Omitted-variable bias and other matters in the defense of the category adjustment model: A reply to Crawford (2019)*

Sean Duffy

Rutgers University-Camden

John Smith

Rutgers University-Camden

Address correspondence to:

John Smith
Department of Economics
Rutgers University-Camden
311 N. 5th Street
Camden, NJ
08102 USA
smithj@camden.rutgers.edu

July 8, 2019

*We thank Roberto Barbera, Deniz Igan, Marcelo Pinheiro, and Adam Sanjurjo for helpful comments. This project was supported by Rutgers University Research Council Grant # FPRC0000016. John Smith thanks Biblioteca de Catalunya.

Abstract

The datasets from Duffy, Huttenlocher, Hedges, and Crawford (2010) [*Psychonomic Bulletin & Review*, 17(2), 224-230] were reanalyzed by Duffy and Smith (2018) [*Psychonomic Bulletin & Review*, 25(5), 1740-1750]. Duffy and Smith (2018) conclude that the datasets are not consistent with the category adjustment model (CAM). Crawford (2019) [*Psychonomic Bulletin & Review*, 26(2), 693-698] offered a reply to Duffy and Smith (2018) that is based on three main points. Crawford proposes regressions that are, in part, based on a "deviation" analysis. Crawford offers a different simulation of data and claims that the techniques employed by Duffy and Smith (2018) are not sufficiently sensitive to detect a specific relationship that is claimed to be consistent with CAM. Crawford also appeals to a figure showing that the responses appear to be biased toward the overall running mean, and presumably not toward recently viewed lines. We show that Crawford's analysis suffers from an omitted-variable bias. Once this bias is corrected, the evidence in support of CAM disappears. When we produce a simulated dataset that is consistent with the specification suggested by Crawford, the techniques of Duffy and Smith (2018) correctly detect the true relationship. Despite the assertion otherwise, the simulated dataset that was analyzed by Crawford is not publicly available. Since the analysis of Crawford (2019) is incorrect, it remains our view that the datasets from Duffy, Huttenlocher, Hedges, and Crawford (2010) do not appear to be consistent with CAM or any Bayesian model of judgment.

Keywords: judgment, omitted-variable bias, category adjustment model, central tendency bias, recency effects, Bayesian judgments

Duffy, Huttenlocher, Hedges, and Crawford (2010), hereafter DHHC, report on two experiments where participants estimate the lengths of lines. These experiments were designed to test the *category adjustment model* (CAM, Huttenlocher, Hedges, & Vevea, 2000), a Bayesian model of judgments. DHHC report that their analysis provides evidence consistent with CAM: that there is a bias toward the running mean of the lengths of lines and not toward the lengths of recently viewed lines.

Duffy and Smith (2018), hereafter DS, reanalyze the DHHC datasets. DS find errors in the analysis of DHHC. Also, by employing more sophisticated techniques than DHHC, DS obtain very different results. Specifically, DS find significant recency effects and several specifications where the running mean is not significantly related to judgment. Further, since it is well-known that Bayesians have beliefs that converge to the truth (Savage, 1954; Blackwell & Dubins, 1962; Edwards, Lindman, & Savage, 1963), DS conduct multiple specifications of different measures of learning across trials. DS does not find evidence of learning across trials in the DHHC datasets. Finally, DS produce a simulated dataset that is consistent with a key feature of CAM and their methods correctly identify this simulated data as consistent with CAM. DS conclude that the DHHC datasets are not consistent with CAM. Further, since there is no evidence of learning, it is not clear how the DHHC datasets are consistent with any model of Bayesian judgment.

This journal recently published a reply to DS. In it, Crawford (2019) offers a defense of DHHC, in particular, and CAM, in general. The reply is based on three main points. Crawford proposes regressions that are, in part, based on a “deviation” analysis. This deviation analysis is claimed to be significantly different than and superior to those in DS. Crawford offers a different simulation of data and claims that the techniques employed in DS are not sufficiently sensitive to detect this specific relationship that is claimed to be consistent with CAM. Crawford also appeals to a figure showing that the responses appear to be biased toward the overall running mean, and presumably not toward recently viewed lines. Below, we address each of these three points.

Background on DHHC datasets

DHHC report on two experiments where participants were directed to judge the length of lines with 19 possible stimulus lengths, ranging from 80 to 368 pixels, in increments of 16 pixels. We refer to the line that is to be estimated as the *target line*.[1]

Participants were presented with the target line then the target line disappeared. Subsequently, an initial adjustable line appeared. Participants would manipulate the length of this adjustable line until they judged its length to be that of the target line. We refer to this response as the *response line*. We refer to the first experiment as DHHC1 and the second as DHHC2. Both datasets are available on https://osf.io/gnefa/.

The reader is referred to DHHC for further details about the design or execution of the experiments. The reader is referred to DS for further details about the datasets or matters related to the analysis.

We define the *Running mean* variable to be the average of the lengths of lines from the previous trials. As DS investigates a recency bias, we refer to the average of the previous n targets as *Last(n)*. Since Crawford focuses on the previous three targets, we also refer to the average of the previous 3 targets as *Last(3)*.

Omitted-variable bias

Suppose that a true relationship between variables Y, X_1, and X_2 is as follows:

[1] We refer to the actual length of the line as the *Target* rather than the *Stimulus* because there are many other possibly relevant stimuli: the initial length of the adjustable line, the length of the previously seen lines, the width of line, the color of line, etc.

(1) $$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon.$$

However, also suppose that we run a regression of the following form:

(2) $$Y = \beta'_0 + \beta'_1 X_1 + \varepsilon.$$

We say that X_2 is an *omitted variable* because it should be included in a regression involving Y. This omitted variable can cause an *omitted-variable bias* if the following two conditions hold: (i) X_1 and X_2 are correlated and (ii) X_2 is a determinant of Y, in that β_2 in expression (1) is not equal to zero.

If these two conditions hold then the estimates of β'_1 in expression (2) will be biased and will not accurately characterize how Y changes in response to a change in X_1. In the case of an omitted-variable bias, the estimates of β'_1 will incorrectly attribute to X_1 the effect on Y that is actually due to the X_2 variable.[2]

Deviation analysis of Crawford (2019)

The primary specification of DS is:[3]

(3) $$\text{Response} = \beta_0 + \beta_1 \text{Target} + \beta_2 \text{Running Mean} + \beta_3 \text{Last}(n) + \varepsilon.$$

On page 695, Crawford writes, "An alternative and more intuitive approach is to model CAM's prediction that bias in responses depends on the deviation of a stimulus from the prior." Crawford thus proposes a different specification. Rather than employ a dependent variable of *Response*, Crawford proposes *Response Bias*, which is defined to be the difference between the

[2] For more on the omitted-variable bias, see Hill, Griffith, and Judge (2001), or nearly any econometrics textbook.
[3] DS consider n equal to 1, 3, 5, 10, 15, and 20.

Response and the Target in that trial.[4] In other words, Response Bias = Response – Target. Crawford also proposes independent variables that account for the deviation of both the Running Mean and the Previous Targets from the Target in that trial. Crawford defines:

$$\text{Run Mean Deviation} = \text{Running Mean} - \text{Target}.$$

The Last(3) Deviation variable is presumably defined in a similar fashion: Last(3) Deviation = Last(3) – Target. Crawford defines these deviation regressions to be:[5]

$$\text{Response Bias} = \beta_0 + \beta_1\text{Run Mean Deviation} + \beta_3\text{Last(3) Deviation} + \varepsilon. \tag{4}$$

On page 695, Crawford writes, “This approach, hereafter called ‘deviations analysis,’ produces different results from the DS analysis. To see why…”

The serious and fundamental problem with the specification in expression (4) is that it suffers from an omitted-variable bias. The case for including Target in the analysis is that it has been known to be significantly related to Response Bias for quite some time (Hollingworth, 1910). Crawford (2019) makes 4 explicit references to the *central tendency bias*, which describes the negative relationship between Response Bias and Target. On page 693, Crawford twice describes the central tendency bias as “well-established.” Figure 1 in Crawford (2019) clearly shows a relationship between Response Bias and Target.

In applied work, there can be disagreement about whether a variable should be included in a regression. However, there can be little doubt that Target should be included in regressions involving either Response Bias or Response.

---

[4] What we refer to as *Response Bias*, Crawford simply refers to as *Bias*. Since DS examine several measures of bias, we employ the terminology of DS.

[5] This is expression (6) in Crawford (2019).

Below we show that Target is a determinant of Response Bias. We also find that Target is correlated with Running Mean Deviation ($r = -0.923$, $p < .001$), Last(3) Deviation ($r = -0.764$, $p < .001$), Running Mean ($r = 0.159$, $p < .001$) and Last(3) ($r = 0.318$, $p < .001$). Therefore, the specification in expression (2) is not valid and any inferences based on that analysis are likely to be incorrect. In summary, we reject the deviation analysis offered by Crawford (2019), which clearly suffers from an omitted-variable bias.

Once we address the matter of the omitted-variable bias, we have the following specification of the deviation analysis proposed by Crawford:

(5) $$\text{Response Bias} = \beta_0 + \beta'_1\text{Target} + \beta_2\text{RunMeanDeviation} + \beta_3\text{Last(3)Deviation} + \varepsilon.$$

We conduct the analysis below both as in expression (3) for n=3 and expression (5). Similar to DS, in order to account for the lack of independence between two observations associated with the same participant, we employ a standard repeated measures technique. We assume a single correlation between any two observations involving a particular participant. However, we assume that observations involving two different participants are statistically independent. In other words, we employ a repeated measures regression with a compound symmetry covariance matrix. Table 1 summarizes this random-effects analysis on the data from DHHC.[6,7]

[6] We note that Table 1 and the regression tables that follow are not consistent with the APA format for regressions. However, the APA format makes it difficult to display multiple specifications because the coefficient estimates and the standard errors are listed in separate columns. Since we prefer to display multiple specifications in each table, we present the regressions in a format, standard in other fields, with a regression in each column.

[7] The left panel in Table 1 is identical to the "Prec 3" specification in Table 2 in DS and is similar to the upper-left panel in Table 2 of Crawford (2019).

Table 1: Random-effects repeated measures regressions on DHHC1 dataset

	DV: Response		DV: Response Bias
Intercept	12.960*	Intercept	12.960*
	(5.909)		(5.909)
Target	0.803***	Target	-0.064**
	(0.005)		(0.024)
Running Mean	0.040	Running Mean Deviation	0.040
	(0.028)		(0.028)
Last(3)	0.092***	Last(3) Deviation	0.092***
	(0.009)		(0.009)
-2 Log L	45154.7	-2 Log L	45154.7

Notes: We provide the coefficient estimates with the standard errors in parentheses. We examine trials 2 through 190. We do not provide the estimates of the covariance parameters. Both regressions have 4696 observations. † indicates significance at $p < .1$, * indicates significance at $p < .05$, ** indicates significance at $p < .01$, and *** indicates significance at $p < .001$. -2 Log L refers to negative two times the log-likelihood.

We summarize the analysis for the second experiment from DHHC in Table 2.[8]

Table 2: Random-effects repeated measures regressions on DHHC2 dataset

	DV: Response		DV: Response Bias
Intercept	19.913**	Intercept	19.913**
	(7.339)		(7.339)
Target	0.784***	Target	-0.088**
	(0.005)		(0.031)
Running Mean	0.059†	Running Mean Deviation	0.059†
	(0.032)		(0.032)
Last(3)	0.068***	Last(3) Deviation	0.068***
	(0.009)		(0.009)
-2 Log L	84249.2	-2 Log L	84249.2

Notes: We provide the coefficient estimates with the standard errors in parentheses. We examine trials 2 through 190. We do not provide the estimates of the covariance parameters. Both regressions have 8505 observations. † indicates significance at $p < .1$, * indicates significance at $p < .05$, ** indicates significance at $p < .01$, and *** indicates significance at $p < .001$. -2 Log L refers to negative two times the log-likelihood.

Once the omitted-variable bias is corrected, the deviation analysis regressions are identical to the regressions employed in DS. Specifically, the running mean coefficient estimates

[8] The left panel in Table 2 is identical to the “Prec 3” specification in Table 6 in DS and is similar to the upper-right panel in Table 2 of Crawford (2019).

and the previous lines estimates are identical. In summary, we find evidence of a bias toward the recent lines but not toward the running mean.

Crawford states that the deviation regressions are fundamentally different than the regressions presented in DS. However, this is not true. Not only are the coefficient estimates for intercept, running mean, and previous lines identical, but so are the fit statistics. In the appendix, we also test the robustness of this finding and our results are not changed.[9] This result is true in general and in the appendix we offer such an argument.

The only novel aspect of the deviation analysis offered by Crawford is the presence of an omitted-variable bias. Once this oversight is corrected, the purported evidence in support of CAM vanishes.

Another Simulated Dataset

To address the concern that the techniques of DS are not sufficiently sensitive to detect a bias toward the running mean, DS produce a simulated dataset that has a bias toward the running mean but not toward recent lines. That simulation was calibrated to have fit statistics similar to those in the analysis of the original data. Under those conditions, the DS technique correctly identifies the simulated relationship with running mean but not towards recent targets.[10]

Crawford asked whether the techniques of DS could detect a relationship that has *both* a bias toward the running mean and toward the recent lines. When describing this possibility, Crawford notes that there is heteroskedasticity in the data. Specifically, the Last(3) variable has a

[9] We have conducted fixed-effects regressions by estimating a unique categorical variable for each participant, rather than the random-effects regressions. The results, presented in Tables A1 and A2, are not changed from Tables 1 and 2. We also conduct the random-effects on the first half of the DHHC1 dataset (before the mean of the distribution changed), presented in Table A3, and our results are not changed.
[10] Tables 3 and A13 from DS.

relatively large variance that is roughly constant across trials. However, the variance of the Running Mean variable decreases across trials.[11] We note that this heteroskedasticity also existed in the DS simulation and the DS techniques successfully identified the simulated relationship. Additionally, DS performed analyses on the non-simulated data that controlled for possibility of heteroskedasticity and found that it does not affect the results.[12]

Despite that DS accounted for heteroskedasticity in the non-simulated data and correctly identified the relationship in the DS simulation, Crawford writes on page 697, “it is possible to produce simulated data with baked-in bias toward the running mean in which the DS analysis fails to identify the running mean effect.” Crawford produces a simulation specified by:[13]

$$\text{Response} = 0.8*\text{Target} + 0.1*\text{Running Mean} + 0.1*\text{Last}(3) + \varepsilon,$$

where ε is normally distributed with a zero mean and a standard deviation[14] of 25. In the analysis of the data, it is unclear to us why Crawford exclusively reported regressions restricted to the first and second halves of trials but did not report the analysis across every trial. This decision is curious because the analysis summarized in Table 3 of DS analyzes judgments across every trial. Therefore, analyzing data across every trial would seem to be the appropriate comparison. Regardless, we proceed by evaluating the first half of trials, the second half of trials, and data from every trial.

[11] Crawford asserts that this noise is an advantage (not a disadvantage) for the significance of the Last(3) related coefficient estimates.
[12] Footnote 10 in DS.
[13] Crawford (2019) reports a specification of “Response = 1.8*Target + 0.1*Running Mean + 0.1*Last3 + ε” although we suspect that the actual coefficient of Target is 0.8, not 1.8.
[14] Crawford (2019) does not report the standard deviation, but from the code it seems to be 25.

It is also not clear to us why where are 3517 dof (and 3518 dof) in the analysis of Crawford. In order to correspond to the non-simulated dataset, our simulated dataset has 4725 observations.[15]

We conduct analyses similar to the regressions summarized in the left panels of Tables 1 and 2. We summarize the analysis of our simulated data in Table 3.

Table 3: Random-effects repeated measures regressions of Response on simulated dataset

	First Half Trials 2-95	Second Half Trials 96-190	All trials Trials 2-190
Intercept	0.097	-8.326	-0.981
	(1.951)	(8.600)	(2.443)
Target	0.804***	0.803***	0.803***
	(0.007)	(0.007)	(0.004)
Running Mean	0.106***	0.126***	0.102***
	(0.017)	(0.031)	(0.011)
Last(3)	0.090***	0.108***	0.099***
	(0.013)	(0.010)	(0.007)
Observations	2350	2375	4725
-2 Log L	21837.5	22157.5	43985.1

Notes: We provide the coefficient estimates with the standard errors in parentheses. We do not provide the estimates of the covariance parameters. † indicates significance at $p < .1$, * indicates significance at $p < .05$, ** indicates significance at $p < .01$, and *** indicates significance at $p < .001$. -2 Log L refers to negative two times the log-likelihood.

As can be seen in Table 3, our techniques are sufficiently sensitive to detect the relationship involving both the Running mean and the Last(3) variables. Further, we detect the relationship when we analyze data from every trial, but also when we restrict attention to either the first half or the second half of the data.

It is not clear to us why Crawford reported that our analysis could not detect the "baked-in bias." Unfortunately, we have been unable to obtain the data that Crawford analyzed.[16]

[15] Our simulated data is available: https://osf.io/79ntw/.
[16] Despite that Crawford (2019) states in the abstract, "Code and data available: https://osf.io/tkqvn," the data from the simulation are not apparently available.

Therefore, we are not able to compare the Crawford simulated data with ours. Crawford also does not report the fit-statistics from the analysis and this prevents even a rough comparison. Regardless, when we analyze our simulated data that was produced according to the specification suggested by Crawford, our techniques correctly identify the simulated relationship.

Figure 1 in Crawford (2019)

The third point made in Crawford relates to the relationship between Response Bias and Target in DHHC1. In the first half of the experiment, one treatment judged lines from a distribution with a large mean and the other treatment judged lines from a distribution with a small mean. In the second half of the experiment, the distributions switched: those previously viewing lines from the large distribution viewed lines from the distribution with a small mean, and vice versa.

Figure 1 in Crawford (2019) depicts the relationship between Response Bias and Target within the large and small distribution treatments, restricted to the first or second half of trials. The figure appears to show that in the first half, the large mean and small mean treatments are different but in the second half they do not appear to be different. Crawford interprets this as uniquely consistent with CAM. On page 695, Crawford writes, "Figure 1 shows that during the second half of trials, responses are biased toward the long-run mean rather than the short-run mean. The pattern seems incompatible with the DS paper's conclusion that estimates are biased toward recent stimuli rather than the long-run mean."

Because figures cannot account for multiple factors, we conduct a regression analysis, restricted to either the first or second half of trials. Since Crawford excludes the first 20 trials in the second half, we conduct our regressions by also excluding the first 20 trials in the first half.

We add the variable *Large mean*, which is assumed to have a value of 1 if the trial was in the large mean distribution, and a value of 0 otherwise.

In order to aid the interpretation of our estimates, we normalize the Target variable by subtracting the aggregate distribution mean (224). We refer to this variable as *Target Norm*. To allow for the possibility that the relationship between Target Norm and Response Bias depends on the distribution treatment, we include specifications with the interaction between Large Mean and Target Norm. We employ the repeated measures techniques used above. We summarize this analysis Table 4.

Table 4: Random-effects repeated measures regressions of Response Bias

	First Half Trials 21-95		Second Half Trials 116-190	
Intercept	-10.325**	-10.272*	-1.165	-1.374
	(3.659)	(3.683)	(3.793)	(3.811)
Target Norm	-0.213***	-0.212***	-0.185***	-0.190***
	(0.009)	(0.0129)	(0.009)	(0.013)
Large Mean	18.944**	18.936**	-1.338	-1.355
	(5.103)	(5.103)	(5.502)	(5.501)
Target Norm*Large Mean	-	-0.002	-	0.010
		(0.018)		(0.018)
-2 Log L	17883.0	17889.2	18064.5	18070.4

Notes: We provide the coefficient estimates with the standard errors in parentheses. Every regression has 1866 observations. We do not provide the estimates of the covariance parameters. † indicates significance at $p < .1$, * indicates significance at $p < .05$, ** indicates significance at $p < .01$, and *** indicates significance at $p < .001$. -2 Log L refers to negative two times the log-likelihood.

The analysis summarized in Table 4 corresponds to Crawford's interpretation of Figure 1 in Crawford (2019): there is a significant and negative relationship between Target and Response Bias in both halves, but the Large Mean variable is only significant in the first half of trials.

Next, we conduct specifications that include the Running Mean and Last(3) variables that are normalized, as was done above for Target. We refer to these variables as *Running Mean Norm* and *Last(3) Norm*. This analysis is summarized in Table 5.

Table 5: Random-effects repeated measures regressions of Response Bias

	First Half Trials 21-95			Second Half Trials 116-190		
Intercept	0.935	-5.4373	-0.524	2.097	3.592	-1.624
	(5.866)	3.7416	(2.631)	(4.022)	(3.873)	(2.862)
Target Norm	-0.209***	-0.210***	-0.207***	-0.181***	-0.181***	-0.183***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Large Mean	-3.802	9.456†	-	-7.795	-10.830†	-
	(10.805)	(5.313)		(6.108)	(5.708)	
Running Mean Norm	0.166	-	-	0.120	-	-
	(0.118)			(0.085)		
Last(3) Norm	0.0972***	0.104***	0.111***	0.104***	0.106***	0.099***
	(0.016)	(0.016)	(0.0152)	(0.016)	(0.016)	(0.016)
-2 Log L	17846.6	17846.2	17854.4	18030.6	18029.6	18038.4

Notes: We provide the coefficient estimates with the standard errors in parentheses. Every regression has 1866 observations. We do not provide the estimates of the covariance parameters. † indicates significance at $p < .1$, * indicates significance at $p < .05$, ** indicates significance at $p < .01$, and *** indicates significance at $p < .001$. -2 Log L refers to negative two times the log-likelihood.

When the Last(3) Norm variable is included, neither the Running Mean Norm nor the Large Mean variables are significant. This analysis suggests that neither of these variables are successful at explaining Response Bias. We also note that these results are robust to specifications that have fixed-effects repeated measures. In summary, Figure 1 of Crawford (2019) seems to suffer from a visual version of an omitted-variable bias. But when recency effects are included, evidence in favor of CAM again disappears.

Conclusions

Crawford (2019) offers a response to DS and a defense of CAM by making three main points. First, Crawford proposes regressions that are, in part, based on a “deviation” analysis.

However, the Crawford specification suffers from an omitted-variable bias. Once this flaw is corrected, the evidence in favor of CAM disappears. Second, Crawford offers a different simulation of data and claims that the techniques employed in DS are not sufficiently sensitive to detect the relationship in support of CAM. When we produce a simulated dataset[17] consistent with the specification in Crawford (2019) our methods correctly identify the true relationship. Despite the assertion otherwise, the Crawford simulated dataset is not publicly available and we are therefore not able to compare the two simulated datasets. Third, Crawford appeals to a figure showing that the responses appear to be biased toward the overall running mean, and presumably not toward recently viewed lines. However, when we run a careful analysis, we find that it is driven by a recency bias, rather that as suggested by Crawford, behavior consistent with CAM. In summary, once the mistakes in Crawford (2019) are remedied, there remains no evidence in support of CAM.

It is also notable what does not appear in the defense of CAM. Crawford (2019) does not present an analysis of responses across trials that would suggest learning. It is well-known that Bayesians, even those with different prior beliefs, will have beliefs that converge to the truth (Savage, 1954; Blackwell and Dubins, 1962). Therefore, DS examined the joint hypothesis that participants learn the distribution and they employ this information in their judgments. These investigations appeared in 70 specifications across 14 tables in the body and the appendix of DS.[18] Despite these efforts, no evidence of learning was found. Crawford neither disputes these efforts nor presents valid evidence of learning across trials. Since there appears to be no learning, it is not clear how the DHHC datasets could be consistent with CAM or any Bayesian model of judgment.

[17] Available here: https://osf.io/79ntw/.
[18] Summarized in Tables 4, 5, 7, 8, A3, A4, A5, A6, A7, A8, A9, A10, A11, and A12 in DS.

On page 697, Crawford wrote: “It (CAM) makes assumptions that are simplifying and skeletal (e.g., treating trace memory and priors as normally distributed) and never claims to capture the whole complexity of human judgment.” It is not our claim that CAM (or any model) needs to explain *everything* in order to be valuable. However, CAM (or any model) does need to explain *something* better than any other alternate explanation. It is not the case that there is only *some* evidence in favor or CAM. Rather, there is apparently *no* (valid) evidence in favor of CAM. Based on our analysis, CAM does not seem to explain the data better than a simple recency bias: there is a clear recency bias and there is no evidence for the joint hypothesis that the participants learn the distribution and employ this in their judgements.

In order to come to the conclusion that CAM is a valuable model of judgment there should be a set of valid falsifiable tests that would support CAM over other explanations. We have subjected the DHHC dataset to many falsifiable tests and we do not find any evidence consistent with CAM.[19] A recency bias is a simpler model that better explains the data. As such, it remains our view that evidence for CAM is a statistical illusion that occurs when researchers analyze data averaged across trials and do not consider a recency bias.

---

[19] Duffy and Smith (2019) does likewise with a dataset designed to replicate Huttenlocher, Hedges, and Vevea (2000) and they do not find any evidence consistent with CAM.

References

Blackwell, D., & Dubins, L. (1962). Merging of opinions with increasing information. *Annals of Mathematical Statistics, 33*, 882-886.

Crawford, L. E. (2019). Reply to Duffy and Smith's (2018) reexamination. *Psychonomic Bulletin & Review, 26*(2), 693-698.

Duffy, S., Huttenlocher, J., Hedges, L. V., & Crawford, L. E. (2010). Category effects on stimulus estimation: Shifting and skewed frequency distributions. *Psychonomic Bulletin & Review, 17*(2), 224-230.

Duffy, S., & Smith, J. (2018). Category effects on stimulus estimation: Shifting and skewed frequency distributions—A reexamination. *Psychonomic Bulletin & Review*, *25*(5), 1740-1750.

Duffy, S., & Smith, J. (2019). On the Category Adjustment Model: Another look at Huttenlocher, Hedges, and Vevea (2000). Working paper, Rutgers University-Camden.

Edwards, W., Lindman, H., & Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review, 70*(3), 193-242.

Hill, R.C., Griffiths, W.E., & Judge, G.G. (2001). *Undergraduate Econometrics*. Wiley and Sons: New York.

Hollingworth, H. L. (1910). The central tendency of judgment. *The Journal of Philosophy, Psychology and Scientific Methods*, 7(17), 461-469.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology: General, 129*, 220-241.

Savage, L.J. (1954). *The Foundations of Statistics*. Wiley, New York. Reprinted in 1972 by Dover, New York.

Appendix

The robustness of Tables 1 and 2

In Table A1, we summarize the fixed-effects versions of Table 1.

Table A1: Fixed-effects repeated measures regressions on DHHC1 dataset

	DV: Response		DV: Response Bias
Intercept	17.104*	Intercept	17.104*
	(6.864)		(6.864)
Target	0.804***	Target	-0.077**
	(0.005)		(0.025)
Running Mean	0.024	Running Mean Deviation	0.024
	(0.030)		(0.030)
Last(3)	0.095***	Last(3) Deviation	0.095***
	(0.009)		(0.009)
-2 Log L	44963.8	-2 Log L	44963.8

Notes: We provide the coefficient estimates with the standard errors in parentheses. We examine trials 2 through 192. We do not provide the estimates of the covariance parameters. Both regressions have 4696 observations. † indicates significance at $p < .1$, * indicates significance at $p < .05$, ** indicates significance at $p < .01$, and *** indicates significance at $p < .001$. -2 Log L refers to negative two times the log-likelihood.

The fixed-effects specification does not change the results From Table 1. In Table A2, we summarize the fixed-effects version of Table 2.

Table A2: Fixed-effects repeated measures regressions on DHHC2 dataset

	DV: Response		DV: Response Bias
Intercept	-3.784	Intercept	-3.784
	(6.618)		(6.618)
Target	0.783***	Target	-0.109**
	(0.005)		(0.035)
Running Mean	0.039	Running Mean Deviation	0.039
	(0.036)		(0.036)
Last(3)	0.069***	Last(3) Deviation	0.069***
	(0.009)		(0.009)
-2 Log L	83867.0	-2 Log L	83867.0

Notes: We provide the coefficient estimates with the standard errors in parentheses. We examine trials 2 through 190. We do not provide the estimates of the covariance parameters. Both regressions have 8505 observations. † indicates significance at $p < .1$, * indicates significance at $p < .05$, ** indicates significance at $p < .01$, and *** indicates significance at $p < .001$. -2 Log L refers to negative two times the log-likelihood.

The fixed-effects specification does not change the results in Table 2.

To address the possibility that the results only hold in the first half of trials in DHHC1 (before the means of the distributions switched), we offer an analysis where only trials in the first half of DHHC1 are considered. This analysis is summarized in Table A3.

Table A3: Random-effects repeated measures regressions on first half of DHHC1 dataset

	DV: Response		DV: Response Bias
Intercept	19.477*	Intercept	19.477*
	(8.438)		(8.438)
Target	0.798***	Target	-0.091*
	(0.008)		(0.035)
Running Mean	0.010	Running Mean Deviation	0.010
	(0.039)		(0.039)
Last(3)	0.101***	Last(3) Deviation	0.101***
	(0.0148)		(0.0148)
-2 Log L	22263.7	-2 Log L	22263.7

Notes: We provide the coefficient estimates with the standard errors in parentheses. We examine trials 2 through 95. We do not provide the estimates of the covariance parameters. Both regressions have 2334 observations. † indicates significance at $p < .1$, * indicates significance at $p < .05$, ** indicates significance at $p < .01$, and *** indicates significance at $p < .001$. -2 Log L refers to negative two times the log-likelihood.

Only analyzing the first half of trials-before the mean has switched-does not change the results of Table 1.

The general relationship between expressions (3) and (5)

The primary analysis in DS used the specification:

$$\text{(A1)} \qquad \text{Response} = \beta_0 + \beta_1\text{Target} + \beta_2\text{RunningMean} + \beta_3\text{Last(n)} + \varepsilon.$$

Crawford claimed that this specification was not sufficiently sensitive to detect evidence of CAM. Crawford proposes that the dependent variable should be Response Bias, not Response.

To accomplish this, we subtract Target from both sides of the equation:

$$\text{Response Bias} = \text{Response} - \text{Target} = \beta_0 + (\beta_1 - 1)\text{Target} + \beta_2\text{RunningMean} + \beta_3\text{Last(n)} + \varepsilon.$$

Next we take the "deviation" of the two primary independent variables of interest: Running Mean and the Last(n):

$$\text{Response Bias} = \beta_0 + (\beta_1 - 1 + \beta_2 + \beta_3)\text{Target} + \beta_2(\text{Running Mean - Target}) + \beta_3(\text{Last(n) - Target}) + \varepsilon.$$

We can rewrite this expression in using the deviation variables defined in the body of the paper. We note that this

This is identical to expression (5) where n=3 in the body of the paper:

(A2) $$\text{Response Bias} = \beta_0 + \beta'_1\text{Target} + \beta_2\text{RunMeanDeviation} + \beta_3\text{Last(n)Deviation} + \varepsilon.$$

In general, the coefficient estimates of β_2 and β_3 will be identical in each of the regressions above. Note that the coefficient for RunningMeanDeviation in (A2) is identical to the coefficient for RunningMean in (A1). Also note that the coefficient for Last(n)Deviation in (A2) is identical to the coefficient for Last(n) in (A1). When Crawford attributes significant differences between the estimates in the deviation analysis and the techniques used by DS, these differences are exclusively drive by the omitted-variable bias.