Original Article

QJEP

Quarterly Journal of Experimental Psychology
1–10

DOI: 10.1177/1747021819833904
qjep.sagepub.com
SAGE

Goal-dependent flexibility in preferences formation from rapid payoff sequences

Michael Brusovansky, Nira Liberman and Marius Usher

Abstract

The formation of attitudes or preferences for alternatives that consist of rapid numerical sequences has been suggested to reflect either a summation or an averaging principle. Previous studies indicate the presence of two mechanisms, accumulators (that mediate summation) and population-coding (that mediate averaging), which operate in preference formation tasks of rapid numerical sequences, and are subject to task-demands and individual differences. Here, we test whether participants can flexibly control the preference mechanism they deploy as a function of the reward contingency. Towards this aim, participants in two studies ($N_1 = 21$, $N_2 = 23$) made choices between the same sequence alternatives in two task-framing sessions, which made the reward dependent on the sequence-sum or sequence-average, respectively. The results demonstrate that although participants show an overall bias in favour of averaging, they are also remarkably flexible in deploying an averaging or a summation type mechanism that matches the reward contingency

Keywords

Decision making; intuitive preferences; averaging; summation; adaptivity; flexibility

Received: 27 September 2018; revised: 30 January 2019; accepted: 4 February 2019

Introduction

Intuitive preferences and attitudes based on sequential information are critical in determining human behaviour towards objects, products, or persons. For example, people form attractiveness impressions of stocks on the basis of sequences of returns (Betsch, Plessner, Schwieren, & Gütig, 2001), and of other people based on samples of observed behaviours or described traits (Anderson, 1981). Much research has examined the controlled information integration of attitudes and came to view it as a capacity-constrained sequential adjustment of an estimate towards a criterion (Hogarth & Einhorn, 1992). Yet, a number of studies have demonstrated that attitudes can also be formed as a result of implicit, non-symbolic, and automatic evaluation (Bargh, 1990; Schneider & Shiffrin, 1977). For example, Betsch and colleagues (2001) have demonstrated that human observers formed accurate attitudes towards alternatives (stocks—characterised as sequences of returns), which were presented as distractors and which the observers had no explicit goal to evaluate (see also Betsch, Kaufmann, Lindow, Plessner, & Hoffmann, 2006). More recently, Brusovansky, Vanunu, and Usher (2017) have shown that participants evaluated the subjective-value (attractiveness) of rapid sequences of numbers (of variable length, presented at two items per second, exceeding the capacity for symbolic computations[1]), which represented athletes competing in an athletics contest, in a way that reflected the mean of the numbers in the sequence. Critically, participants were not able to discount deviant numbers that were enclosed in a salient red frame and that were said to represent "computer-errors," although they expressed confidence that they did. As the ability to act on negation is one of the characteristics of controlled processing (Deutsch, Gawronski, & Strack, 2006; Gawronski & Bodenhausen, 2006), these results support an automatic/intuitive mechanism of integrating values to form evaluations. We note, though, that this criterion for automaticity is less strict than the one used in Betsch et al.,

School of Psychological Sciences and Sagol School of Neuroscience, Tel Aviv University, Tel Aviv, Israel

Corresponding authors:
Michael Brusovansky, School of Psychological Sciences and Sagol School of Neuroscience, Tel Aviv University, Haim Levanon 55, P.O. Box 39040, Tel Aviv-6997801, Israel.
Email: michaelbrus25@gmail.com

Marius Usher, School of Psychological Sciences and Sagol School of Neuroscience, Tel Aviv University, Haim Levanon 55, P.O. Box 39040, Tel Aviv-6997801, Israel.
Email: marius@post.tau.ac.il

which involved lack of any intention to process the presented numbers. We will return to the issue of automaticity and intentional encoding in the "Discussion" section.

A central question that was subject to some debate is the principle that drives such intuitive preferences. While some studies (Betsch et al., 2006; Betsch et al., 2001) suggested that value integration follows a *summative* principle, others showed results consistent with an *averaging* principle (Anderson, 1981; Brusovansky et al., 2017; Kahneman, 2011).[2] For example, Brusovansky and colleagues (2017) demonstrated a so-called "Jordan-effect" whereby participants gave higher ratings (to indicate "Hall of Fame" potential) to sequences (corresponding to careers of basketball players) with a high mean, compared with the same sequences to which few extra numbers were added (i.e., few extra seasons, increasing the sum but reducing the mean).[3] These results suggest that people might favour averaging even when summation is more appropriate. Furthermore, Brusovansky et al. (2017) examined the discrepancy between studies that supported summation and those that supported averaging, and proposed a critical factor that could explain the different results: the presentation and evaluation format. Accordingly, averaging dominates preferences in a *one-by-one* format, in which the values corresponding to each alternative appear sequentially and each alternative is evaluated before the next one is presented. By contrast, in a *grouped* format, in which several alternatives are presented together and then evaluated together, although averages still influence the attractiveness evaluations, sequence-length (and thus summation) also has a strong influence on evaluation (Experiment 4 in Brusovansky et al., 2017).

Summation and averaging (at a non-symbolic level) also seem to be supported by different mechanisms at the neuronal level. Specifically, summation may be mediated by neural accumulators (often assumed in models of choice and evidence integration; Brown & Heathcote, 2008; Roe, Busemeyer, & Townsend, 2001; Usher & McClelland, 2001, 2004; Vickers, 1970), whereas averaging may be supported by a population-coding mechanism (Brezis, Bronfman, Jacoby, Lavidor, & Usher, 2016; Brezis, Bronfman, & Usher, 2015; Malmi & Samson, 1983). Assuming that people can deploy both summation and averaging (as a non-symbolic process), which of them would be deployed in a specific case may be determined by characteristics of both individuals and the task. For example, a population-coding mechanism of averaging may be more difficult to deploy when the alternatives are presented and evaluated together, as this requires the maintenance and the separation of multiple reward distributions simultaneously (Brusovansky et al., 2017; Vanunu, Pachur, & Usher, 2018).

To sum up, the literature so far suggests that people are able to employ either averaging or summation when they form intuitive preferences or attitudes based on rapid sequential information, and that both averaging and summation can be performed automatically (in the "General Discussion" section, we offer a more detailed discussion of automaticity and specifically address the potential roles of conscious and intentional processing). The two mechanisms seem to rely on different neural mechanisms and to be favoured by different formats of information presentation and preference elicitation.

It remains unknown, however, whether in conditions that favour intuitive processing (e.g., rapid sequential presentation), participants can selectively deploy the mechanism (summation vs. averaging) that fits best with the rational considerations implied by the task framing. In a previous study from our lab (Brusovansky et al., 2017), averaging has been shown to contaminate performance in tasks that (*arguably*) call for summation (the "Jordan-effect"), suggesting that averaging is dominant relative to summation. In other words, averaging was performed not only spontaneously, when one does not have a goal to do so, but also, possibly, when one has a different, incompatible goal of assessing the sum. However, in that study, alternatives were presented one by one (which favours averaging; Experiments 1-3) and, in addition, it is possible that some of the participants thought that it is rational to average (see Control-Experiment in Supplementary Material). Moreover, these studies did not "give summation a fair chance" in that they did not look at whether summation too might contaminate averaging in tasks that call for averaging. In light of this, the question of whether averaging contaminates summation more than vice versa remains open.

Whereas in previous studies it was (intentionally) left somewhat ambiguous whether looking for the higher average or the higher sum is more appropriate (to contrast spontaneous inclinations), in the present study we set to manipulate the task framing to unequivocally support either averaging or summation. Towards this aim, we carried out two within-participants experiments, in which all participants were tested in two sessions, both of which presented the same choice problems but with different instructions (framing). Thus, we sought to contrast averaging and summation processes by introducing a procedure that allows both of these mechanisms to be deployed and by using explicit task framing that favoured either averaging or summation, while rendering the other strategy irrelevant. We also used a presentation mode that was in-between the one-by-one and the grouped conditions.[4] Specifically, we presented participants with rapid reward sequences of two slot-machines and told them to choose one assuming that they will get to play the chosen machine once (in which case, we thought, choice should rationally reflect average and ignore sum) or receive the sum of the rewards they observed (in which case, we thought, choice should rationally reflect the sum and ignore the average).

Table 1. Condition and total number of trials in Experiments 1 and 2.

Type of comparison and condition	Alternatives' sums	Condition label	No. of trials in Experiment 1	No. of trials in Experiment 2
(1) "40-6" vs. "60-9"	240 vs. 540	*Filler 1 (increasing average, N, and sum)*	20	10
(2) "40-6" vs. "60-6"	240 vs. 360	*Filler 2 (increasing average and sum)*	35	15
(2) "40-9" vs. "60-9"	360 vs. 540	*Filler 2 (increasing average and sum)*	35	15
(3) "40-6" vs. "40-9"	240 vs. 360	*Same-Average/Different-Sum*	35	35
(3) "60-6" vs. "60-9"	360 vs. 540	*Same-Average/Different-Sum*	35	35
(4) "40-9" vs. "60-6"	360 vs. 360	*Same-Sum/Different-Average*	50	40
(5) "40-11" vs. "60-5"	440 vs. 300	*Averages and Sums in Opposition*	–	50
Total number of trials			210	200

We examine (1) whether participants would be able to choose the strategy that is favoured by the framing, and (2) whether they would do so to the same extent with summation versus averaging. We examine Questions (1) and (2) both in situations in which one option dominates the other (i.e., the two alternatives have equal averages but different sums or, alternatively, equal sums but different averages) and in situations in which the two options trade-off average and sum (i.e., one option has higher average and lower sum, whereas the other option has higher sum and lower average). We also examine (3) individual differences. Specifically, we look at whether participants show (a) consistency in using summation versus averaging, (b) consistency in sensitivity to instructions that favour summation versus averaging, or (c) neither (i.e., show no consistency in either sticking to a particular strategy or in being attentive to framing). Two experiments tested these questions.

Experiment 1

The experiment presented rapid pairs of numerical sequences (describing rewards produced by two slot-machines), which we created so as to contrast between the sequence-average and the sequence-sum. To do this, we manipulated between the two alternatives either the sequence-average, the number of sequence values, or both. The same sequences were presented in the two sessions under different goal framings.

Method

Participants. In total, 21 students from Tel Aviv University (17 women, age: 18-36, M=23.10) participated in the experiment in exchange for course credit and payment that depended on performance. On average, participants received 20 NIS (~ US$6).

Materials. In each of two experimental sessions (sums vs. averages framings), participants were presented a series of decisions, in which they had to choose between two slot-machines. Each slot-machine sequence was presented as a series of either six or nine numbers, representing the possible winnings of a slot-machine. The average winnings could be either 40 or 60, so each slot-machine sequence is denoted by its average and length, that is, "40-9." These values were selected to maintain a similar ratio for the sequence-average and sequence-length manipulation (1.5).

The pairs that were presented in the experiment were constructed as all six possible combinations of pairs of different slot-machines. These were organised in four conditions (see Table 1): (1) "40-6" versus "60-9"—20 trials, (2) "40-6" versus "60-6" and "40-9" versus "60-9"—35 trials each, (3) "40-6" versus "40-9" and "60-6" versus "60-9"—35 trials each, and (4) "60-6" versus "40-9"—50 trials, for a total of 210 experimental trials (see Table 1). Condition (1) is a filler condition, as one of the alternatives ("60-9") dominates the other ("40-6") in both sequence-average (60 vs. 40) and sequence-length (9 vs. 6), or the sum (540 vs. 240). Condition (2) also involves a dominating alternative, but this time the average changes by a factor of 1.5, while the sequence-length is kept constant (thus the sum also increases by a factor of 1.5). Conditions (3) and (4) contain the critical comparisons. In Condition (3), only the sequence-length (and thus the sum) is increased by a factor of 1.5 (240-360 or 360-540), whereas the average stays the same (40 or 60), and in Condition (4), only the average increases by a factor of 1.5 (40-60), whereas the sum stays the same (360).

Procedure. Participants came for two 1-hr sessions—one with an averages framing and one with a sums framing. The order of the sessions was counterbalanced. The emphasis on sums versus averages was carried out in the instructions stage of each session, by telling participants how they should think of the slot-machines when making their choices. In the sums session, participants were told that when indicating their preference, they should assume they would receive all of its winnings that were presented on that particular trial. In the averages session, participants were told that they should assume they would play the slot-machine once and thus receive a single winning from it. To make sure participants take the goal-framing instructions seriously, they were also told that at the end of each

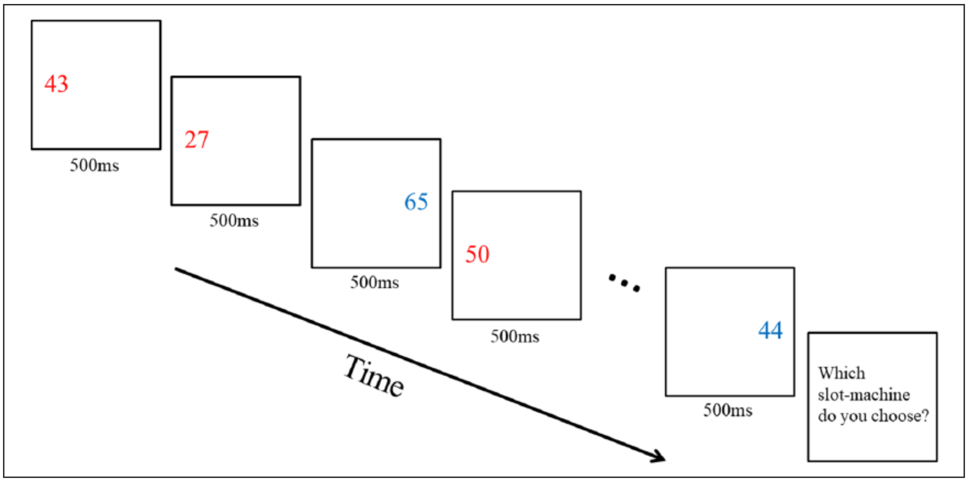

Figure 1. Timeline of a single trial in Experiments 1 and 2. Each square corresponds to the display screen (background colour was black instead of white), shown for 500 ms, and ending with a probe to select the preferred alternative (left vs. right).

session the computer will randomly choose one of the 210 trials they did, and the participants will receive a winning according to the slot-machine they chose on that trial, according to the rules of the session—all the chosen slot-machine's winnings, or a single one.

Before the experimental trials, participants completed one practice trial. For each experimental trial, the two slot-machines were assigned randomly to either the right side or the left side of the screen. The winnings of the slot-machine that appeared on the left always appeared in red colour and the winnings of the slot-machine that appeared on the right always appeared in blue colour—this was done to increase the segregation between the two alternatives. The allocation of the colours and spatial location on the screen for the two alternatives was chosen randomly before each trial. The numbers appeared sequentially on the screen, in a random order (500 ms/item, Inter-Stimulus-Interval of 100 ms), until all the winnings from both slot-machines were presented, and then participants had to indicate the slot-machine they choose on this trial (see Figure 1). The keyboard keys used for choosing the slot-machines on each trials were "q" and "p." After every 40 experimental trials, participants were granted a short break. Each session took approximately 60 min.

Analysis and contrasts. In the Filler Conditions (1) and (2), we expected that participants will prefer the dominating alternative. This "choice-accuracy" is thus reported as a measure of task engagement. The critical contrast involves the Conditions (3) and (4), in which either the average or the sum (but not both) distinguish between the two alternatives. Thus, to measure the preferences' sensitivity to the goal framing, we contrast the two sessions in each of these critical conditions. If participants are sensitive to goal framing, we expect that preference for the "High-Sum" alternative (Condition 3) will increase in the sum session, and also that the preference for the "High-Average" alternative (Condition 4) will increase in the average session.

Results

Because participants sometimes missed the correct response keys, the actual number of valid trials differed somewhat from participant to participant (on average, it was 208.62 out of 210). Therefore, for each possible comparison type, we looked at the percentages of choices for each alternative, instead of the actual number of choices.

Condition (1), increasing average, N, and sum: "40-6" versus "60-9"

In these trials, the "60-9" alternative dominates the "40-6" alternative both in terms of average and in terms of sum. Indeed, the majority of participants chose it at above-chance-level rate (50%), both under the averages framing, 91.99%, $t(20) = 20.69$, $p < .001$, Cohen's $d = 4.52$, and under the sums framing, 92.57%, $t(20) = 27.69$, $p < .001$, Cohen's $d = 6.05$, with no difference between the two framings, $t(20) = -0.27$, $p = .791$, Cohen's $d = 0.06$.

Condition (2), increasing average and sum: "40-6" versus "60-6" and "40-9" versus "60-9"

Here, too, alternatives "60-6" and "60-9" dominate alternatives "40-6" and "40-9," both in terms of average and in terms of sum. The majority of participants chose the dominating alternatives ("60-6," "60-9") at above-chance-level rate (50%), both under averages framing, 88.11%, $t(20) = 18.33$, $p < .001$, Cohen's $d = 4.00$, and under sums framing, 88.02%, $t(20) = 20.38$, $p < .001$, Cohen's $d = 4.45$, with no difference between the two framings, $t(20) = 0.06$, $p = .950$, Cohen's $d = 0.01$.

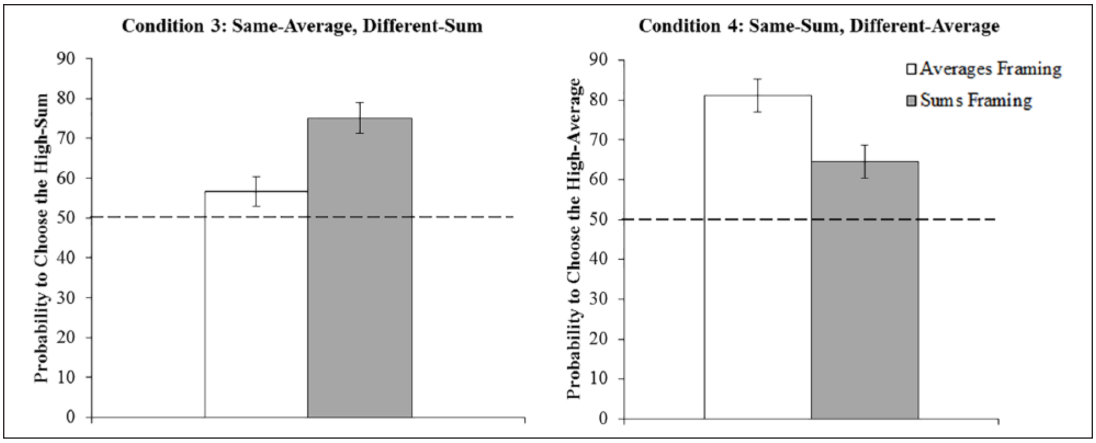

Figure 2. Percentages of choices in Experiment 1 for the "High-Sum" alternatives ("40-9" and "60-9," when compared with "40-6" and "60-6"—left panel) and of choices for the "High-Average" alternative ("60-6," when compared with "40-9"—right panel), as a function of the framings used. Dashed lines represent chance-level probability of choice. Error bars correspond to within-participant 95% confidence intervals (Cousineau, 2005).

Condition (3), Same-Average/Different-Sum: "40-6" versus "40-9" and "60-6" versus "60-9"

Both comparisons involve increasing the sum by a factor of 1.5 (240-360 and 360-540), while maintaining the same average. Therefore, we calculated for each participant the percentage of choosing the dominated alternative with six items ("Low-Sum"), over the dominating alternative with nine items ("High-Sum"). Overall, participants preferred the "High-Sum" alternatives at above-chance-level rate (50%), both under the averages framing, 56.67%, $t(20) = 2.33$, $p = .031$, Cohen's $d = 0.51$, and under the sums framing, 75.06%, $t(20) = 8.64$, $p < .001$, Cohen's $d = 1.89$. Critically, this preference for the dominating (higher sum) alternatives was higher under the sums framing, $t(20) = –4.60$, $p < .001$, Cohen's $d = 1.00$.

Condition (4), Same-Sum/Different-Average: "60-6" versus "40-9"

Overall, participants preferred the dominating "60-6" alternative ("High-Average") at above-chance-level rate (50%), both under the averages framing, 81.20%, $t(20) = 7.46$, $p < .001$, Cohen's $d = 1.63$, and under the sums framing, 64.61%, $t(20) = 4.34$, $p < .001$, Cohen's $d = 0.95$. Critically, the participants showed higher preference for this average-dominating alternative under the averages framing, $t(20) = 3.83$, $p = .001$, Cohen's $d = 0.84$ (see Figure 2, right panel).

Discussion

The results for Dominated-Trial Conditions (1) and (2) demonstrate that the participants engaged well in the task, as indicated by the high "choice-accuracy" (around 90% on average). More important, as shown in Figures 2 and 3, the comparison between the goal-framing sessions in the Critical Conditions (3) and (4) demonstrates that the participants are able to switch their reliance from averages to sums when evaluating the same alternatives, in accordance to the task rationale. When both alternatives had the same average but one alternative had a higher sum (Condition 3), participants preferred the "High-Sum" alternative more when they were instructed that the reward is based on all the rewards from a slot-machine (thus making the sum relevant and the average irrelevant). However, when they were instructed that rewards are based on a single sample from a slot-machine (thus making the average relevant and the sum irrelevant), their preference for the "High-Sum" alternative was not higher than for the "Low-Sum" alternative.

Quite the opposite pattern emerged when the alternatives had the same sum, but one alternative had a higher average (Condition 4). Here, participants showed more preference to the "High-Average" alternatives when they were instructed that the reward is based on a single sample, rather when they were instructed that the reward is based on all the presented rewards. However, under the sums framing, participants still preferred the "High-Average" alternative to the "Low-Average" alternative, despite the fact that they both had the same sum. This indicates that the averages affected preferences to some degree, even when they were not relevant for the goal and were not contingent on reward.

One possibility is that under the conditions of our experiment, the processing of the average outcome is more habitual and automatic. Note, however, that we lacked here the Critical Condition (5) that sets the sums and the averages in opposition.

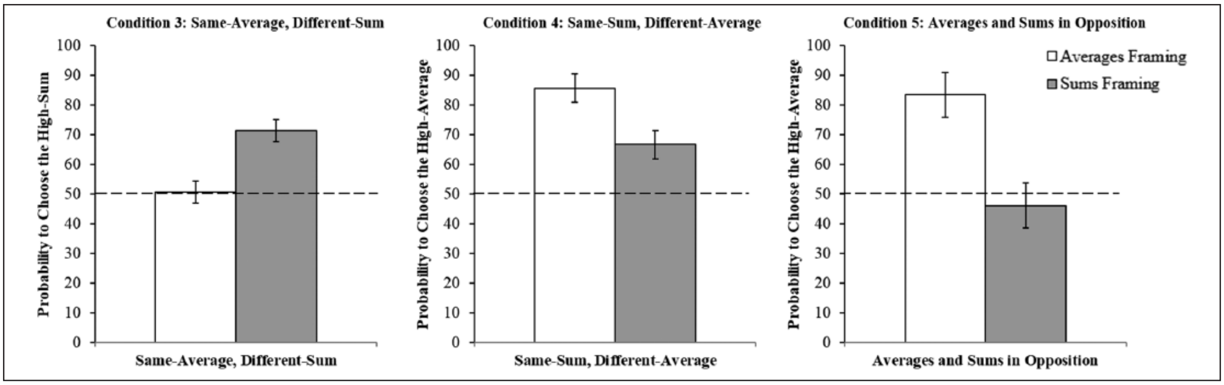

Figure 3. Percentages of choices in Experiment 2, as a function of the framings used, for Left panel—the "High-Sum" alternatives ("40-9" and "60-9"), when compared with "40-6" and "60-6," Middle panel—the "High-Average" alternative ("60-6"), when compared with "40-9," and Right panel—the "High-Average, Low-Sum" alternative ("60-5"), when compared with "40-11." Dashed lines represent chance-level probability of choice. Error bars correspond to within-participant 95% confidence intervals.

Experiment 2

In Experiment 2, we sought to replicate the findings of Experiment 1 and to test, in addition, whether a full "switch" between averaging and summation can be obtained when the two choice alternatives present a full trade-off between sum and average. We therefore included a new condition with one alternative in each choice pair having a higher average but a lower sum, and the other alternative a lower average but a higher sum.

Method

Participants. In total, 23 students from Tel Aviv University (20 women, age: 18-25, $M=22.0$) participated in the experiment in exchange for course credit and payment that depended on performance. On average, participants received 20 NIS (~ US$6).

Materials. Experiment 2 used the same framing manipulation and the same pairs of alternatives as Experiment 1, except that a different number of trials was allotted for each condition, and Condition 5 was added (see Table 1). Condition 5 involved a comparison of the alternative "40-11" with the alternative "60-5," which increases the average (from 40 to 60), but lowers the sum (from 440 to 300), thus allowing for a full reversal test—participants who are sensitive to the framing manipulation should prefer the "40-11" alternative under the sums framing but switch their preference to the "60-5" alternative under the averages framing. The number of trials for each condition was changed, to allow the addition of this condition without increasing the overall time of the experiment.

Procedure. The procedure of Experiment 2 was similar to Experiment 1. Each session took approximately 60 min.

Analysis and contrasts. Experiment 2 included the same Conditions 1 to 4 from Experiment 1, and therefore, similar results were expected in those conditions. The new Condition (5) involved setting the average and the sum in opposition, so we expected that if participants are sensitive to goal framing, then they should prefer the "High-Average" alternative ("60-5") under the averages framing, and prefer the "High-Sum" alternative ("40-11") under the sums framing.

Results and discussion

Like in Experiment 1, we examined the percentages of choices for each alternative, instead of the actual number of choices, because participants sometimes missed the correct response keys and the actual number of valid trials differed from participant to participant. On average, the number of valid trials per participant was 198.61 out of 200.

Condition (1), increasing average, N, and sum: "40-6" versus "60-9"

The majority of participants chose the dominating "60-9" alternative at above-chance-level rate (50%), both under the averages framing, 90.50%, $t(22)=11.88$, $p<.001$, Cohen's $d=2.48$, and under the sums framing, 88.42%, $t(22)=11.64$, $p<.001$, Cohen's $d=2.43$, with no difference between the two framings, $t(22)=-0.60$, $p=.56$, Cohen's $d=0.12$.

Condition (2), increasing average and sum: "40-6" versus "60-6" and "40-9" versus "60-9"

The majority of participants chose the dominating alternatives ("60-6," "60-9") at above-chance-level rate (50%), both under the averages framing (87.88%, $t(22)=11.84$, $p<.001$, Cohen's $d=2.47$, and under the sums framing, 82.95%, $t(22)=11.86$, $p<.001$, Cohen's $d=2.47$, with no difference between the two framings, $t(22) = 1.68$, $p = .108$, Cohen's $d = 0.35$.

Condition (3), Same-Average/Different-Sum: "40-6" versus "40-9" and "60-6" versus "60-9"

Similarly to Experiment 1, we calculated for each participant the average percentage of choosing the dominating alternative with nine items ("High-Sum") over the dominated alternative with six items ("Low-Sum"). Participants preferred the "High-Sum" alternatives at above-chance-level rate (50%) under the sums framing, 71.26%, $t(22) = 7.37$, $p < .001$, Cohen's $d = 1.54$. However, under the averages framing, this high-sum preference was not higher than chance level, 50.57%, $t(22) = 0.30$, $p = .769$, Cohen's $d = 0.06$, and it was lower than the preference for the "High-Sum" alternatives under the sums framing, $t(22) = –5.81$, $p < .001$, Cohen's $d = 1.21$.

Condition (4), Same-Sum/Different-Average: "60-6" versus "40-9"

Overall, Participants preferred the dominating "60-6" alternative ("High-Average") at above-chance-level rate (50%) both under the averages framing, 85.57%, $t(22) = 9.23$, $p < .001$, Cohen's $d = 1.92$, and under the sums framing, 66.57%, $t(22) = 4.07$, $p = .001$, Cohen's $d = 0.85$. Critically, the preference for this dominating-average alternative was higher under the averages framing, $t(22) = 4.13$, $p < .001$, Cohen's $d = 0.86$ (see Figure 3, middle panel).

Condition (5), Averages and Sums in Opposition: "60-5" versus "40-11"

Participants preferred the "60-5" alternative ("High-Average/Low-Sum") at above-chance-level rate (50%) only under the averages framing, 83.32%, $t(22) = 7.22$, $p < .001$, Cohen's $d = 1.50$. Under the sums framing, however, this preference was not different from chance level. Moreover, the preference for the high-average/low-sum alternative was higher under the averages framing than under the sums framing, 46.04%, $t(22) = 5.19$, $p < .001$, Cohen's $d = 1.08$ (see Figure 3, right panel).

Individual differences in sensitivity to sums and averages. While these results reflect some sensitivity to task framing at the group level, we may ask whether this sensitivity differs between individuals. To answer that question, we examined, for each participant, three indexes of sensitivity. The first two indexes relied on Conditions (3) Same-Average/Different-Sum and (4) Same-Sum/Different-Average (see Table 1). Specifically, we calculated, for each participant, Index-1 as the proportion of choices of the "High-Sum" alternatives in Condition (3) under the sum framing minus the same proportion under the average framing. Similarly, we calculated for each participant Index-2, as the proportion of "High-Average" choices in Condition (4) under the averages framing, minus the same proportion under the sums framing.

As we show in the left panel of Figure 4, these two indexes of sensitivity to task framing are highly correlated ($r = .713$, $p < .001$). Thus, participants who were affected by framing in choosing the "High-Sum" alternatives in Condition (3) were also affected by framing in choosing the "High-Average" alternative in Condition (4). We combined these two indexes together by averaging them, thus getting a single measure for each participant, based on the Conditions (3) and (4).

Finally, we computed another index of task-framing sensitivity, *Index-3*, based on the opposition Condition (5). This index is computed as the proportion of choices of the "High-Average/Low-Sum" alternative under the averages framing, minus the proportion of choices of the same alternative under the sums framing. As shown in the right panel of Figure 4, we found a very high correlation between the Index-3 measure of sensitivity and the one computed from Conditions (3) and (4): $r = .929$, $p < .001$.

The variability in sensitivity among participants raises the possibility that the non-significant preference of the "High-Sum" alternative in Condition (5) of the sums session (see Figure 3, right panel, grey bar) is due to the variability in the task-framing sensitivity among participants. Indeed, when we look at this same preference (Condition 5), separately for participants who showed higher sensitivity to framing and participants who showed lower sensitivity (median split), we see that the former group shows a preference reversal, whereas the latter group shows only a small modulation of choice with task framing (see Figure 5). In both groups, we also see a residual bias in favour of averaging.

General discussion

We have carried out two experiments, designed to test participants' ability to adhere to the task's rational considerations by switching between summation and averaging when facing decisions between the same rapid numerical sequence alternatives under two payoff-contingency conditions. In the first experiment, we showed that participants are indeed sensitive to the task rationale, and that they modulate their preferences according to whether it is rational to choose according to an averaging principle or according to a summation principle. When the two alternatives had the same sum but one had a higher average, participants were more likely to choose the alternative with the higher average when the task's rational consideration was to choose according to the averages. By contrast, when the two alternatives had the same average but one had a higher sum,

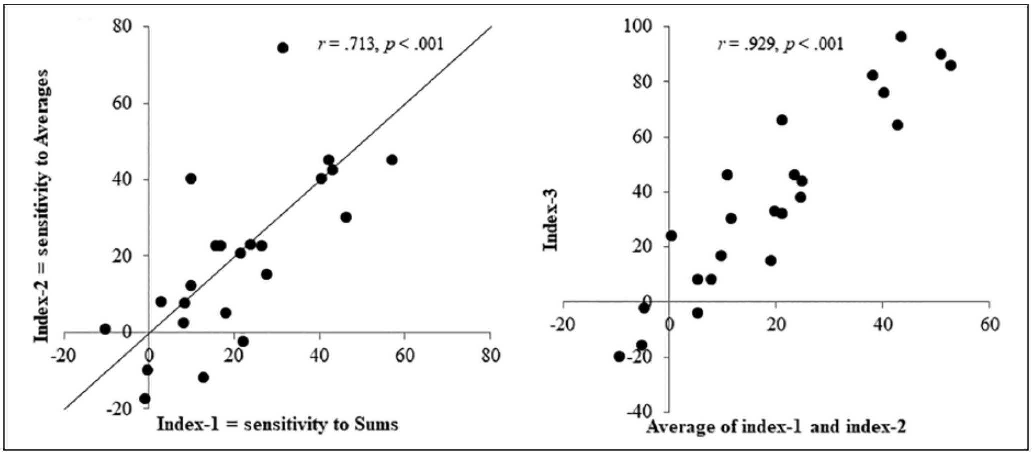

Figure 4. Individual differences in sensitivity to sums and to averages. Left panel: correlation between Index-1 (Condition 3, *x*-axis) and Index-2 (Condition 4, *y*-axis). Right panel: correlation between the average of Index-1 and Index-2 (*x*-axis), with Index-3 (Condition 5, *y*-axis).

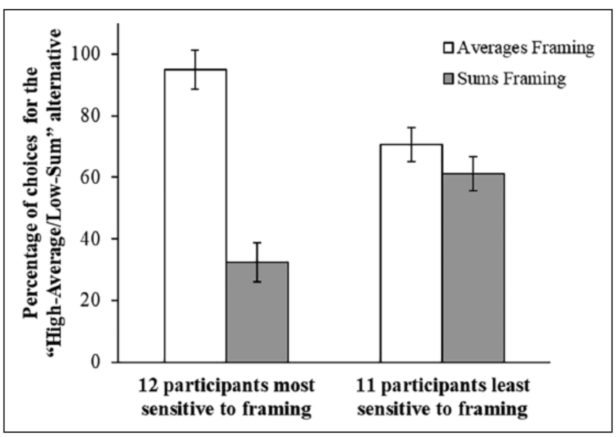

Figure 5. Percentages of choices of the "High-Average/ Low-Sum" alternative ("60-5"), when contrasted with the "Low-Average/High-Sum" alternative ("40-11"), as a function of framing sensitivity, for participants who are more sensitive or less sensitive (median split) to task framing. Error bars correspond to within-participant 95% confidence intervals.

participants were more likely to choose the alternative with the higher sum when the task's rational consideration was to choose according to the sums.

Experiment 2 provided a direct replication to the results of Experiment 1, and introduced a new condition, in which the averages and sums of the alternatives were set in opposition. While this opposition condition did not result in a "full switch"—a preference for the "High-Average, Low-Sum" alternative under the averages framing, and preference for the "High-Sum, Low-Average" alternative under the sums framing (the latter was in the predicted direction but not significantly different from chance; Figure 4, right panel)—there was a strong preference modulation by goal framing (about 40% in preference), indicating that participants did modulate their preference according to the task rationale. The results also indicate that despite the flexible modulation of the preference by the task contingency, which the participants exhibit, there is still a bias in favour of the average (see Figure 5).[5] Of course, the extent of this bias is likely to depend on the parametric modulations of the sum and the average in our Condition (5). In the opposition condition, we constructed alternatives with similar ratios for averages and sums (60/40 vs. 440/300, respectively)—see Condition (5) in Table 1—but a more extreme ratio favouring the sums could eliminate this bias. Moreover, we can also see a similar bias in favour of the average when examining the dominant conditions: the extent of preferring the high-average alternative under the sum framing was greater than the extent of preferring the high-sum alternative under the average framing (3 and 4; see Figures 2 and 3, left and middle panels).

The average bias we obtain here is consistent with the "Jordan-effect" reported by Brusovansky et al. (2017) in a framing (being selected for the "Hall of Fame") that was intentionally ambiguous with regard to what is more rational for that scenario (although we, and about 70%, a group of 29 participants considered summation to be more rational for this case, when queried about this in an analytical way; Experiments 1-3: Discussion, p. 9). In the present study, we obtained a similar averaging bias, with a much less ambiguous framing. For example, participants attended to the average in the summation framing (for alternatives with equal sums; see Figure 4, middle panel), but they did not attend to the sum in the averaging framing (for alternatives with equal average; see Figure 4, left panel). Moreover, this asymmetry now occurs in a situation in which the alternatives are not presented and evaluated one by one, but rather in pairs.

The averaging bias we obtained needs to be qualified, however, as it might depend on the specific characteristics of our task, which presented participants with multiple, rapid sequences and required them to make repeated and intentional preference judgments. This differs from the task used by Betsch and colleagues (2001) who found a summation dominance with a one-trial/subject design, in which the values of the alternatives were framed as distractors, and thus were not subject to intentional encoding (though, like in our case, they were consciously processed, because participants had to read them aloud). It is thus possible that while averaging dominates under the conditions we used, summation dominates under the conditions used by Betsch et al., and the critical factor is the presence of intentional goals. Alternatively, it may be the case that both summation and averaging can be carried out automatically and unintentionally, and the process that dominates depends on unconscious goals. Future studies could test this idea by manipulating goals unconsciously (Bargh, 1990, 1994) in unintentional design tasks.

The idea we suggest here, that averaging could be carried out automatically, may appear to contradict some commonly held assumptions on numerical processing, if one assumes that to compute an average, one needs to first compute a sum and divide by the sample size, a process that makes averaging a more complex operation than summation. In particular, if averaging requires monitoring the sample size in addition to the sum, it might appear unlikely to be carried out unintentionally or with rapid sequences. We should emphasise, however, that this analytic way to compute an average is not the only one, as suggested by the remarkable ability of participants to evaluate summary statistics of visual and even emotional properties from brief displays (Ariely, 2001; Chong & Treisman, 2003, 2005; Haberman & Whitney, 2009; see Khayat & Hochstein, 2018, for a demonstration of automatic processing of the average with an unintentional task). Within the domain of numerical averaging, people have the capacity to estimate averages of multiple sequences presented at a fast rate quite accurately (Malmi & Samson, 1983; see also Krueger, Rothbart, & Sriram, 1989, for a dual-task design). Furthermore, as suggested by Malmi and Samson (1983), it is possible that this computation is mediated by representing the frequencies of the values and extracting the centre of mass of this frequency representation (see Brezis et al., 2016; Brezis et al., 2015, for a population-coding mechanism that accounts for data in explicit averaging tasks with rapid sequences of up to 10/s).

Another possible account for the dominance of averaging in our studies is that there is a hidden (normative) assumption that participants make with regard to the nature of our task. They may assume that the "nature of the slot-machine" is reflected in the average outcome that it produces, but not so much in the number of rewards, which is simply determined by how many times you play it. In other words, the average is beyond one's control and is a relatively stable property of the environment. In contrast, the number of rewards is totally determined by the player, has nothing to do with the machine, and it can be changed from one time to the other.

Beyond the averaging versus summation asymmetry bias, the main aspect of our results is the fact that intuitive preferences are flexibly modulated by task framing. This flexibility, however, shows a significant degree of individual differences. As shown in Figure 4, the modulation of sensitivity by individual differences is both significant and internally consistent. Thus, while some of the participants demonstrate preferences that match well the normative task payoffs, others appear to show only little adaptive modulations (see Figure 5; both groups of people appear to favour averaging to summation). Future research will be needed to better understand the source of these individual differences and their relation to other sources of individual variability, such as fluid intelligence (Duncan, Emslie, Williams, Johnson, & Freer, 1996), need for cognition (Cacioppo, Petty, Feinstein, & Jarvis, 1996), or having a rational versus intuitive thinking style (Epstein, 1994).

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The study was funded by Israeli Science Foundation (1413/17) and Binational (Israel–USA) Science Foundation (2014612) to MU.

Supplementary material

The Supplementary Material is available at: http://supp.apa.org/psycarticles/supplemental/dec0000087/dec0000087_supp.html

Notes

1. See Brezis, Bronfman, and Usher (2015) for presentation rates of up to 10/s.
2. To be more precise, Kahneman and colleagues have proposed a Peak-End heuristic (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993; Redelmeier & Kahneman, 1996). However, this is a type of recency weighted averaging, which makes similar predictions with averaging (but see Brusovansky, Vanunu, & Usher, 2017, for a comparison that favoured averaging). In this study, we do not distinguish between Peak-End and averaging, but rather we see both as variants to be contrasted with summation.
3. The name "Jordan-effect" alludes to the higher public appreciation of Michael Jordan's basketball career after his retirement in 1998, than after his 2 years of comeback a few years later, when he was not nearly as successful as before his retirement, despite additional achievements during these 2 years, which were still impressive compared with other National Basketball Association (NBA) players, but less so than his own before retirement. See also Diener, Wirtz, and Oishi (2001) for a similar "James Dean" effect.
4. In Brusovansky et al. (2017) and in Vanunu, Pachur, and Usher (2018), the grouped condition involved four

alternatives whose sequences were randomly presented. Here, we present sequences of pairs of values and require the subject to select the one they prefer.

5. Looking at individual subjects, we can see two main groups: (1) a group of three to five perfectly rational subjects who almost always select based on task framing, and (2) the rest of subjects that have mixed preferences between the high-average and the high-sum alternative, with a modulation by task framing and a bias in favour of the average (about half of these subjects are at ceiling in their choices of the high-average alternative under the averages framing, but they do not reach ceiling performance in choices of the high-sum alternative under the sums framing).

References

Anderson, N. H. (1981). Integration theory applied to cognitive responses and attitudes. In R. E. Petty, T. M. Ostrom & T. C. Brock (Eds.), *Cognitive responses in persuasion* (pp. 361–397). Hillsdale, NJ: Lawrence Erlbaum.

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*, 157–162.

Bargh, J. A. (1990). Auto-motives: Preconscious determinants of social interaction. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. *2*, pp. 93–130). New York, NY: Guilford Press.

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control in social cognition. In R. S. Wyer Jr & T. K. Srull (Eds.), *Handbook of social cognition: Basic processes; Applications* (pp. 1–40). Hillsdale, NJ: Lawrence Erlbaum.

Betsch, T., Kaufmann, M., Lindow, F., Plessner, H., & Hoffmann, K. (2006). Different principles of information integration in implicit and explicit attitude formation. *European Journal of Social Psychology*, *36*, 887–905.

Betsch, T., Plessner, H., Schwieren, C., & Gütig, R. (2001). I like it but I don't know why: A value-account approach to implicit attitude formation. *Personality and Social Psychology Bulletin*, *27*, 242–253.

Brezis, N., Bronfman, Z. Z., Jacoby, N., Lavidor, M., & Usher, M. (2016). Transcranial direct current stimulation over the parietal cortex improves approximate numerical averaging. *Journal of Cognitive Neuroscience*, *28*, 1700–1713.

Brezis, N., Bronfman, Z. Z., & Usher, M. (2015). Adaptive spontaneous transitions between two mechanisms of numerical averaging. *Scientific Reports*, *5*, 10415. doi:10.1038/srep10415

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.

Brusovansky, M., Vanunu, Y., & Usher, M. (2017). Why we should quit while we're ahead: When do averages matter more than sums? *Decision*, *6*, 1–16. doi:10.1037/dec0000087

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, *119*, 197–253.

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*, 393–404.

Chong, S. C., & Treisman, A. (2005). Statistical processing: Computing the average size in perceptual groups. *Vision Research*, *45*, 891–900.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*, 42–45.

Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of automaticity: Negation as reflective operation. *Journal of Personality and Social Psychology*, *91*, 385–405.

Diener, E., Wirtz, D., & Oishi, S. (2001). End effects of rated quality of life: The James Dean effect. *Psychological Science*, *12*, 124–128.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: The organization of goal-directed behavior. *Cognitive Psychology*, *30*, 257–303.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist*, *49*, 709–724.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*, 692–731.

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 718–734.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*, 1–55.

Kahneman, D. (2011). *Thinking, fast and slow*. London, England: Penguin Books.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, *4*, 401–405.

Khayat, N., & Hochstein, S. (2018). Perceiving set mean and range: Automaticity and precision. *Journal of Vision*, *18*(9), 23.

Krueger, J., Rothbart, M., & Sriram, N. (1989). Category learning and change: Differences in sensitivity to information that enhances or reduces intercategory distinctions. *Journal of Personality and Social Psychology*, *56*, 866–875.

Malmi, R. A., & Samson, D. J. (1983). Intuitive averaging of categorized numerical stimuli. *Journal of Verbal Learning & Verbal Behavior*, *22*, 547–559.

Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, *66*, 3–8.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision making. *Psychological Review*, *108*, 370–392.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1–66.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.

Usher, M., & McClelland, J. L. (2004). Loss aversion and inhibition in dynamical models of multialternative choice. *Psychological Review*, *111*, 757–769.

Vanunu, Y., Pachur, T., & Usher, M. (2018). Constructing preference from sequential samples: The impact of evaluation format on risk attitudes. *Decision*. Advance online publication. doi:10.1037/dec0000098

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics*, *13*, 37–58.