On Bayesian integration in sensorimotor learning: Another look at Kording and Wolpert (2004)

Duffy, Sean and Igan, Deniz and Pinheiro, Marcelo and Smith, John

27 January 2021

On Bayesian integration in sensorimotor learning: Another look at Kording and Wolpert (2004)*

Sean Duffy† Deniz Igan‡ Marcelo Pinheiro§ John Smith¶

January 27, 2021

Keywords: data reanalysis, memory, Bayesian judgments

Kording and Wolpert (2004)[1], hereafter referred to as KW, describe an experiment where subjects engaged in a repeated task entailing movements of their finger. Subjects strove for accuracy in the stochastic environment and, on some trials, received mid-trial and post-trial feedback. KW claims that subjects learned the underlying stochastic distribution from the post-trial feedback of previous trials. KW also claims that subjects regarded mid-trial feedback that had a smaller visual size as more precise and they were therefore more sensitive to such mid-trial feedback. KW concludes that the observations are consistent with optimal Bayesian learning. Indeed, under mild assumptions, it is well-known that Bayesian learners will have posterior beliefs that converge to the true distribution[2–4].

We note that the KW analysis is based on data that had been averaged across important trial-specific details and averaged across trials. Averaging data disregards possibly valuable information and its dangers have been known for some time[5–8]. Notably, the KW analysis does not exclude non-Bayesian explanations. When we analyze the trial-level KW data, we find that subjects were less–not more–sensitive to mid-trial feedback when it had a smaller

*The authors are listed according to the lexicographic convention of economics.

†Rutgers University-Camden, Department of Psychology

‡International Monetary Fund, Research Department

§Guidehouse Consulting

¶Corresponding Author; Rutgers University-Camden, Department of Economics, 311 North 5th Street, Camden, New Jersey, USA 08102; Email: smithj@camden.rutgers.edu; Phone: +1-856-225-6319; Fax: +1-856-225-6602

visual size. Our trial-level analysis also suggests a recency bias, rather than evidence that the subjects learned the stochastic distribution. In other words, we do not find that the observations are consistent with optimal Bayesian learning. In the KW dataset, it seems that evidence for optimal Bayesian learning is a statistical artifact of analyzing averaged data.

Suppose that Alice is involved in a repeated guessing task. At the start of every trial, a number is drawn from a symmetric, unimodal stochastic distribution with a mean of 0. Alice neither knows the distribution nor the draw, yet strives to guess a number closest to the draw in that trial. After each guess, Alice learns the draw in the trial. Therefore, while Alice is not given the distribution, it can be learned by accounting for the previous trials. However, instead of incorporating the information from previous trials, suppose that Alice flips a coin. If the coin lands heads, Alice responds 100 and if the coin lands tails, Alice responds -100. If the analyst averages Alice's responses, they will appear to be correct. However, the analyst would not be justified in claiming that Alice has learned the distribution.

Now, suppose that Bob is in the same environment. Bob responds with the revealed draw from the previous trial. The average of Bob's responses will also approach the mean of the distribution. However, the analyst again would not be justified in claiming that Bob has learned the distribution. Therefore, before the analyst can claim that a subject has learned a distribution, responses must be analyzed across trials and the analyst must rule out that the responses are due to a recency bias.

Finally, suppose that Chris is in the same environment as Alice and Bob. But Chris is additionally given a stochastic mid-trial hint about the draw in that particular trial. Any claims about the effect of this hint should compare the specific information provided in that trial with the response in that trial. Drawing conclusions from averaged mid-trial hints are particularly problematic if the accuracy of these mid-trial hints varies across experimental conditions. Similar to the problems with the inferences involving Alice and Bob, averaging across mid-trial hints can also lead to unwarranted conclusions.

Experimental design

Subjects attempted to hit a target (which had a known and visible position) with a cursor (which had an imperfectly known position). Subjects controlled the cursor with their finger. Subjects could not view the location of their finger because it was situated below a screen. At the start of every trial, the cursor was made invisible to subjects and shifted a distance that was drawn from a normal distribution with a mean of 1 cm to the right and standard deviation of 0.5. This trial-specific outcome is referred to as the *Lateral shift*. See Figure 1 for a graphical characterization of the design.

The target was a vertical distance of 20 cm above the start position. Subjects slid their finger from the start position and strove to hit the target with the cursor.

When their finger was a vertical distance of 10 cm above the start position, subjects possibly obtained mid-trial feedback about the location of the cursor.

The location of these objects could be characterized by their position in the standard two-dimensional (x, y) coordinates. If the target would have coordinates $(0, 0)$, the start position of the finger would be $(0, -20)$, and the mean initial position of the cursor would be $(1, -20)$. Given that the Lateral shift has a mean of 1, the final finger location that would account for this shift would be $(-1, 0)$. We refer to the x-coordinate of the final finger location, as the *Final finger position*. Likewise, we refer to the x-coordinate of the final cursor location, as the *Final cursor position*.

There were 10 subjects who each performed 2000 trials. There were 4 within-subject conditions that were randomly assigned across trials.

In Condition 1, subjects received mid-trial feedback, in the form of a dot, suggesting the location of the cursor. Subjects in Condition 1 also obtained post-trial feedback indicating the final location of the cursor. Condition 1 was programmed to occur in 50% of trials. In Conditions 2 and 3, subjects received mid-trial feedback in the form of a cloud of 25 dots. The variance in the spatial locations of the dots was larger in Condition 3 than in Condition 2. Neither Conditions 2 nor 3 provided post-trial feedback about the final cursor location. Conditions 2 and 3 were both programmed to occur in 16.7% of trials. In Condition 4,

the subjects neither received mid-trial feedback nor post-trial feedback. Condition 4 was programmed to occur in 16.7% of trials.

Finally, note that KW analyze only the last 1000 of the 2000 trials.

Main results of KW

Proponents of Bayesian models argue that subjects optimally incorporate new information by using Bayes' Rule to account for their imperfect information, imperfect memory, and other possible imperfections.

As the result of the visual sizes of the mid-trial feedback, KW claims that subjects regarded the mid-trial feedback of Condition 1 as more informative than in Condition 2, and the mid-trial feedback of Condition 2 as more informative than in Condition 3. As evidence of this, KW claims that responses were more sensitive to mid-trial feedback in Condition 1 than Condition 2 and Condition 2 more than Condition 3. See Figures 2a and 2b in KW.

KW also claims that subjects learned the average of the distribution by noting the mean responses in Condition 4. See page 244 in KW.

Effect of mid-trial feedback

KW claims that evidence of optimal Bayesian learning comes from the observation that subjects were more sensitive to the mid-trial feedback when it occupied a smaller visual area. Below, we reproduce a table version of Figures 2a and 2b in KW. The variable *Deviation*, defined to be the difference between the Target position and the Final cursor position, is the dependent variable. The Lateral shift is the independent variable. These regressions are summarized in Table 1.

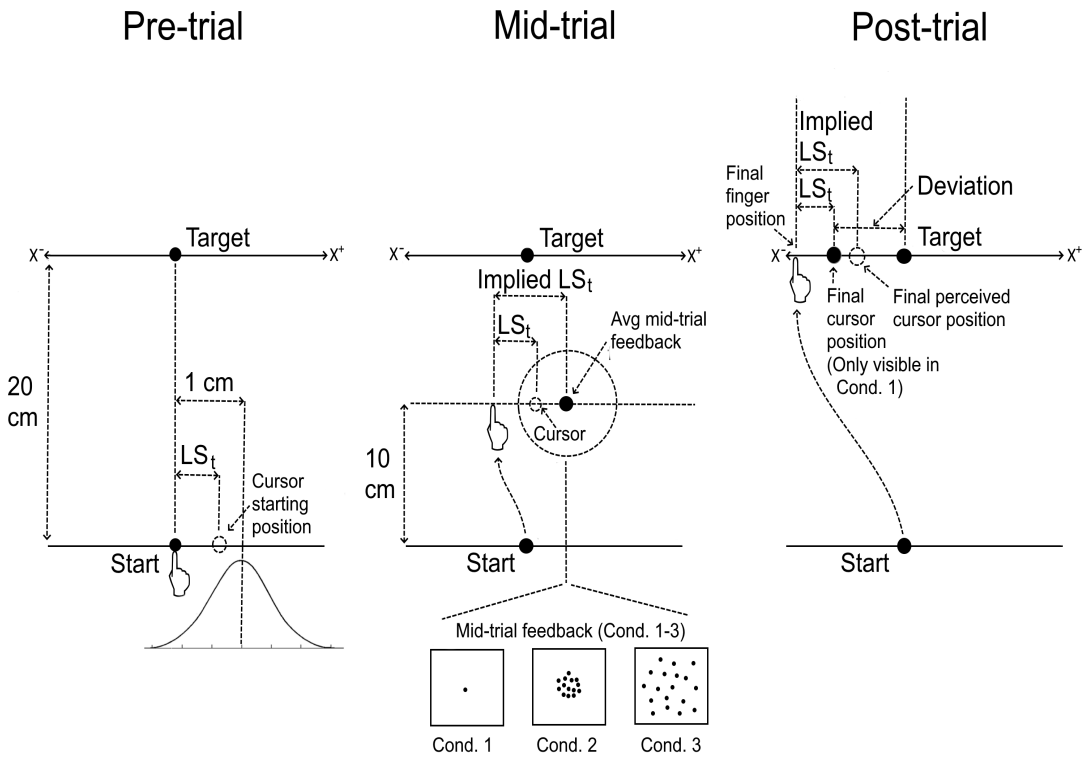

Figure 1: The Pre-trial panel shows the finger at the start position and the invisible cursor is shifted with a mean of 1 cm to the right. LS_t is the Lateral shift in trial t, which is fixed throughout the trial. The Mid-trial panel shows the finger and the invisible cursor 10 vertical cm from the start position. In Conditions $1-3$, the subject received feedback about the location of the cursor. Mid-trial feedback in Condition 1 was a dot, in Condition 2 was a cloud of dots, and in Condition 3 was a larger cloud of dots. Condition 4 trials did not receive mid-trial feedback. The difference between the Finger position and the Average mid-trial feedback is the Implied lateral shift for trial t. The Post-trial panel shows the finger and the cursor 20 vertical cm from the start position. This Final cursor position was visible only in Condition 1. We also depict the Final perceived cursor position, which would have been the final location of the cursor if the Average mid-trial feedback coincided with the actual cursor. Finally, we denote the Deviation as the difference between the Target position and the Final cursor position.

Table 1: Regressions of Deviation for last 1000 trials

	Cond. 1	Cond. 2	Cond. 3	Cond. 4
Intercept	-0.3550^{***}	-0.5739^{***}	-0.6843^{***}	-0.9931^{***}
	(0.01607)	(0.03065)	(0.03293)	(0.03624)
Lateral shift	0.3297^{***}	0.5777^{***}	0.6902^{***}	1.0199^{***}
	(0.01451)	(0.02808)	(0.02938)	(0.03240)
AIC	7172.9	2732.2	2877.2	3393.5
Observations	4916	1627	1615	1673

We provide the coefficient estimates and, in parentheses, the standard errors. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$, for two-sided p-values.

The results of Table 1 resemble those of Figures 2a and 2b in KW. These results are also robust to including fixed-effects for intercepts and slopes. The results are further robust to including all 2000 trials, not simply the last 1000 (Supplemental analysis Table $A1$).

KW interprets this as evidence that subjects were using the apparent precision of mid-trial feedback as inversely related to the visual size of the cloud and that they account for these differences in their judgments. However, before such an inference can be made, it is crucial to understand the relationship between the mid-trial feedback and the position of the (invisible) cursor.

KW suggests that the mid-trial feedback accurately revealed the cursor position. For example, Figure 1 in KW shows the mid-trial feedback at the same location as the cursor. KW also write, "Midway through the movement (10 cm), feedback of the cursor centered at the displaced finger position was flashed for 100 ms" (page 246).

Here we investigate the accuracy of the mid-trial feedback. Since the mid-trial feedback in Conditions 2 and 3 took the form of a cloud of dots, we summarize the feedback by taking the average of the x-coordinates of the dots in the cloud. We run regressions with the *Average mid-trial feedback* as the dependent variable. We use the *Mid-trial cursor position*, defined to be Mid-trial finger position plus the Lateral shift, as the independent variable. These regressions are summarized in Table 2.

Table 2: Regressions of Average mid-trial feedback for last 1000 trials

	Condition 1	Condition 2	Condition 3
Intercept	-0.0096^{***}	-0.926^{***}	0.046^{***}
	(0.0007)	(0.0005)	(0.0005)
Mid-trial cursor position	1.0023^{***}	1.0062^{***}	1.0042^{***}
	(0.0008)	(0.0006)	(0.0006)
AIC	−16348.2	−8026.0	−8212.6
Observations	4916	1627	1615

We provide the coefficient estimates and, in parentheses, the standard errors. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$, for two-sided p-values.

Accurately reflecting the cursor position would lead to an Intercept coefficient estimate of 0 and a Mid-trial cursor position coefficient estimate of 1. We find evidence that the mid-trial feedback was not accurately reflecting the location of the cursor. The Intercept coefficient estimates are significantly different than 0. The Mid-trial cursor position coefficient estimates are significantly different than 1 for Condition 1 ($t(4914) = 2.86$, two-sided $p = 0.004$), Condition 2 ($t(1625) = 10.51$, two-sided $p < 0.001$), and Condition 3 ($t(1613) = 7.39$, two-sided $p < 0.001$). Additionally we find that the relationship between mid-trial feedback and the mid-trial cursor position varied across conditions. These results are also robust to including all 2000 trials (Supplemental analysis Table *A*2).

We further note that subjects in Conditions 2 and 3 were not able to learn this relationship since they did not receive post-trial feedback. Therefore, the analyses summarized in our Table 1 and Figures 2a,b in KW are not sufficient to infer a negative relationship between sensitivity to mid-trial feedback and the visual size of the mid-trial feedback.

In order to investigate such a relationship, we compare the location of the mid-trial feedback in that trial with the perceived location of the final cursor position, as implied by the mid-trial feedback in that trial. The mid-trial finger position and the mid-trial feedback generates information about the apparent Lateral shift in that trial. Since this is the only trial-specific information about the Lateral shift, we define the *Final perceived cursor position* to be the Final cursor position if the Mid-trial feedback reflected the true Lateral shift. We summarize the regressions with Final perceived cursor position as dependent variable and

Average mid-trial feedback as independent variable in Table 3.

Table 3: Regressions of Final perceived cursor position for last 1000 trials

	Condition 1	Condition 2	Condition 3
Intercept	-0.064^{***}	-0.572^{***}	-0.022
	(0.007)	(0.017)	(0.014)
Average mid-trial feedback	0.320^{***}	0.438^{***}	0.478^{***}
	(0.008)	(0.015)	(0.017)
AIC	6424.1	2418.3	2710.0
Observations	4916	1627	1615

We provide the coefficient estimates and, in parentheses, the standard errors. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$, for two-sided p-values.

These coefficient estimates suggest that if Average mid-trial feedback were to change by 1 *cm* then the changes in the Final perceived cursor position would be 0.320 *cm* for Condition 1, 0.438 *cm* for Condition 2, and 0.478 *cm* for Condition 3. In other words, there was a positive relationship between the Average mid-trial feedback coefficients and the visual size of the mid-trial feedback. Therefore, we find evidence of a positive–not a negative–relationship between sensitivity to mid-trial feedback and the visual size of the mid-trial feedback cloud. The Supplemental analysis also contains specifications that use the data from every trial (Table *A*3), those with the dependent variable of Final cursor position (Tables *A*4 and *A*5), and those with the dependent variable of Final finger position (Tables *A*6 and *A*7). Our results, that trials in Conditions 2 and 3 were more sensitive to mid-trial feedback than trials in Condition 1, are not changed.

On the evidence of learning in Condition 4

On page 244, KW writes, "...when feedback was withheld (σ_∞), subjects pointed 0.97 ± 0.06 *cm* (mean $\pm$ s.e.m. across subjects) to the left of the target showing that they had learned the average shift of 1 *cm* experienced over the ensemble of trials." However, based on a two-sided single-sample t-test ($t(1672) = 1.65$, $p = 0.0999$) and a two-sided signed-rank test ($S = 38484$, $p = 0.0515$), we actually find weak evidence that the Final finger position was different than

1 *cm*. To further explore this matter, we characterize the Final finger position for every trial, the last 1000 trials, and the last 500 trials in Table 4.

Table 4: Mean Final finger position by condition

	Cond. 1	Cond. 2	Cond. 3	Cond. 4	Eq. of Var (4 vs. others)
Every trial	−0.99400	−0.96511	−0.94688	−0.93434	$F(3230, 15808) = 1.24$
	(0.63421)	(0.64263)	(0.65414)	(0.71294)	$p < 0.001$
Last 1000	−1.01964	−0.98527	−0.99555	−0.97324	$F(1672, 8157) = 1.22$
	(0.60005)	(0.59564)	(0.60725)	(0.66476)	$p < 0.001$
Last 500	−1.00432	−0.96890	−0.97591	−0.97539	$F(836, 4123) = 1.34$
	(0.58522)	(0.57000)	(0.58461)	(0.67316)	$p < 0.001$

We provide the means and, in parentheses, the standard deviations. We also perform the equality of variance folded F-test on Condition 4 as different from the other trials. We report the F-statistic and the two-sided p-values.

We find that the mean final finger position in Condition 4 is farther from 1 *cm* than the other conditions, restricted to every trial and the last 1000 trials. We also note that the standard deviation of Condition 4 appears to be larger than those in the other conditions. In order to test this conjecture, we perform an Equality of Variance Test that trials in Condition 4 have a different variance than trials in the other conditions. Despite that the Condition 4 trials do not receive mid-trial feedback, there appears to be more noise in Condition 4 trials than in the other trials.

Because standard deviation is a global measure of noise, we consider a local measure, which compares adjacent responses within the same condition. We take the absolute value of the difference between the Final finger position and the previous Final finger position, within in the same condition. We refer to this variable as *Absolute previous within condition deviation*. In Table 5, we summarize the Absolute previous within condition deviation restricted by condition.

Table 5: Absolute previous within condition deviation

	Cond. 1	Cond. 2	Cond. 3	Cond. 4	Wilcoxon (4 vs. others)
Every trial	0.63910	0.61739	0.61214	0.66455	$z(3221, 15779) = 2.6429$
	(0.51630)	(0.50078)	(0.49744)	(0.56052)	$p = 0.008$
Last 1000	0.61328	0.58073	0.58200	0.62392	$z(1673, 8158) = 1.5638$
	(0.48913)	(0.45209)	(0.47055)	(0.52581)	$p = 0.118$
Last 500	0.61125	0.55503	0.58125	0.62673	$z(837, 4124) = 0.9337$
	(0.47270)	(0.43754)	(0.46075)	(0.57290)	$p = 0.351$

We provide the means and, in parentheses, the standard deviations. We also perform a Wilcoxon two-sample test on the difference between Condition 4 and the other conditions. We report the z-statistic and the two-sided p-values.

Our results suggest that trials in Condition 4 were either noisier than the other trials or they were not significantly less noisy. This appears to contradict the claim that mid-trial feedback is affecting the responses, as the condition without mid-trial feedback did not have significantly less noise than the conditions with mid-trial feedback.

In order to learn whether the trajectory of the noise across trials in Condition 4 was different from those in the other trials, we conduct regressions with Absolute previous within condition deviation as the dependent variable. We employ the *Trial count* variable, which is the ordinal trial number within the set of considered trials. Condition 4 is a dummy variable that obtains a value of 1 if the trial was in Condition 4 and a 0 otherwise. We also include the Trial count interaction with Condition 4. This is summarized in Table 6.

Table 6: Regressions of Absolute previous within condition deviation

	Every Trial	Last 1000	Last 500
Intercept	0.698^{***}	0.628^{***}	0.615^{***}
	(0.008)	(0.011)	(0.015)
Trial count	-0.00007^{***}	-0.00005^{**}	-0.00008
	(0.000007)	(0.000019)	(0.000052)
Condition 4	$0.0342^{\dagger}$	-0.034	-0.0445
	(0.021)	(0.026)	(0.0366)
Trial Count*Condition 4	0.000001	0.00012^{*}	0.00031^{*}
	(0.000018)	(0.00005)	(0.00013)
AIC	28926.6	13790.7	6922.7
Observations	19000	9831	4961

We provide the coefficient estimates and, in parentheses, the standard errors.

AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$, for two-sided p-values.

We find that, while the noise in the conditions other than Condition 4 was diminishing across trials, we do not find evidence of this in Condition 4 trials. In fact, it appears as if there are specifications where the noise in Condition 4 was increasing across trials.

Here we seek to better understand the noise in Condition 4. We run regressions with the Final finger position as the dependent variable, as Final finger position is an estimate of the Lateral shift. We include an independent variable that is the running mean of the Lateral shifts previously observed in post-trial feedback. We refer to this variable as *RM prev. viewed lateral shift.* We also include an independent variable that is the previously viewed Lateral shift in post-trial feedback. We refer to this variable as *Prev. viewed lateral shift.* This analysis is summarized in Table 7.

Table 7: Regressions of Final finger position for last 1000 trials in Condition 4

	(1)	(2)	(3)
Intercept	1.3035	0.6854	1.0180
	(0.8923)	(0.03567)	(0.8722)
RM prev. viewed lateral shift	−0.3329	−	−0.3352
	(0.8992)		(0.8784)
Prev. viewed lateral shift	−	0.2924***	0.2924***
		(0.03244)	(0.03245)
AIC	3387.1	3314.6	3312.8

We provide the coefficient estimates and, in parentheses, the standard errors. Each regression has 1673 observations. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$, for two-sided p-values.

We find that there is a bias toward the previously viewed Lateral shift, but there is not a bias toward the running mean of the previously viewed Lateral shifts. Therefore, what can be viewed as learning based on the analysis of averaged data, appears to be a recency bias. The Supplemental analysis also contains specifications that use the data from every trial (Table *A*8), and our results are not changed. There are also other examples where a recency bias accounts for observations that were previously claimed to be evidence of Bayesian learning[9−10].

In summary, we do not find evidence that the subjects learned the mean of Lateral shift. We find that responses in Condition 4 (without mid-trial feedback) were not less noisy than responses in the other conditions. Moreover, there was not a decrease in the noise across trials that would be consistent with learning. Further, subjects appear to have exhibited a recency bias rather than learning the mean of the distribution. In conclusion, we do not find evidence of optimal Bayesian learning in the responses of Condition 4.

Conclusions

We have reanalyzed the data from KW, and we do not find evidence in support of claims that subjects were behaving in a manner that was consistent with optimal Bayesian learning. When we compare the mid-trial feedback in a particular trial and the outcome in that trial, we find that subjects were not more sensitive to mid-trial feedback because it had a smaller visual size. We also find that the responses exhibited a recency bias rather than evidence of learning.

Our effort is another instance of the reanalysis of a dataset that was previously regarded as providing evidence in favor of Bayesian models, yet was previously scrutinized only by techniques that analyzed averaged data[9−14]. As we do here, those other efforts also found that evidence in favor of Bayesian models is a statistical artifact of analyzing averaged data. When data are analyzed by accounting for relevant trial-specific details, the evidence in favor of optimal Bayesian learning disappears[9−10].

Our findings support the view that, before strong conclusions are made, statistical analyses should rule out alternate explanations. Further, researchers should exercise caution about analyzing data that have been averaged across meaningful experimental details.

References

1. Körding, K.P. & Wolpert, D. Bayesian integration in sensorimotor learning. *Nature* **427**, 244–247 (2004).

2. Savage, L.J. The Foundations of Statistics. Wiley, New York (1954).

3. Blackwell, D. & Dubins, L. Merging of opinions with increasing information. *Annals Math Stat* **33**, 882–886 (1962).

4. Edwards, W., Lindman, H. & Savage, L.J. Bayesian statistical inference for psychological research. *Psychol Rev* **70**, 193–242 (1963).

5. Sidman, M. A note on functional relations obtained from group data. *Psychol Bull* **49**, 263–269 (1952).

6. Hayes, K. J. The backward curve: A method for the study of learning. *Psychol Rev* **60**, 269–275 (1953).

7. Estes, W. K. The problem of inference from curves based on group data. *Psychol Bull* **53**, 134–140 (1956).

8. Siegler, R. S. The perils of averaging data over strategies: An example from children's addition. *J Exp Psychol Gen* **116**, 250–264 (1987).

9. Duffy, S. & Smith, J. Category effects on stimulus estimation: Shifting and skewed frequency distributions—A reexamination. *Psychon Bull Rev* **25**, 1740–1750 (2018).

10. Duffy, S. & Smith, J. On the Category Adjustment Model: Another look at Huttenlocher, Hedges, and Vevea (2000). *Mind Soc* **19**, 163–193 (2020a).

11. Huttenlocher, J., Hedges, L.V. & Vevea, J.L. Why do categories affect stimulus judgment? *J Exp Psychol Gen* **129**, 220–241 (2000).

12. Duffy, S., Huttenlocher, J., Hedges, L.V. & Crawford, L.E. Category effects on stimulus estimation: Shifting and skewed frequency distributions. *Psychon Bull Rev* **17**, 224–230 (2010).

13. Crawford, L.E. Reply to Duffy and Smith's (2018) reexamination. *Psychon Bull Rev* **26**, 693–698 (2019).

14. Duffy, S. & Smith, J. Omitted-variable bias and other matters in the defense of the category adjustment model: A comment on Crawford (2019). *J Behav Exp Econ* **85**, 101501 (2020b).

Methods

If the subject failed to traverse the vertical 20 cm in the allotted time ($2s$) then the trial is referred to as Condition 5. KW does not mention the possibility of Condition 6. However, the dataset contains trials denoted as Condition 6, which similar to Condition 5, appear to be errors or incomplete trials. It appears that trials in Condition 5 and Condition 6 were excluded from the analysis of KW. In trials that were subsequently relabeled as either Condition 5 or Condition 6, it is not possible to infer the initially designated condition ($1-4$).

The original dataset had $20,100$ trials. Upon inspection, every trial number that was a multiple of 201, was a repeat of the previous trial. We therefore deleted these identical trials so that the number of trials matches that reported by KW.

The distribution of conditions were: 9523 in Condition 1 (47.6%), 3146 in Condition 2 (15.7%), 3140 in Condition 3 (15.7%), 3231 in Condition 4 (16.2%), 610 in Condition 5 (3.05%), and 350 in Condition 6 (1.75%).

KW only analyze the final 1000 trials per subject. The distribution of conditions within the last 1000 trials were: 4916 in Condition 1 (49.2%), 1627 in Condition 2 (16.3%), 1615 in Condition 3 (16.2%), 1673 in Condition 4 (16.7%), 106 in Condition 5 (1.06%), and 63 in Condition 6 (0.63%).

Trials were constrained to be 2 seconds or less. The original dataset provided a characterization of the finger trajectory within every trial. These 2 seconds are partitioned into 200 equal segments whereby a measurement of the finger position is specified at the end of every 0.01 seconds. We generate our variables from the original dataset, which was downloaded from https://crcns.org/data-sets/movements/dream in June 2018. KW refers to Conditions $1-4$, respectively, as σ_0, σ_M, σ_L, and σ_∞. We refer to the conditions as they appear in the dataset.

Data availability

Our dataset is available at https://osf.io/ag2qm.

Acknowledgements We thank Roberto Barbera and Joseph Martin for comments. We thank Huy Nguyen for assistance with the data. John Smith thanks Biblioteca de Catalunya. This research was supported by Rutgers University Research Council Grant #FPRC0000016. The views expressed in this paper are those of the authors and do not necessarily represent those of the International Monetary Fund or its policy.

Author contributions D.I., M.P., and J.S. produced the dataset. J.S. analyzed the data and wrote the paper. S.D., D.I., and M.P. contributed to the conceptualization, the exposition, the statistics, and the interpretation of the results. S.D. and J.S. produced the figure.

Competing interests The authors declare no competing interests.

Supplemental analysis

We conduct an analysis similar to Table 1 but we include every trial.

Table *A*1: Regressions of Deviation for every trial

	Cond. 1	Cond. 2	Cond. 3	Cond. 4
Intercept	-0.3997^{***}	-0.5898^{***}	-0.6344^{***}	-0.9363^{***}
	(0.0130)	(0.0244)	(0.0254)	(0.0282)
Lateral shift	0.4015^{***}	0.6212^{***}	0.6889^{***}	1.0020^{***}
	(0.0117)	(0.0220)	(0.0227)	(0.0254)
AIC	16051.3	5875.0	6076.4	6996.1
Observations	9523	3146	3140	3231

We provide the coefficient estimates and, in parentheses, the standard errors. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$, for two-sided p-values.

These results are qualitatively similar to those in Table 1. We conduct an analysis similar to Table 2 but we include every trial.

Table $A2$: Regressions of Average mid-trial feedback for every trial

	Condition 1	Condition 2	Condition 3
Intercept	-0.01072^{***}	-0.9273^{***}	0.04475^{***}
	(0.000378)	(0.000395)	(0.000438)
Mid-trial cursor position	1.0031^{***}	1.0066^{***}	1.0041^{***}
	(0.0005)	(0.0005)	(0.0005)
AIC	-35854.2	-15039.6	-14384.1
Observations	9523	3146	3140

We provide the coefficient estimates and, in parentheses, the standard errors. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$, for two-sided p-values.

These results are qualitatively similar to those in Table 2. We conduct an analysis similar to Table 3 but we include every trial.

Table $A3$: Regressions of Final perceived cursor position for every trial

	Condition 1	Condition 2	Condition 3
Intercept	-0.035^{***}	-0.483^{***}	0.034^{***}
	(0.005)	(0.013)	(0.011)
Average mid-trial feedback	0.357^{***}	0.496^{***}	0.501^{***}
	(0.007)	(0.011)	(0.012)
AIC	14757.1	5147.5	5570.1
Observations	9523	3146	3140

We provide the coefficient estimates and, in parentheses, the standard errors. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$, for two-sided p-values.

The results in Table 3 are not changed, as Condition 1 trials were less sensitive to mid-trial feedback than were trials in Conditions 2 or 3. We conduct of these results by employing the Final cursor position as dependent variable.

Table $A4$: Regressions of Final cursor position for last 1000 trials

	Condition 1	Condition 2	Condition 3
Intercept	-0.05431^{***}	0.3481^{***}	-0.06741
	(0.006593)	(0.01738)	(0.01408)
Average mid-trial feedback	0.3148^{***}	0.4316^{***}	0.4736^{***}
	(0.008034)	(0.01450)	(0.01665)
AIC	6329.1	2402.5	2697.0
Observations	4916	1627	1615

We provide the coefficient estimates and, in parentheses, the standard errors. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$, for two-sided p-values.

We find that subjects in Conditions 2 and 3 were more sensitive to mid-trial feedback than in Condition 1. We conduct an analysis similar to Table *A*4 but we include every trial.

Table *A*5: Regressions of Final cursor position for every trial

	Condition 1	Condition 2	Condition 3
Intercept	-0.02418^{***}	0.4377^{***}	-0.01024
	(0.005342)	(0.01351)	(0.01052)
Average mid-trial feedback	0.3517^{***}	0.4888^{***}	0.4962^{***}
	(0.006445)	(0.01119)	(0.01202)
AIC	14574.0	5098.5	5526.9
Observations	9523	3146	3140

We provide the coefficient estimates and, in parentheses, the standard errors. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$, for two-sided p-values.

Our results are not changed. We conduct another robustness check with Final finger position as dependent variable.

Table *A*6: Regressions of Final finger position for last 1000 trials

	Condition 1	Condition 2	Condition 3
Intercept	-1.0153^{***}	-0.9188^{***}	-1.0171^{***}
	(0.008582)	(0.02039)	(0.01514)
Average mid-trial feedback	-0.05231^{***}	0.07983^{***}	0.1337^{***}
	(0.01046)	(0.01702)	(0.01790)
AIC	8920.4	2923.3	2930.9
Observations	4916	1627	1615

We provide the coefficient estimates and, in parentheses, the standard errors. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$, for two-sided p-values.

Here we find that trials in Condition 2 and 3 have a positive relationship between mid-trial feedback and final finger position. However, we find a negative and significant coefficient for Condition 1. This suggests that mid-trial feedback in Condition 1 negatively affects the final

finger position in the last 1000 trials. We conduct an analysis similar to Table $A6$ but we include every trial.

Table $A7$: Regressions of Final finger position for every trial

	Condition 1	Condition 2	Condition 3
Intercept	-0.9939^{***}	-0.8421^{***}	-0.9690^{***}
	(0.006520)	(0.01569)	(0.01155)
Average Mid-trial feedback	-0.00212	0.1459^{***}	0.1618^{***}
	(0.007866)	(0.01299)	(0.01320)
AIC	18369.0	6036.9	6113.7
Observations	9523	3146	3140

We provide the coefficient estimates and, in parentheses, the standard errors. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$, for two-sided p-values.

We find evidence that subjects were sensitive to mid-trial feedback in Conditions 2 and 3, but not in Condition 1. Similar to Table $A6$, we find a negative coefficient estimate for Condition 1, but it is not significantly different than 0.

We conduct an analysis similar to Table 7 but we include every trial.

Table $A8$: Regressions of Final finger position for every trial in Condition 4

	(1)	(2)	(3)
Intercept	0.5833^{*}	0.6824^{***}	0.5460^{*}
	(0.2344)	(0.0279)	(0.2309)
RM prev. viewed lateral shift	0.3542	–	0.1386
	(0.2354)		(0.2328)
Prev. viewed lateral shift	–	0.2560^{***}	0.2546^{***}
		(0.02525)	(0.02536)
AIC	6973.3	6878.8	6879.5

We provide the coefficient estimates and, in parentheses, the standard errors. Each regression has 3227 observations. AIC refers to the Akaike information criterion. *** denotes $p < 0.001$, ** denotes $p < 0.01$, * denotes $p < 0.05$, and † denotes $p < 0.1$, for two-sided p-values.

The results from Table 7 are not changed.