Journal of Behavioral Decision Making
J. Behav. Dec. Making, **22**: 378–389 (2009)
Published online 16 March 2009 in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/bdm.638

When the Best Appears to Be Saved for Last: Serial Position Effects on Choice

YE LI and NICHOLAS EPLEY*
University of Chicago, USA

ABSTRACT

Decision-makers often evaluate options sequentially due to constraints on attention, timing, or physical location of the options. Choosing the best option will therefore often depend on people's memories of the options. Because imperfect recall introduces uncertainty in earlier options, judgments of those options should regress toward the category mean as memory decays over time. Relatively desirable options will therefore tend to seem less desirable with time, and relatively undesirable options will tend to seem less undesirable with time. We therefore predicted that people will tend to select the first option in a set when choosing between generally undesirable options, and will tend to select the last when choosing between generally desirable options. We demonstrate these serial position effects in choices among paintings, American Idol audition clips, jellybeans, and female faces, provide evidence of its underlying mechanism, and explain how these findings build on existing accounts. Copyright © 2009 John Wiley & Sons, Ltd.

KEY WORDS memory; choice; order effects; decision making; preferences

INTRODUCTION

Making decisions requires people to choose between options. In an ideal world, people's options would be presented simultaneously for easier evaluation, but in the real world, people's options are often presented individually over time. People cannot simultaneously visit multiple houses for sale, interview multiple job candidates, or eat at multiple restaurants, but instead must consider each object in isolation, and subsequently evaluate their experiences after seeing all the options. People's choice of the best option in these cases will therefore be based on how they reconstruct their evaluation of the stimulus at the time of judgment (Bettman, Luce, & Johnson, 1998; Hastie & Park, 1986; Payne, Bettman, & Johnson, 1992; Slovic, 1995). Any systematic biases in that reconstruction process are likely to produce systematic biases in evaluation and choice.

* Correspondence to: Nicholas Epley, Booth School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA. E-mail: epley@chicagobooth.edu

Perhaps the most basic bias introduced by this reconstruction process is that one's immediate online evaluation of a stimulus is likely to be imperfectly correlated with one's more distant recall of an experience (Johnson & Sherman, 1990; Van Boven & Ashworth, 2007). Whenever two variables are imperfectly correlated, extreme scores on one variable tend to be matched by less extreme scores on the other variable (Galton, 1886). An intense positive experience, for instance, is likely to seem more neutral in memory, just as an intense negative experience is likely to seem more neutral in memory (Van Boven, White, & Huber, 2008). And students who perform exceptionally well on an exam are likely to underestimate their performance relative to others whereas students who perform exceptionally poorly are likely to overestimate their performance (Burson, Larrick, & Klayman, 2006; Kruger & Dunning, 1999; Moore & Small, 2007). This regression to the mean is a statistical truism determined by the strength of the correlation between two variables. Factors that weaken the correlation should increase the degree of regression observed.

When options in a choice set are presented sequentially over time, both the simple delay of time as well as interference from other cognitively demanding tasks are likely to increase uncertainty in one's evaluations of more distant options compared to more recent options, and should therefore increase the amount of regression to an overall category mean. If, for instance, a person interviews three equally outstanding job candidates over 3 weeks, he or she is likely to remember that the most recent candidate was highly qualified but be relatively less certain about more distantly interviewed candidates. Reconstructing one's evaluation of a candidate recalled with less certainty is likely to produce a regressive evaluation closer to the average job candidate—in this case a relatively more negative evaluation. Because one's more certain evaluation of the most recent candidate is not disadvantaged by this regression to the mean, he or she may be more likely to be hired than the earlier candidates.

Such regression in sequentially presented options has very clear implications for people's decisions depending on the overall desirability of the options. In particular, it predicts that when making choices among generally desirable options and such as particularly outstanding job candidates and, those that appear early in the sequence are likely to be evaluated less favorably than options that appear more recently in the series. In contrast, when making choices among undesirable choice sets and such as particularly mediocre job candidates and, those that appear early in the sequence should be evaluated more favorably compared to those that appear later in the sequence. People therefore systematically prefer (and choose) options presented later in a sequence when selecting among generally desirable options, but prefer (and choose) options presented earlier in a sequence when selecting among generally undesirable options.

Existing research demonstrates that this pattern of choices may be produced by a very different psychological mechanism of feature-based comparisons. In particular, this research demonstrates that people tend to compare choice options based on unique features rather than on shared features. A series of options with uniquely negative features will therefore become progressively less desirable across the sequence of options whereas options that have uniquely positive features will become progressively more desirable across the sequence (Bruine de Bruin, 2005, 2006; Bruine de Bruin & Keren, 2003; Dhar & Sherman, 1996; Hodges, 1997; Houston, Sherman, & Baker, 1989; Mantel & Kardes, 1999). For example, Bruine de Bruin and Keren (2003) presented participants with lists of traits describing three different apartments. In the positive condition, each apartment had four common negative traits (e.g., each had old carpeting) and five unique positive traits (e.g., one was near the park whereas another had nice roommates). In the negative condition, each apartment instead had four common positive traits but five unique negative traits. Participants rated later apartments as more attractive in the positive condition and less attractive in the negative condition. This ''direction-of-comparison'' effect was found for evaluations measured immediately after each stimulus, at the end of the sequence, and even when the stimuli were presented simultaneously.

Our regression hypothesis builds on these results by proposing a more general mechanism that goes beyond what a feature-based comparison mechanism would predict because it is not limited to contexts, domains, or presentation modes that enable or require explicit feature-based comparisons. We therefore

DOI: 10.1002/bdm

tested our hypotheses in experiments using stimuli that are unlikely to be decomposed into their constituent features, such as the taste of a jellybean or the attractiveness of a face.

Our regression hypothesis also makes unique predictions about when order effects will emerge in sequential choices and when they will not. Regressive evaluations are likely to depend critically on memory decay, as details of an evaluation are lost and need to be constructed on the basis of more general category-dependent information. This mechanism therefore predicts that the order effects in choice emerge at the time of choosing rather than at the time of stimulus encoding. Serial order effects should therefore disappear if all options are presented simultaneously, or if the preference for each option is recorded immediately following its presentation. Existing feature-based comparison accounts, in contrast, predict that serial order influences evaluations at the time a stimulus is encoded.

Finally, our regression hypothesis predicts that the order effects in evaluation produced by time delays in presentation should remain even if the ability to engage in explicit comparisons is eliminated, as long as the passage of time is maintained. Feature-based comparison mechanisms require a within-participant comparison of relevant options in a choice set in order to produce differing evaluations, such as the same interviewer comparing a second job candidate directly with a first job candidate. A regression-based mechanism predicts that serial order effects are produced not by the explicit comparison of choice options but rather by regression in the evaluation of stimuli over a time delay. Thus, the same order effects would emerge *between* participants as long as the time delay in evaluation is maintained, such as with one interviewer evaluating a single job candidate interviewed more recently (or second) and another interviewer evaluating another single job candidate interviewed more distantly (or first). We believe that both feature-based comparisons and regressive evaluations operate in choices in everyday life. Our scientific purpose is not to invalidate the former mechanism, but rather to investigate the existence of the latter mechanism.

We tested our hypothesis that people will prefer later options in desirable choice sets and earlier options in undesirable choice sets, based on regressive evaluations, in six experiments. In each experiment, participants saw desirable or undesirable sets of stimuli and indicated which they liked the best (either by reporting their liking for the stimulus or choosing their favorite). Some experiments (1a–1c and 3) manipulated the desirability of the choice set by altering the stimuli themselves, whereas others (2a, 2b) held the choice set constant and manipulated desirability by comparison to an ‘‘example’’ set of highly desirable or undesirable stimuli. These experiments tested a regressive evaluation mechanism in two different ways: (1) By precluding regressive evaluations by either presenting all stimuli simultaneously or eliciting preferences immediately following presentation of each stimulus (Experiments 2a and 2b), and (2) by retaining the passage of time but eliminating the possibility of explicit comparisons by presenting stimuli in isolation and manipulating the time between stimulus presentation and evaluation (Experiment 3). These experiments collectively demonstrate that the best option appears to be saved for last when choosing among desirable choice sets, but the best seems to come first when choosing among undesirable choice sets.

EXPERIMENTS 1A–1C

Participants viewed good or bad paintings (Experiment 1a), listened to good or bad songs (Experiment 1b), and ate good or bad jellybeans (Experiment 1c). A short filler task followed each stimulus in order to separate them in consistent intervals over time, to control the amount of thought paid to the critical stimuli during the intervening time periods, and to increase the memory decay over the short time delay. When finished viewing, listening, or eating, participants evaluated how much they liked each stimulus, and chose which one was their favorite of the choice set. We predicted that participants would be more likely to prefer the later option in the good choice sets, but be more likely to prefer the early option in the bad choice sets.

DOI: 10.1002/bdm

Method

Participants

Participants were anonymous visitors to a public website ($N = 120$, Experiment 1a), and University of Chicago undergraduates and staff ($N = 48$, Experiment 1b; $N = 60$, Experiment 1c).

Procedure

In Experiment 1a, participants completed an online questionnaire in which they were shown either three desirable or three undesirable paintings, in a fully counterbalanced order. These paintings were, respectively, the three most liked and three most disliked paintings from a pretest of website visitors ($N = 705$) who collectively evaluated a large pool of paintings. Each painting appeared on the screen for 8 seconds. After the first and second paintings, participants completed anagrams for 2 minutes as a filler task. After the final painting, participants were asked to select which painting they liked the most (1st, 2nd, or 3rd). Participants also rated how much they liked each painting by clicking the appropriate point on a continuous line ranging from ‘‘strongly dislike’’ to ‘‘strongly like’’ with the midpoint labeled ‘‘neutral’’. The order of these two measures was counterbalanced. Finally, participants reported how well they remembered each painting from 0 (*not at all*) to 10 (*very well*).

In Experiment 1b, participants in the laboratory listened to 20-second clips from either three good or three bad songs, in a fully counterbalanced order. The three good clips were taken from (once) popular songs (‘‘The Beatles—Penny Lane,’’ ‘‘Buddy Holly—Everyday,’’ and ‘‘Jimmy Cliff—I Can See Clearly Now’’). The three bad clips were taken from ‘‘The Worst of American Idol Auditions: Seasons 1–4.’’ After the first and second clips, participants solved riddles for 90 seconds as a filler task. After the final clip, participants indicated which song they liked the most (A, B, or C), and rated how much they liked each clip from 0 (*not at all*) to 7 (*very much*).

In Experiment 1c, participants were approached at a dining hall and tasted either three good or three bad jellybeans, in a fully counterbalanced order. The good jellybeans were three flavors of ordinary Jelly Belly brand jellybeans (very cherry, raspberry, and blueberry). The bad jellybeans were three flavors of the Bertie Bott’s Every Flavor Beans (dirt, earwax, and grass). After tasting the first and second jellybeans, participants generated anagrams for 1 minute as a filler task. After tasting the final jellybean, participants indicated which jellybean they liked the most (A, B, or C), and then rated how much they liked each jellybean from 1 (*not at all*) to 10 (*very much*).

Results and discussion

Responses from one outlier on the liking measures were excluded from Experiment 1a, leaving 119 participants in that analysis. Neither filler task performance in any of the experiments nor question order in Experiment 1a meaningfully influenced any of the results. Standardized ratings and choice percentages for all experiments are reported in Table 1. In all of the experiments reported in this paper, our main prediction is for an interaction between presentation order and stimulus desirability. We tested this interaction in all experiments in an unordered 3 (presentation order: first, second, third) $\times$ 2 (stimulus desirability: good vs. bad) ANOVA, followed by focused contrasts to test the predicted pattern of that interaction.

In Experiment 1a (see Figure 1), participants reported liking the desirable paintings more overall than the undesirable paintings $F(1, 234) = 66.55$, $p < .0001$, and also reported remembering the most recent painting better ($M = 7.35$) than the earlier paintings (Ms $= 6.26$ and 6.29, 1st and 2nd, respectively), $F(2, 234) = 10.11$, $p < .0001$. More important, participants’ liking of the paintings showed the predicted order $\times$ desirability interaction, $F(2, 234) = 3.74$, $p < .05$. People tended to like the last painting more than the first among the desirable set, $F(1, 234) = 3.09$, $p < .10$, but tended to like the first more than the last among the undesirable set, $F(1, 234) = 4.40$, $p < .05$. Participants’ choices mirrored these liking ratings, with

DOI: 10.1002/bdm

Table 1. Participants' choice percentages and average ratings of stimuli for all conditions in all reported experiments

		Choice percentage			Ratings (*z*-scores)		
	Position	1	2	3	1	2	3
Study 1a: Paintings (*n* = 119)	Bad	45	32	23	−0.36	−0.43	−0.64
	Good	25	32	42	0.35	0.50	0.58
Study 1b: Song clips (*n* = 48)	Bad	50	33	17	−0.32	−0.60	−0.83
	Good	17	29	54	0.30	0.54	0.91
Study 1c: Jellybeans (*n* = 60)	Bad	23	47	30	−0.36	−0.36	−0.87
	Good	27	27	47	0.35	0.40	0.83
Study 2a: Paintings							
Sequential (*n* = 34)	Bad	63	31	6	0.08	−0.14	−0.75
	Good	17	28	56	−0.32	−0.26	0.39
Online (*n* = 34)	Bad	32	53	16	−0.08	0.02	−0.26
	Good	28	28	44	−0.04	0.23	0.45
Simultaneous (*n* = 37)	Bad	6	29	65	−0.23	0.11	0.49
	Good	18	47	35	0.20	0.35	−0.26
Study 2b: Faces							
Sequential (n = 60)	Bad	50	32	18	0.17	0.15	−0.19
	Good	22	16	63	−0.06	0.03	0.49
Online (*n* = 62)	Bad	45	21	34	0.06	−0.09	−0.16
	Good	40	33	27	0.27	−0.01	0.05
Simultaneous (*n* = 59)	Bad	36	29	36	−0.43	−0.41	−0.28
	Good	32	44	24	0.22	0.10	−0.04
Study 3: One painting at a time (*n* = 360)	Bad				−0.26	−0.56	−0.73
	Good				0.44	0.45	0.65

Note: Ratings are reported as standardized *z*-scores.

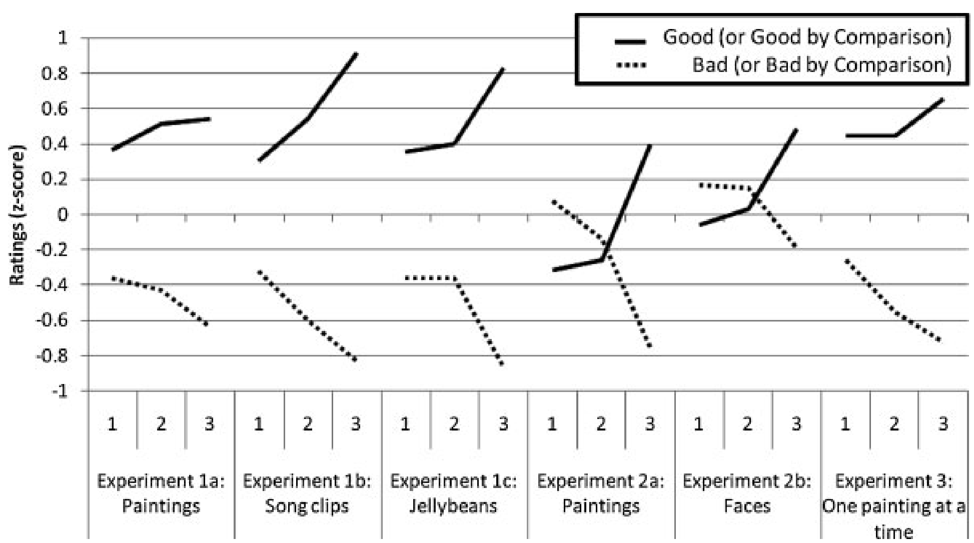

Figure 1. Participants' average ratings of stimuli by position (1, 2, 3) and good or bad stimulus condition for the sequential presentation conditions in all reported experiments. *Note*: Data are presented as standardized z-scores

DOI: 10.1002/bdm

participants being more likely to choose the last painting as the best among the desirable set but the first among the undesirable set, $\chi^2(2) = 6.52$, $p < .05$.

Participants in Experiment 1b also reported liking the good songs overall more than the bad songs $F(1, 92) = 55.05$, $p < .00001$. More important, liking ratings again showed the same order × desirability interaction, $F(2, 92) = 7.74$, $p < .001$. Participants liked the last good song better than the first good song, $F(1, 92) = 9.15$, $p < .01$, but liked the first bad song more than the last bad song, $F(1, 92) = 6.41$, $p < .02$. Participants' choices again mirrored these liking ratings, with participants tending to report that the last song was the best among the desirable set, but that the first song was the best among the undesirable set, $\chi^2(2) = 8.83$, $p < .02$.

Finally, participants in Experiment 1c reported liking the good jellybeans better overall than the bad jellybeans $F(1, 116) = 71.00$, $p < .00001$, again qualified by the predicted desirability × order interaction, $F(2, 116) = 6.92$, $p < .01$. Participants liked the last good jellybean better than the first good jellybean, $F(1, 116) = 5.10$, $p < .05$, but liked the first bad jellybean better than the last, $F(1, 116) = 5.78$, $p < .05$. Participants' choices tended to mirror these liking ratings with participants tending to choose the last jellybean as the best among the desirable set, but the first jellybean as being better than the last among the undesirable set, $\chi^2(2) = 8.37$, $p < .02$. The only unusual aspect of these choices is the relatively high number of people choosing the second jellybean among the undesirable set. Because this finding was unpredicted and unreplicated, we do not speculate about it further.

These experiments support our hypothesis of a systematic order effect in choice dependent on the desirability of the choice set. From paintings to songs to jellybeans, people evaluating desirable options tended to like (and choose) the last one in a series the best, but people evaluating undesirable options tended to do precisely the opposite. Experiment 1a provided weak evidence consistent with the regressive evaluation mechanism by simply demonstrating that people report better recall of the most recent item.

These experiments are also consistent, however, with other plausible interpretations. As discussed earlier, people may be engaging in an explicit comparison process as they are presented with each choice option. It is also possible that their actual sensory experience is changing over time, either because people are learning to differentiate between stimuli as they acquire more experience or because their sensory experiences are actually accumulating over time.

Notice that the key difference compared to a regressive evaluation mechanism is that these alternative interpretations occur at the time of stimulus encoding, whereas regressive evaluation occurs as a result of uncertainty about more distant stimuli at the time of recall and choice. We tested our mechanism more directly in Experiments 2a and 2b, eliminating this uncertainty by presenting some participants with all of their options simultaneously at the time of judgment (after first presenting them in a serial order), or by having some participants evaluate each stimulus immediately upon presentation. We predicted that these manipulations would eliminate the systematic effect of order on choice observed in Experiment 1.

EXPERIMENTS 2A AND 2B

Good and bad stimuli differed by identity in Experiment 1, with the good stimuli being objectively different from the bad stimuli. This introduces the possibility that differences in objective features of good versus bad stimuli could account for the results we observed. To avoid any such alternative interpretations, and to more conclusively demonstrate our hypotheses, the critical stimuli in Experiment 2 were identical stimuli rendered good or bad only by comparison to a baseline series of stimuli that were objectively better or worse. Participants in this procedure first saw a baseline series of good or bad stimuli, followed by neutral experimental stimuli. Based on simple contrast effects, we expected that the neutral stimuli would appear less desirable after seeing a good baseline series and more desirable after seeing a bad baseline series. Participants evaluated paintings in Experiment 2a and female faces in Experiment 2b. We again expected that participants

DOI: 10.1002/bdm

who saw the stimuli sequentially would prefer the last in the series of good (by comparison) paintings or faces, but would prefer the first in the series of bad (by comparison) paintings or faces. Consistent with a mechanism based on regressive evaluations, we predicted this effect would not emerge among participants who either saw the stimuli simultaneously at the time of judgment in the simultaneous condition or people who evaluated the stimuli immediately upon seeing the stimuli in the online condition (Hastie & Park, 1986).

Method

Participants

University of Chicago students ($N = 138$, Experiment 2a; $N = 112$, Experiment 2b) participated in a laboratory setting for monetary compensation.

Procedure

In Experiment 2a, participants were told that they would be evaluating paintings and that they would first see sample paintings to acclimate them to the task. Participants saw either six good paintings or six bad paintings (as judged by a separate sample of 36 participants) for 8 seconds each, and were told to use these as a ‘‘baseline.’’

In the sequential condition, participants completed the experiment following the same procedure as in Experiment 1a. In the online condition, participants rated how much they liked each painting immediately after seeing it (and before completing the same filler task as participants in the sequential condition). In the simultaneous condition, participants first saw each painting in a series just as participants in the sequential condition did, but then saw all three paintings on the screen side-by-side while making their evaluations.

In Experiment 2b, participants evaluated the attractiveness of female faces. Participants in the good (by comparison) condition first saw a baseline series of women with craniofacial syndrome (a genetic disorder that causes severe facial deformity), and those in the bad (by comparison) condition first saw a series of 2005 Miss Universe contestants. All participants then saw a series of female faces of average attractiveness (based on ratings from hotornot.com). Participants were randomly assigned to the sequential, simultaneous, or online conditions following the same procedure as Experiment 2a.

Results and discussion

As shown in Table 1, the sequential conditions in both Experiments 2a and 2b replicate the basic order effects already described, even though the stimuli in the good and bad conditions were objectively identical (see also Figure 1). This order effect was not observed in either experiment, however, in the simultaneous or online evaluation conditions.

In Experiment 2a, participants' evaluations in the sequential condition showed the predicted order $\times$ desirability interaction, $F(2, 64) = 6.18$, $p < .01$. Participants in the good (by comparison) condition tended to like the last painting more than the first, $F(1, 64) = 4.89$, $p < .05$, but those in the bad (by comparison) condition liked the first painting more than the last, $F(1, 64) = 4.89$, $p < .05$. Participants' choices mirrored these evaluations, with participants in the good (by comparison) condition generally selecting the last painting as their favorite and those in the bad (by comparison) condition generally selecting the first as their favorite, $\chi^2(2) = 11.05$, $p < .01$. Neither the simultaneous nor online condition showed this same pattern, indicated by a significant three-way interaction, $F(4, 198) = 5.79$, $p < .001$. The only significant effect of order and desirability was a significant effect in the *opposite* direction among participants in the simultaneous evaluation condition. Here, participants in the bad (by comparison) condition liked the recent

DOI: 10.1002/bdm

paintings less than the earlier paintings, whereas participants in the good (by comparison) condition liked the recent paintings more than the earlier paintings, $F(2, 64) = 3.54$, $p < .05$. We will not speculate on the reason for this unpredicted and unreplicated effect, but note that our prediction was simply for a different pattern of results in these conditions compared to the sequential condition, consistent with the regressive evaluation mechanism.

In Experiment 2b, participants' evaluations in the sequential condition again showed the predicted order × desirability interaction, $F(2, 116) = 6.57$, $p < .01$. Participants in the good (by comparison) condition rated the last face as more attractive than the first, $F(1, 116) = 8.66$, $p < .01$, whereas those in the bad (by comparison) condition rated the first face as more attractive than the last, $F(1, 116) = 3.23$, $p < .10$. Participants' choices again mirrored these liking manipulations, with participants in the good (by comparison) condition generally selecting the last face as the most attractive and those in the bad (by comparison) condition generally selecting the first face as the most attractive, $\chi^2(2) = 12.26$, $p < .01$.

Once again, the simultaneous and online conditions did not show this same pattern of evaluations, as indicated by a significant three-way interaction, $F(4, 350) = 3.23$, $p < .02$.

These two experiments extend our research by demonstrating a systematic effect of order on evaluation and choice depending on the desirability of the overall choice set, but in a context in which the evaluated stimuli were identical across experimental conditions. This procedure makes it clear that our predicted effects are not produced by some artifact of the stimuli being evaluated that differs across experimental conditions, but rather by the perceived desirability of the set itself. These order effects were eliminated in the simultaneous and online evaluation conditions, consistent with a regressive evaluation mechanism rather than alternatives based on explicit comparisons, some form of learning, or altered sensory experiences.

Note that the results for simultaneous and online evaluation appear to contradict Bruine de Bruin and Keren's findings (2003). Their Experiments 1 and 4, respectively, are comparable to our online and simultaneous evaluation conditions. In both the experiments, they find the same pattern of serial position effects as in sequential evaluation. However, this apparent contradiction in results is explainable due to differences in our experimental stimuli. Whereas Bruine de Bruin and Keren's stimuli are based on lists of features that are likely to enable (or enhance) explicit comparisons, the effects using relatively holistic visual stimuli in Experiments 2a and 2b are more readily explained by a mechanism based on regressive evaluations. In fact, it would be rather unnatural for participants to decompose paintings or faces into separate components or dimensions. With different mechanisms at work, different results need not be contradictory.

EXPERIMENT 3

We designed one final experiment to provide a maximally rigorous test of our regressive evaluation mechanism, and to distinguish it further from existing accounts of serial position effects dependent on feature-based comparisons. The critical determinant of choice for our regressive evaluation mechanism is not the presence of other stimuli, but rather the relative decay in memory of a given stimulus and the corresponding increase in uncertainty in recalled judgment. Although the presence of other stimuli in a choice set may enhance the choice effects observed in the preceding experiments due to retroactive interference, they should not depend on them. Thus, unlike the preceding experiments in which participants viewed either three good or bad stimuli in the first, second, *and* third positions, participants in Experiment 3 saw only *one* good or bad stimulus in the first, second, *or* third position while maintaining the two filler tasks. In place of the missing stimuli, participants simply pressed a button to continue the experiment. This design thus treats position as a between-participants variable by maintaining the time delay between presentation and evaluation, but eliminates the possibility of comparisons between stimuli. This design therefore precludes any explanation based on comparisons between stimuli (e.g., Bruine de Bruine & Keren, 2003), simply because each participant only sees one stimulus. If, however, memory decay over time leads to

DOI: 10.1002/bdm

regressive evaluations, then the stimuli presented in the "third" position between participants should, on average, be evaluated most positively among good stimuli but least positively among bad stimuli.

Method

Participants and procedure

Three hundred sixty anonymous visitors to a public website evaluated either one pleasant painting or one unpleasant painting. The procedure was identical to Experiment 1a except that we presented a painting in only one of the three possible positions (before, between, or after the two anagram tasks) mirroring the first, second, and third positions in Experiment 1a. Participants evaluated how much they liked the painting they saw, but had no choice to make. The painting presented and the position in which it was presented were fully counterbalanced across participants.

Results and discussion

As predicted by regressive evaluations, the pattern of results *across* participants mimicked that observed *within* participants in Experiment 1a (see Figure 1). Participants' evaluations of the paintings once again showed the predicted position × desirability interaction, $F(2, 354) = 4.76$, $p < .01$. Participants tended to like a painting in the last position more than a painting in the first position when evaluating good paintings, $F(1, 354) = 9.09$, $p < .01$, but showed a nonsignificant tendency in the opposite direction when evaluating bad paintings, $F(1, 354) = 1.78$, ns. Replicating this basic interaction pattern even in the absence of other choice options demonstrates that the pattern can emerge even when the possibility of comparison processes are completely eliminated.

GENERAL DISCUSSION

Choosing the best option in a decision set requires an evaluation of all options in the set. When options cannot be evaluated simultaneously, evaluations become vulnerable to biases arising from the presentation order. This research investigated one possible bias: As the details of the earlier experiences are lost and must be reconstructed when making a decision, evaluations of those experiences regress toward more moderate evaluations compared to the more recent experiences (Van Boven et al., 2008). When choosing the best of one's options presented over time, people should therefore tend to choose a more recent option when selecting among desirable options but choose a more distant option when selecting among undesirable options. We found this pattern across a variety of stimuli, including paintings, songs, jellybeans, and faces.

The present research complements and extends previous work on serial position effects in choice. Although similar patterns have been documented (Bruine de Bruin, 2005, 2006; Bruine de Bruin & Keren, 2003), these effects appear to be produced by explicit comparisons between shared and unique features that are likely to be more relevant when such features are clearly and saliently known and only when direct analytical comparisons are possible (Dhar & Sherman, 1996; Houston et al., 1989). The regressive evaluation mechanism we document expands these findings to a wider range of choices that require neither decomposable traits nor direct comparisons to function. Many choices of both large and small consequence are based on an overall affective evaluation of the options at hand rather than on analytical feature-based comparisons (Slovic, Finucane, Peters, & MacGregor, 2002). Indeed, in the experiments reported here, disabling the possibility of regressive evaluations by presenting stimuli simultaneously eliminated the effects

DOI: 10.1002/bdm

of presentation order on choice, as did asking participants to evaluate the options immediately following presentation (Experiment 2a and 2b). Instead of influencing how a choice option is encoded, regressive evaluations influence how a choice option is represented at the time of a decision.

We believe that our findings add to the growing recognition of the importance of time on choice and evaluation (e.g., Libby & Eibach, 2002; Trope & Liberman, 2003; Wilson & Gilbert, 2003). For instance, people tend to experience stronger emotions, both positive and negative, when anticipating events than when recalling them (Van Boven & Ashworth, 2007). The emotional impact of past events is blunted by the passage of time and what may once have been strong emotions fade to merely average ones. The affective evaluations measured in our experiments demonstrated similar patterns, and we therefore expect to observe similar findings in domains that do not involve choices at all. For instance, when we asked a sample of 28 students to report either their most positive or most negative personal event from the last 10-week academic quarter, 68% reported events from the last third of the quarter for both valences (vs. 18 and 14% for the first and second thirds, respectively), $\chi^2(2) = 7.79$, $p < .05$.

In accordance with our theory, we also expect the serial position effect to generalize to any attribute that needs to be reconstructed over time. For example, when lifting a series of heavy weights, people can easily recall that the most recently lifted weight was heavy but need to reconstruct the experiences of the earlier weights. This is likely to produce regressive estimates and create a sense that the most recent weight was the heaviest. Indeed, such ‘‘time-order’’ effects are among the oldest demonstrations in psychological science (Fechner, 1860), and have been documented for a wide variety of evaluations including estimates of duration (Vierordt, 1868), auditory intensity (Needham, 1935), line length (Tresselt, 1944), and pitch (Postman, 1947). These results are most widely explained as a process of physiological adaptation that, like serial position effects explained by feature-based comparisons, influences evaluations through a process of comparison at the time of encoding rather than at the time of evaluation. It is important to note, however, that these time–order effects propose physiological adaptation as a mechanism without testing it directly. The results of the experiments we have presented suggest a very different alternative, one that highlights the constructed nature of human judgment more than the comparative nature of experience. Attending more carefully to the mechanisms that produce serial position effects in choice will provide insights into when the order effects in choice are likely to be relatively large in everyday life and when they are not, and perhaps breathe new research life into one of the most long-standing topics in psychology.

The frequency of sequential choices in everyday life suggests that examples of serial position effects are likely to be quite common, even if typically unnoticed. For example, on the popular American Idol television show, the best singer from a series of final contestants is chosen by votes of viewers, but only at the end of each episode after all contestants have performed. Among these final good singers, those who came later in the show fared much better than those who came earlier. In our analysis of one season (Season 5), for each performance after theirs, a contestant’s odds of being voted into the bottom three were 22% higher and odds of being voted off were 12% higher (controlling for actual coded quality). Twenty-one of the last 25 Academy Award Best Picture winners were released in theaters in the second half of the year in consideration, 12 of which were released in the final 3 weeks of the year. Such effects in everyday life may stem from strategic attempts to capitalize on the kinds of regressive evaluations we have demonstrated, but they may also stem from them. Those charged with making decisions ranging from hiring to buying to voting may be well advised to remember that the best is not always saved for last.

ACKNOWLEDGEMENTS

The authors thank Gabriela Gondim, Megan Kolasinski, and Carlos Lozano for assistance conducting this research. The authors thank Wandi Bruine de Bruin and William Goldstein for helpful comments on an earlier version of this paper.

DOI: 10.1002/bdm

REFERENCES

Bettman, J. R., Luce, M. R., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of Consumer Research*, *25*, 187–217.

Bruine de Bruin, W. (2005). Save the last dance for me: Unwanted serial position effects in jury evaluations. *Acta Psychologica*, *118*, 245–260.

Bruine de Bruin, W. (2006). Save the last dance II: Unwanted serial position effects in figure skating judgments. *Acta Psychologica*, *123*, 299–311.

Bruine-de-Bruin, W., & Keren, G. B. (2003). Order effects in sequentially judged options due to the direction of comparison. *Organizational Behavior and Human Decision Processes*, *92*, 91–101.

Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology*, *90*, 60–77.

Dhar, R., & Sherman, S. J. (1996). The effect of common and unique features in consumer choice. *Journal of Consumer Research*, *23*, 193–203.

Fechner, G. T. (1860). *Elemente der Psychophysik* (Elements of Psychophysics). Leipzigm, Germany: Breitkopf & Härtel.

Galton, F. (1886). Regression towards mediocrity in hereditary stature. *Journal of the Anthropological Institute*, *15*, 246–263.

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, *93*, 258–268.

Hodges, S. D. (1997). When matching up features messes up decisions: The role of feature matching in successive choices. *Journal of Personality and Social Psychology*, *72*, 1310–1321.

Houston, D. A., Sherman, S. J., & Baker, S. M. (1989). The influence of unique features and direction of comparison on preferences. *Journal of Experimental Social Psychology*, *25*, 121–141.

Johnson, M. K., & Sherman, S. J. (1990). Constructing and reconstructing the past and the future in the present. In E. T. Higgins , & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2, pp. 482–526). New York: Guilford Press.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*, 1121–1134.

Libby, L. K., & Eibach, R. P. (2002). Looking back in time: Self-concept change affects visual perspective in autobiographical memory. *Journal of Personality and Social Psychology*, *82*, 167–179.

Mantel, S. P., & Kardes, F. R. (1999). The role of direction of comparison, attribute-based processing, and attitude-based processing in consumer preference. *Journal of Consumer Research*, *25*, 335–352.

Moore, D. A., & Small, D. A. (2007). Error and bias in comparative judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology*, *92*, 927–989.

Needham, J. G. (1935). The effect of time interval upon the time error at different intensive levels. *Journal of Experimental Psychology*, *18*, 539–543.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1992). Behavioral decision research: A constructive processing perspective. *Annual Review of Psychology*, *43*, 87–131.

Postman, L. (1947). Time-errors as a function of the method of experimentation. *American Journal of Psychology*, *60*, 101–108.

Slovic, P. (1995). The construction of preference. *American Psychologist*, *50*, 364–371.

Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2002). The affect heuristic. In T. Gilovich , D. Griffin , & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 397–420). New York: Cambridge University Press.

Tresselt, M. E. (1944). Time-errors in successive comparison of simple visual objects. *American Journal of Psychology*, *57*, 555–558.

Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review*, *110*, 401–421.

Van Boven, L., & Ashworth, L. (2007). Looking forward, looking back: Anticipation is more evocative than retrospection. *Journal of Experimental Psychology: General*, *136*, 289–300.

Van Boven, L., White, K., & Huber, M. (2008). Immediacy bias in perceptions of the intensity of different emotions over time. Manuscript submitted for publication (copy on file with author).

Vierordt, K. (1868). *Der Zeitsinn nach Versuchen* (The time-sense according to experiments). Tübingen, Germany: Laupp.

Wilson, T. D., & Gilbert, D. T. (2003). Affective forecasting. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 35, pp. 345–411). San Diego, CA: Academic Press.

DOI: 10.1002/bdm

Authors' biographies:

Ye Li is a doctoral candidate in Managerial and Organizational Behavior at the Booth School of Business, University of Chicago. His research interests fall under the umbrellas of judgment and decision-making and behavioral economics.

Nicholas Epley is a Professor of Behavioral Science at the Booth School of Business, University of Chicago. He is primarily interested in intuitive judgment, decision-making, and social cognition.

Authors' addresses:

Ye Li and **Nicholas Epley**, Booth School of Business, University of Chicago, 5807 South Woodlawn Avenue, Chicago, IL 60637, USA.

DOI: 10.1002/bdm