experience (E)	Raw D-E gap‡	$\chi^2(1)$ for diff. D versus E†
Experiment 1 ($N = 101$, between-ss)				($n = 50$)	($n = 51$)		
1	10, .9; 0	9	Sure	66	86	+20	5.73*
2	10, .8; 0	8	Sure	84	80	−4	0.22
3	10, .6; 0	6	Sure	78	71	−7	0.73
4	6, .5; 0	3	—	64	67	—	0.08
Experiment 2 ($N = 152$, between-ss)				($n = 74$)	($n = 78$)		
1	20, .05; 0	1	Risky	48	31	+17	4.81*
2	10, .9; 0	9	Sure	43	60	+17	4.47*
3	20, .15; 0	3	Risky	51	40	+9	1.74
4	10, .8; 0	8	Sure	47	55	+8	0.95
5	8, .75; 0	6	Sure	57	55	−2	0.10
6	10, .3; 0	3	Risky	40	51	−11	1.66
Experiment 3 ($N = 71$, within-ss)				($n = 71$)	($n = 71$)		
1	10, .9; 0	9	Sure	48	59	+11	1.78
2	4, .9; −6	3	Sure	39	58	+19	4.12*
3	8, .9;−12	6	Sure	39	58	+19	5.12*
4	10, .8; 0	8	Sure	47	58	+11	1.60
5	12, .8; −8	8	Sure	41	62	+21	7.25**
6	12, .2; −8	−4	Risky	55	51	+4	0.27

V_1 is Value (points) for risky outcome, otherwise V_2. V_s is Value of sure outcome.
†McNemar χ^2 for paired observations used in Experiment 3.
‡Signed difference in choice percentages between conditions. Positive differences are consistent with the expectation that rare events are given greater weight for described choices than for experienced ones.
$^{*}p < .05$; $^{**}p < .01$.

DOI: 10.1002/bdm

of the described or experienced problems participants then completed the same problems in the other mode of presentation (experience or description, respectively). Participants were not informed that the two sets of problems were the same in both halves of the experiment, encountered them in a different random order and were not told how many problems to expect.

Results and discussion

Do we find a description-experience gap?

Table 1 shows the pattern of choice for each problem. Positive values in the penultimate column indicate where the difference in choice proportions is in the expected direction (i.e. consistent with lower weighting of small probabilities for experienced choices than for described ones). The gap is in the predicted direction for the majority of problems (11/15) and is significant in 40% of problems. This replication of the D-E gap is noteworthy as our problem descriptions used frequencies, not percentages as per most previous demonstrations of the gap. A key feature of Table 1 becomes evident upon emphasising that the problems for each experiment are listed according to the probability of the least likely outcome for the risky option (ascending chance). For rare event probabilities of .1, the gap (difference in percentages) is always in the predicted direction, is typically 15–20%, and is usually statistically significant. For rare event probabilities of .2, the gap is usually in the predicted direction, and is less frequently significant for individual problems with an average difference around 8%. When the least frequent event occurred with probability >.2 no reliable D-E gaps were apparent. Data from Erev et al. (2009) show a similar pattern: They also examined the D-E gap across multiple problems (a larger number of problems, but fewer participants per problem).

This observation is important for two reasons. First, whilst all demonstrations of, and all explanations for, the D-E gap assume that the phenomenon is restricted to choices involving small probabilities, it is important that this point is not lost. The discrepancy between described and experienced choice is not necessarily universal. With respect to patterns of choice, it is restricted to problems where at least one of the outcomes is relatively rare. Moreover, these data provide an answer to the question: *How rare is a rare event?* An outcome probability around .1 was small enough to deliver a significant D-E gap most of the time in a moderately large sample of participants, whereas the gap appeared only some of the time for event probabilities around .2. This concurs with the derivation of prospect theory's decision weight function from described decisions, where overweighting is inferred for probabilities up to .3, but is minimal in the range of .2–.3 (e.g. Prelec, 1998). It also aligns with the observation that (in experienced samples) the distribution of the number of occurrences of the least likely outcome exhibits greatest skew when this event is 'rare'. For instance, in Experiment 1 the skewness for the percentage of times that the worst outcome (0 points) was observed was +0.83 in Problem 1 (where its probability was .1). This was the only problem with a significant D-E gap in Experiment 1. In contrast, observed outcomes in Problems 2–4 showed relatively little skew: +0.33, +0.14 and –0.20, respectively. Thus for these problems, participants were about as likely to see the rare event too often as they were to encounter it too infrequently, relative to the expected frequency based on its *a priori* probability (.2–.5 in these problems). If the D-E gap only emerges when there is skew in the *observed* sample, it weakens the need for a 'cognitive' account of the D-E gap in decisions from sampling.

The second reason why this observation is important is that it emphasizes that the size of the D-E gap observed in a given experiment will depend greatly upon the particular set of choice problems examined. This is illustrated in Table 2, which reports the percentage of choices that are consistent with overweighting small probabilities. (This can be determined by considering that overweighting a rare attractive outcome should increase preference for a risky option, whereas overweighting the worst outcome should decrease preference for a risky option.) Examination of the 'all participant' data shows that the mean of these scores was always greater in the description condition than in the experience condition. This is consistent with relatively greater weight for small probabilities in described choices than for experienced ones, and is indicative of the predicted D-E gap. However, this effect is very small and not significant in Experiment 1 ($F < 1$),

DOI: 10.1002/bdm

Table 2. Examination of the description–experience (D-E) gap: Percentage of choices (standard deviation) consistent with overweighting small probabilities

Age group	Presentation mode		Combined	Ave. D-E gap† (%, D-E)
	Description	Experience		
(a) Experiment 1*				
Adult	32 (30)	36 (26)	34 (31)	−4
Child	22 (21)	13 (18)	17 (20)	+9
All participants	27 (26)	24 (28)		+3
(b) Experiment 2				
Adolescent	48 (23)	42 (24)	45 (23)	+6
Child	50 (22)	42 (23)	46 (23)	+8
All participants	49 (22)	42 (23)		+7
(c) Experiment 3				
Adolescent	56 (28)	46 (21)	51 (26)	+10
Young adolescent	57 (25)	40 (18)	49 (24)	+17
All participants	57 (27)	43 (20)		+14

*Expectations about over/under-weighting probabilities make no specific prediction for Problem 4 in Experiment 1. For consistency with Problems 1–3, choosing the sure thing in Problem 4 was coded as consistent with overweighting. Consequently, values for Experiment 1 also represent the percentage of non-risky choices.
†Positive differences reflect the expectation that rare events are given greater weight for described choices than for experienced ones.

small/medium but not quite significant in Experiment 2 ($p = .069$) and medium-sized and significant in Experiment 3 ($p < .001$).[3] This aligns with the fact that the proportion of problems with rare event probabilities $\leq .1$ increases across experiments (1/4, 1/3 to 1/2), as does the proportion of problems where this probability is $\leq .2$ (1/2, 2/3 and 1, respectively). Thus comparisons of the D-E gap in *different* experiments are of limited interest, and may even be misleading, *unless* care is taken to take into account the particular 'mix' of problems used in each study. (A corollary of this is that a model that is 'tuned' to a particular subset of problems may fit the data well in one experiment, but not in another.) Consequently, it is of particular interest when a factor moderates the size of the D-E gap for a *given* problem or set of problems—then we are brought closer to understanding the 'gap'. Therefore, we turn to an examination of whether age moderates the size of the D-E gap.

Does the size of the D-E gap vary with age?

The size of the D-E gap is shown separately for each participant group in Table 2. In each experiment, the 'gap' is larger for the younger group. However, the gap was never significantly larger for the younger age group than for the older one: The two-way (presentation mode by age group) interaction was non-significant in Experiment 1, $F(1,97) = 1.79$, $p = .191$, partial $\eta^2 = .02$, Experiment 2 ($F < 1$) and Experiment 3 ($F < 1$). There are a number of interpretations of this (approximately) null effect. One possibility is that the difference in working memory capacity between our two groups was simply too small to engender meaningful discrepancies that depend upon memory capacity. A related possibility is that the mental sampling of individual payouts is a poor model for decisions from sampling. For instance, belief updating models of

[3]Between-subjects analysis of Experiment 3 revealed that the D-E gap was larger the first time that participants encountered the problems (mean ϕ of 0.17 vs. 0.12). It is therefore possible that, in comparison to experiments employing a between-subjects design, the within-subjects design used here *slightly* reduces the size of the D-E gap. However, this order effect is small and non-significant—so would appear to make little difference to the data. Camilleri and Newell (2009) also demonstrated a D-E gap in within-subjects data.

DOI: 10.1002/bdm

sequential information acquisition assume that only two values need be considered at any one time: One value representing the current belief about an option, plus the most recent datum observed for that option (e.g. Busemeyer & Townsend, 1993; Hogarth & Einhorn, 1992; Sarin & Vahid, 2001). This places minimal demands upon working memory. Such models would therefore predict little relation between memory capacity and beliefs or choices.

Are sequence order effects apparent, and do they vary with age?

Sequence order effects were examined for experience-based choices because memory-based sampling accounts of EBDM predict recency effects. These should be more marked in individuals with lower working memory capacity (e.g. young children): Hence, order effects were analysed separately for each age group. Order effects were analysed using a strategy employed by Hertwig et al. (2004) and Rakow et al. (2008). The 10 observations for each option were split into two sets: The first five ('first half') and the last five observations ('second half'). The average value of the observations was calculated for each half sequence and for the entire sequence of 10 observations. We then determined whether the choice made was consistent with choosing the higher value option. This was done separately for each half sequence, and for the whole sequence. If expected value (EV) was equal, the sure thing was predicted (an arbitrary resolution of the 'tie'). The number of choices predicted by this procedure was calculated separately for each participant. These data are summarized as percentages in Table 3. Comparison of the percentage of choices predicted by each half sequence can be used to infer order effects. If the second half predicts better than the first half, then a recency effect is implied: Because recent observations predict choices better than earlier observations.

Table 3 shows an interesting pattern of effects. In each experiment, the younger group exhibits comparatively more recency—or, equivalently, the older group shows comparatively more primacy. Indeed, in Experiment 1, the adult group showed a significant primacy effect (first half predicts better than second half), $F(1,49) = 4.05$, $p = .0496$. Importantly, the two-way age group by sequence half interaction was significant in this experiment, $F(1,49) = 4.12$, $p = .048$, partial $\eta^2 = .08$—so the order effect for adults differed significantly from that for children. The older group in Experiment 3 also showed signs of a primacy effect. However, this was not significant, $t(36) = 1.64$, $p = .111$, as was also the case for the two-way interaction ($p = .312$). The order effects were not significant in Experiment 2. Therefore, whilst there is a coherent pattern of age-related differences in sequence order effects across the experiments, age was only a significant moderator when the difference between age groups was greatest.

These results add to those from previous studies that show the ephemeral nature of sequence order effects in EBDM. For instance, Hertwig et al. (2004) observed recency for decisions from sampling, whereas Hau,

Table 3. Examination of order effects: Percentage of choices predicted by the value of observations by sequence half

	Sequence examined			
Age group	First half	Second half	Entire sequence	Implied order effect*
(a) Experiment 1				
Adult	69	55	61	Primacy
Child	50	56	51	Recency
(b) Experiment 2				
Adolescent	64	64	67	None
Child	60	66	62	Recency
(c) Experiment 3				
Adolescent	64	56	64	Primacy
Young adolescent	58	57	60	None

*Order effect implied by comparison of the percentages shown in the columns labelled 'first half' and 'second half'.

DOI: 10.1002/bdm

Pleskac, Keifer, and Hertwig (2008) found no clear evidence of recency for the same set of problems. Rakow et al. (2008) observed recency when participants controlled the sampling process ('active sampling') but not when participants were passive observers of the same set of outcomes. The importance of Experiment 1 is that it shows that not only do sequence order effects vary according to the procedural characteristics of the task; they can also vary with participant characteristics. Moreover, certain combinations of task and participant characteristics can result in primacy effects in EBDM—a rare example of which is seen in Experiment 1. One possibility is that when presented with a task with a highly concrete or transparent structure, our adult participants tended to discount or ignore later observations on the grounds of redundancy or irrelevance.

What other age-related effects are found in EBDM?

Age related effects were examined for three other aspects of behaviour: Risk preference, the predictability of choice from sampled information in EBDM, and sampling patterns in EBDM.

(a) *Risk preference*. In each experiment, younger participants were more likely to choose the risky option than older participants. This difference was medium-sized and significant in Experiment 1, $F(1,97) = 11.07$, $p = .001$, partial $\eta^2 = .10$, small and significant in Experiment 2, $F(1,148) = 4.05$, $p = .046$, partial $\eta^2 = .03$, but non-significant in Experiment 3 where the age difference between participant groups was least, $F(1,69) = 1.39$, $p = .243$. This is in keeping with several developmental investigations of risk-taking behaviour and risk preference: Risk aversion often increases with age (Harbaugh, Krause, & Vesterlund, 2002; Reyna & Farley, 2006). Similarly we observed greater differences in risk preferences when the age differences between groups was larger. This age-related difference did not vary systematically or significantly with the mode of presentation (description or experience). This continuity across conditions represents a potentially interesting similarity between described and experienced decisions. We examined this further for Experiment 3, where participants had made both described and experienced choices. Even though sampling variability ensured that experienced and described problems were not identical, some consistency in risk preferences was apparent. This was shown by a near-significant correlation between the number of risky options chosen in the experience condition and the number of risky options chosen in the description condition ($r = .22$, $p = .067$; $r = .25$ for adolescents, $r = .16$ for young adolescents).[4] Koritzky and Yechiam (2009) also report a similar (slightly stronger) degree of consistency in risk preference across the D-E divide. Consistent individual differences in risk preference across different tasks and domains are surprisingly illusive (Yates, 1992). Therefore, although the relationship is not strong, such consistency is noteworthy, as correlation across tasks potentially signifies shared process(es) or dispositions.

(b) *Predictability on the basis of observed value*. Inspection of Table 3 shows that, in each experiment, the older group's choices were more closely aligned to the EV of observations than those of the younger participants. This accords with performance in the Iowa Gambling Task, where the proportion of good options that participants select increases with age through childhood and adolescence (Huizenga, Crone, & Jansen, 2007). It is inconsistent with Reyna's proposal that children are more calculative than adults in their decision making (Reyna & Ellis, 1994; Reyna & Farley, 2006). The age effects that we observe would be predicted on the basis of any account of cognitive development, general or specific. With increasing age, participants should become more able to assimilate observations, which could reflect an increase in general ability and/or improvement in some specific ability (e.g. memory skill). However, given the rather large difference in cognitive capacity that would be expected between our groups, these differences are surprisingly modest. For instance, even where the age difference was largest (Experiment 1) the older group were not (quite) significantly more sensitive to observed EV, $F(1,49) = 3.09$, $p = .085$, partial $\eta^2 = .06$. This

[4]Evidence of consistency in risk preferences across both problem formats was also apparent for the adolescent participants in Experiment 2 (who also completed both conditions—see footnote 2). The correlation between the number of risky options chosen in the experience condition and the number of risky options chosen in the description condition was significant, $r = .30$, $p = .008$.

DOI: 10.1002/bdm

Table 4. Mean number of alternations in the sampling phase by age group

	Age group		Difference (younger minus older)			
	Younger	Older	Effect size (*d*)	*t*-value	*df*	*p*-value
(a) Experiment 1	12.33	3.87	1.62, very large	5.78	49	<.001
(b) Experiment 2	9.92	5.96	0.79, large	3.40	73	.001
(c) Experiment 3	7.03	4.80	0.47, medium	2.00	69	.0499

may imply that decisions from samples do not place particularly high cognitive demands on the decision maker. Alternatively, if choices are driven by risk attitudes that eschew EV maximisation, EV will be a poor predictor of choice. For instance, choices in Experiment 1 were highly risk seeking, especially for the younger participants. If participants are drawn strongly to the option with the best possible outcome, observed EV can play only a small role in choice when options are equated for *a priori* EV.

(c) *'Sampling style'—the alternation between options when acquiring information.* This represents the strongest, and perhaps most intriguing age-related effect in these experiments. We calculated the number of alternations between options during the pre-choice sampling phase for each EBDM problem presented to each participant. The minimum number of alternations possible is one (10 picks from one option followed by 10 from the other), and the maximum number of alternations possible is 19 (sampling alternately from the two options). There was a considerable discrepancy between the age groups on this measure: A difference that was substantial and statistically significant in each experiment (see Table 4). In Experiment 1, adult participants made only one alternation in 63% of samples, whereas children did this only 8% of the time. Children made nine or more alternations (i.e. on average, alternating at least every two 'picks') 73% of the time, with a strict alternating strategy (i.e. 19 alternations) being followed on 21% of occasions. In contrast, adults made nine or more alternations only 15% of the time, and followed a strict pattern of alternation on only 7% of occasions. In Experiment 2, the minimum number of alternations (i.e. one) was twice as common for adolescents as for children (39 vs. 18% of games), whereas strict alternation when sampling was twice as common for children (16 vs. 8%). Furthermore, even though the age difference was small, a single alternation when sampling was more common for the older group in Experiment 3 (62 vs. 30%). The effect sizes in Table 4 show a clear 'dose-response' relationship: The greater the age-difference between the groups ('dose') the larger the mean difference in the number alternations ('effect').

This greater tendency to flip-flop between alternatives when exploring options is reminiscent of the generally increased tendency in younger participants towards 'matching' in probability learning tasks (i.e. alternating between options; Derks & Paclisanu, 1967). However, to our knowledge, this is the first time that such an individual difference in the approach to option exploration has been noted in decisions from sampling. This phenomenon was therefore felt to be worthy of further investigation.

EXPERIMENT 4—MANIPULATING SAMPLING STYLE IN EXPERIENCE-BASED CHOICE

Experiments 1–3 failed to show clear age-related differences in the size of the D-E gap, but did point towards other potentially important age-related effects. First, older participants were more risk averse than children. Second, the experience-based choices of younger children showed comparatively more recency (less primacy) than those of older participants. Third, EV could more readily predict these choices in the case of older participants (only a slight effect). Fourth, and most significantly, with increasing age there is a clear decreasing tendency to alternate between the options during pre-decisional information acquisition.

Further analysis indicated that these age-related differences in sampling style might play a role in some of the other effects. More chaotic patterns of information acquisition (with moderate numbers of alternations)

DOI: 10.1002/bdm

were associated with lower predictability in choice in Experiment 2. This was supported by a significant quadratic trend in the appropriate regression analysis ($\Delta R^2 = .064$; $p = .030$). In Experiment 3, the strongest primacy effects in choice occurred for infrequent alternation. Such effects make sense when one considers that sampling style should affect how easily new observations are assimilated or remembered, and how directly options can be compared. For instance, consider the following three possible styles of sampling, and the option evaluation processes implied by them. (1) Alternating only once (or infrequently) presumably helps the decision maker to evaluate each option separately before comparing one option against the other. (2) Alternating on every (or almost every) trial should allow the decision maker to evaluate the *difference* between the options directly, without necessarily forming separate valuations. (3) Intermediate numbers of alternations presumably represent more chaotic patterns of observation, which make tallying or totalling more difficult. If forgetting is selective, a chaotic sampling pattern should result in sequence order effects: Primacy if later observations are forgotten, recency if earlier ones are difficult to recall. These three modes of sampling were manipulated in Experiment 4.

Method

Two groups of participants were recruited: 46 male school pupils from Years 7 to 12 (mean age of 14.4 years, $SD = 1.8$ years, range 11–17 years with fairly uniform distribution), and 58 first year psychology undergraduate students (18 male, mean age of 20.7 years, $SD = 4.5$ years). The apparatus and procedure were identical to the experience condition of Experiment 3 except that participants were allocated to one of three conditions that determined the order in which they sampled the options prior to choice. Participants in the alternating condition followed a pattern of strict alternation between the options. In the random condition, the order of all 20 practice picks was randomized. In the exhaustive condition, 10 observations were made from one box, then 10 from the other. Nine (risky vs. sure thing) choice problems were used, all with rare event probabilities of .1 or .2. The EV of the risky and sure options was equal for three problems, higher in the risky option for three problems, and lower for the risky option for three problems. The samples were fixed such that rare events occurred with the correct expected frequency (e.g. 2 in 10 occasions when Pr(rare event) = .2) though the order of outcomes was randomized (including which box, ‘risky’ or ‘sure’, was selected for the first ‘pick’). Participants received payment contingent on the outcome of one of their choices (selected at random). Additionally, after making their choice for the final problem, participants were given a surprise test of memory for their observations in this final problem. They were asked to report the value of the outcomes that they had observed, which box each value was associated with, and the number of times that each outcome had been seen. Thus participants were not ‘cued’ with the information that they had observed one option with a single outcome value, and another with two outcome values; nor were they reminded of the values that they had seen or the number of observations made (cf. Ungemach, Chater, & Stewart, 2009). This approach was chosen to provide a stringent test of memory for recent observations. Memory was tested only after completion of the last problem—otherwise participants may have been tempted to treat the experiment as a memory task rather than as a choice task. Three standard measures of intellectual capacity/efficiency were obtained for the undergraduate participants: Raven’s advanced progressive matrices, digit span and working memory span. Time constraints precluded giving these tasks to the school students—but age was recorded to serve as a proxy measure of intellectual capacity.

Results and discussion

For each participant, we calculated the number of risky options chosen (/9), the number of choices that maximized EV (/6) and the number of choices consistent with overweighting small probabilities (/9). For each of these dependent variables (analysed separately), there was no significant effect of sampling mode (alternating vs. random vs. exhaustive): All $p > .21$, all partial $\eta^2 \leq .03$. These three dependent measures were

DOI: 10.1002/bdm

Table 5. Experiment 4: Recall (percentage correct) for frequencies observed prior to choice

	Condition (pattern for data acquisition)		
	Alternating	Random	Exhaustive
Risky option fully correct	34	24	34
Risky option fully or partially correct	56	54	51
Sure thing fully correct	44	38	49
Sure thing fully or partially correct	53	70	60

not significantly correlated with cognitive capacity (undergraduates) or with age (high school participants), all $|r| < .16$, all $p > .24$. The method for this experiment permitted a more sensitive analysis of order effects than that used in Experiments 1–3. The within-sequence position of a rare event was regressed onto choice (separate analysis for each problem). However, sequence–order effects were generally small and no clear differences between conditions in recency/primacy effects were apparent.

Recall for the final choice problem was scored separately for each option (sure and risky). Responses were coded as *fully correct* if the values seen and their respective frequencies were stated accurately (one frequency for the sure, and two for the risky option). Answers were scored as *partially correct* if the total number of points seen was reported (without correct frequencies), or when only one frequency was correctly reported for the risky option, or if the participant failed to report the frequency for the sure thing correctly but did indicate that the (correct) value was the only outcome that he/she observed from that box. Table 5 shows that, as predicted, recall was lowest in the random condition when the strictest scoring criterion was used (i.e. fully correct). However, these differences were not significant, even when the random condition was tested against the combined data from the other two conditions ($p = .290$ for the risky, and $p = .406$ for the sure option). When a more relaxed scoring criterion was used (i.e. partially or fully correct—allowing for 'gist memory') no deficit was apparent in the random condition. A total memory score was obtained for each participant by awarding two points for each fully correct answer and one for each partially correct answer (i.e. maximum of 4). In adult participants, memory score correlated moderately with digit span ($r = .32$, $p = .014$), weakly with working memory span ($r = .18$, $p = .194$) and was uncorrelated with Raven's matrices performance ($r = .00$). For the school-age participants, age barely predicted memory score ($r = .13$, $p = .393$). Mean memory score was higher for adults than for the school-age participants, but not significantly so, $F(1,98) = 1.70$, $p = .196$, partial $\eta^2 = .02$. Neither the effect of condition upon memory score, nor the age-group-by-condition interaction was significant ($F\text{s} < 1$).

GENERAL DISCUSSION—LESSONS FROM A DEVELOPMENTAL APPROACH

On the basis that working memory capacity increases through childhood and adolescence, we adopted a developmental approach as a strategy to test whether memory constraints are important in the appearance of the description-experience (D-E) gap. We did observe the D-E gap for choice problems involving relatively rare outcomes—but the size of this gap was not moderated by the age of the participant to any great extent. The interpretation of null findings is problematic, but is less so when accompanied by readily interpretable significant effects in the same investigation (Abelson, 1995). For instance, we found that younger participants made riskier choices than older participants in both described and experienced choices. This is a common (though not universal) finding in other studies (e.g. Harbaugh et al., 2002; Reyna & Farley, 2006), though, to the best of our knowledge, has not previously been demonstrated simultaneously for both described and experienced choices within the same investigation. Importantly, the average effect of age on risk preference was similar for both problem formats, and consequently the D-E gap did not vary much with age. Thus, whilst

DOI: 10.1002/bdm

we did observe age-related effects, there was little evidence that age-related differences in memory capacity play a substantial role in moderating the size of the D-E gap.

Our developmental investigation of EBDM yielded an unexpected and intriguing result. There were considerable age-related differences in the manner in which observations were collected prior to experience-based choice. Specifically, adult participants generally collected data separately for each option, whereas children generally made frequent alternations between the options in the data acquisition phase. Correlational analysis of Experiments 2 and 3 suggested that some of the modest age-related effects that we observed might be partly explained by these strong age-related differences in information acquisition. However, none of these relationships were corroborated when participants were randomly assigned to one of three modes of data acquisition (Experiment 4). Perhaps surprisingly, varying the mode of acquisition (quite markedly) has little effect upon memory for observations or patterns of choice. One unavoidable difficulty with Experiment 4 is that being *assigned* to collect data in a particular way may not have quite the same effect as *choosing* to collect data in that way. For instance, both Newell and Rakow (2007) and Rakow *et al.* (2008) found differences between active and passive observation in EBDM (e.g. between choosing to make observations and seeing observations sought by another participant). Therefore, it is possible that an experimental approach where observation patterns are determined for the participant may not equate perfectly with situations where decision makers chose these patterns of acquisition for themselves.

We therefore have an intriguing state of affairs. We have an interesting new finding: Namely, that there is a very strong developmental trend in patterns of information acquisition in EBDM. However, there is, as yet, no compelling evidence that this large effect is associated with other age-related effects or with patterns of choice (as might have been expected). However, this result highlights the importance of considering idiosyncratic patterns of data acquisition, especially where these are difficult to dissociate from patterns of choice. For instance, in tasks such as the Iowa gambling task (IGT), probability learning tasks (e.g. Goodnow, 1955) or decisions from feedback (e.g. Barron & Erev, 2003), participants acquire information and make choices simultaneously. Several studies report poorer choice performance by children (in comparison to adults) on the IGT (see Boyer, 2006). Our results highlight that the inability to settle on the best option may, at least in part, reflect a propensity to explore options with frequent movements between the options. Thus, what is thought to be a chaotic pattern of choice may be more appropriately classified as a chaotic pattern of information acquisition (see Yechiam & Druyan, 2007, for an example of a group of adult participants who display chaotic patterns of acquisition/choice in the IGT).

Our studies confirm that EBDM can be investigated successfully in child and adolescent populations, and, therefore, that a developmental approach can be used to test theoretical claims. In many respects, older and younger participants differed little in their choices or in the characteristics of their decision making. However, we observed a strong age-related trend in ‘sampling style’ during pre-decisional information acquisition. Younger participants were also more likely to choose a risky option over a sure thing, which occurred to a similar degree for described and experienced choices. The size of the D-E gap was seen to vary systematically and predictably with payoff probabilities: Increasing as the probability of the least likely outcome decreased. However, contrary to the conjecture that limitations in memory capacity should increase the tendency to give less weight to rare outcomes in EBDM, the D-E gap was not reliably larger for younger participants.

ACKNOWLEDGEMENTS

The work reported in Experiment 2 was supported by a University of Essex Teaching-and-Learning Bursary to the second author. The final versions of this manuscript were completed while the first author was a Visiting Fellow in the School of Psychology at the University of New South Wales, and supported by a Short Visit grant from The Royal Society. The authors thank the staff and pupils of Anson and Barham Primary Schools in London for their participation in Experiments 1 and 2. The authors thank Craig Scott for assistance with data collection and data management for Experiment 3. They thank Chris Graham and the students of Raleigh House at the Royal Hospital School for their participation in Experiment 4.

REFERENCES

Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Barron, G., & Erev, I. (2003). Small feedback-based decisions and their limited correspondence to description-based decisions. *Journal of Behavioral Decision Making*, *16*, 215–233.

Boyer, T. W. (2006). The development of risk-taking: A multi-perspective review. *Developmental Review*, *26*, 291–345.

Boyer, T. W. (2007). Decision-making processes: Sensitivity to sequentially experienced outcome probabilities. *Journal of Experimental Child Psychology*, *97*, 28–43.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision Field Theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459.

Camilleri, A. R., & Newell, B. R. (2009). Within-subject preference reversals in description- and experience-based choice. *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, N. Taatgen, J. van Rijn, J. Nerbonne & L. Schomaker (Eds.), Austin, TX, Cognitive Science Society.

Derks, P. L., & Paclisanu, M. I. (1967). Simple strategies in binary prediction by children and adults. *Journal of Experimental Psychology*, *73*, 278–285.

Erev, I., Ert, E., Roth, A., Haruvy, E., Herzog, S., Hau, R., et al. (2009). A choice prediction competition, for choices experience and from description. *Journal of Behavioral Decision Making.*

Fox, C. R., & Hadar, L. (2006). ‘‘Decisions from experience’’ = sampling theory + prospect theory: Reconsidering Hertwig, Barron, Weber & Erev (2004). *Judgment and Decision Making*, *1*, 159–161.

Goodnow, J. J. (1955). Determinants of choice-distribution in two-choice situations. *American Journal of Psychology*, *68*, 106–116.

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.

Hau, R., Pleskac, T. J., & Hertwig, R. (2009). Decision from experience and statistical probabilities: Why they trigger different choices than *a priori* probabilities. *Journal of Behavioral Decision Making.*

Hau, R., Pleskac, T. J., Kiefer, J., & Hertwig, R. (2008). The description-experience gap in risky choice: The influence of sample size and experienced probabilities. *Journal of Behavioral Decision Making*, *21*, 493–518.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*, 534–539.

Hitch, G. J., & Towse, J. N. (1995). Working memory: What develops?. In F. E. Weinert , & W. Schneider (Eds.), *Memory performance and competencies: Issues in growth and development* (pp. 3–21). Mahwah, NJ: Lawrence Erlbaum Associates.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: the Belief Adjustment Model. *Cognitive Psychology*, *24*, 1–55.

Huizenga, H. M., Crone, E. A., & Jansen, B. J. (2007). Decision-making in healthy children, adolescents and adults explianed by the use of increasingly complex proportional reasoning rules. *Developmental Science*, *10*, 814–825.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decisions under risk. *Econimetrica*, *47*, 263–291.

Kareev, Y., Lieberman, I., & Lev, M. (1997). Through a narrow window: Sample size and the perception of correlation. *Journal of Experimental Psychology: General*, *126*, 278–287.

Koritzky, G., & Yechiam, E. (2009). On the robustness of description and experience based decision tasks to social desirability. *Journal of Behavioral Decision Making.*

Levin, I. P., & Hart, S. S. (2003). Risk preferences in young children: Early evidence of individual differences in reaction to potential gains and losses. *Journal of Behavioral Decision Making*, *16*, 397–413.

Newell, B. R., & Rakow, T. (2007). The role of experience in decisions from description. *Psychonomic Bulletin and Review*, *14*, 1133–1139.

Prelec, D. (1998). The probability weighting function. *Econometrica*, *66*, 497–527.

Rakow, T., Demes, K. A., & Newell, B. R. (2008). Biased samples not mode of presentation: Re-examining the apparent underweighting of rare events in experience-based choice. *Organizational Behavior and Human Decision Processes*, *106*, 168–179.

Reyna, V. F., & Ellis, S. C. (1994). Fuzzy-trace theory and framing effects in children’s risky decision making. *Psychological Science*, *5*, 275–279.

DOI: 10.1002/bdm

Reyna, V. F., & Farley, F. (2006). Risk and rationality in adolescent decision making: Implications for theory, practice and public policy. *Psychological Science in the Public Interest*, *7*, 1–44.
Sarin, R., & Vahid, F. (2001). Predicting how people play games: A simple dynamic model of choice. *Games and Economic Behavior*, *34*, 104–122.
Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.
Ungemach, C., Chater, N., & Stewart, N. (2009). Are probabilities overweighted or underweighted when rare outcomes are experienced (rarely)? *Psychological Science*, *20*, 473–479.
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.
Yates J. F. (Ed.). (1992). *Risk taking behavior.* Chichester, England: Wiley.
Yechiam, E., & Druyan, M. (2007). Poor theory of mind is associated with erratic choice behavior but not IQ: Findings in Aspergers and healthy subjects. Warsaw: *SPUDM 21, conference paper.*

Authors' biographies:

Tim Rakow is a Senior Lecturer in Psychology at the University of Essex. His research interests include decisions from experience, pre-decisional information acquisition, probability judgment, expertise, risk communication and decision making in healthcare and medicine.

Saleema B. Rahim was an undergraduate psychology student at the University of Essex when this research was conducted. She is currently undertaking an MSc in Psychological Research at the University of Oxford.

Authors' addresses:

Tim Rakow and **Saleema B. Rahim,** Department of Psychology, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK.

DOI: 10.1002/bdm