

Author response to reviews of

Manuscript CEP-2018-0068

Instance theory predicts information theory: Episodic uncertainty as a determinant of keystroke dynamics

Crump, Lai, & Brosowsky

submitted to *Canadian Journal of Experimental Psychology*

RC: Reviewer Comment AR: Author Response ☐ Manuscript text

Dear Dr. Glen Bodnar,

Thank you for allowing us the opportunity to respond to the issues raised by yourself and the reviewers in the action letter. We have carefully considered each of the comments and addressed each of them. Our response to each comments is detailed below.

We found the comments very helpful for improving our thinking and writing about this topic, and believe our manuscript is much improved.

1. Overview of changes

Before addressing each comment, we provide an overview of the major changes in this revision.

1.0.1 Major findings and conclusions

Following the advice of reviewer 3 (Michael Masson), we recalculated H conditionalized for letter n-1 identity, position and word length as weighted means (by contextual probability). This new measure was similar to the unweighted one, but explained less variance (.7) than the original one (.81). Furthermore, in response to reviewer 1, we also conducted a new regression that added letter position (first vs. other), and showed that it explained more variance (.88), than conditionalized letter uncertainty alone. We also revised our main conclusions. We believe our data 1) do support the idea that a large portion of the variance in mean IKSI is explained by variation in letter uncertainty, but 2) we do not rule out the role of planning processes in producing first-letter slowing effects. We greatly expand our discussion of these points in the GD.

1.0.2 Modeling

We agreed with reviewer 1 that our first submission did not connect deeply enough with prior modeling work. We expanded the modeling section in several ways. First, we explain how Jamieson & Mewhort (2009) used an instance model to account for the hick-hyman law. We relate our model to theirs by also showing that our model can explain the results of Hyman (1953, as Jamieson & Mewhort, 2009, did). We believe this section also does a better job of explaining how we implemented the model. We then applied the same model to the problem of typing. We removed our prior simulation of letter uncertainty (unconditionalized by n-1),

and replaced it with a model simulating mean RTs as a function of letter uncertainty (conditionalized, as suggested by reviewer 1). We also made efforts to be more clear that our modeling results show that the instance theory of automatization makes exact predictions for mean RT as a function of uncertainty (scaled by constant factors).

1.0.3 General discussion

We added several new sections to the general discussion, including a new section discussing major conclusions about planning vs. uncertainty based interpretations of first-letter and mid-word slowing. We discuss the connection between our efforts and Logan's (2019) recent CRU model of typing performance. We elaborate on implications and future directions, and eliminated strong claims about causality.

In sum, thanks to the reviewers, we are much happier with the present revision.

2. Action Editor

RC: Add the required public significance statement.

AR: The public significance statement has been added.

RC: Rather than quoting large sections of your chapter in the method, provide a paragraph description of the key details of the method (including the ethics statement) and move the rest of the details to a Supplemental Materials file. Be sure to clarify that participants performed additional tasks (as noted by R1).

AR: We took this advice. The methods section is now a single paragraph with the highlights (including mention that typists performed multiple typing tasks, specifically typing random strings and English-like non words, in addition to the five English paragraphs). We also note additional details can be found in the supplemental materials document.

RC: The writing is very crisp but you introduce your use of Monte-Carlo simulation twice (p. 16 then p. 17).

AR: Thank you for catching that, now there is only one use.

RC: Change the wording/grammar on page 3 to “measure the structure of the letter sequences that typists type”.

AR: Thank you, changed.

RC: Also, I would ask you to consider making your data and/or methods openly available and in doing so qualifying for one or more of the Open Science Framework badges that we now offer at CJEP. Information on the badges is available near the bottom of our manuscript submission page, here: <http://www.apa.org/pubs/journals/cep/?tab=4>. We invite authors to deposit their data in APA's own repository at <https://osf.io/view/apa/>. This repository will help researchers work on a project privately, with collaborators, or allow them to make parts of or their entire project publicly accessible. The repository will store and archive research data, protocols and materials, with data being made open once it is published in an APA journal. If you choose to submit your data to the repository, please include the dataset's OSF link as part of your author note, and include the data citation in your reference list.

AR: We suspect this request may be part of the boilerplate for action letters. In our original submission, we applied for the Open Data badge. Our data, code, and entire manuscript are completely open, and reproducible (using R). We have now included the OSF link in the author note, and in the references.

3. Reviewer #1

RC: **This manuscript addresses a timely topic that would likely be of interest to readers of this journal. I think the work provides some unique insights on typing performance and how it is influenced by statistics of the environment (e.g., letter and bigram frequencies in naturalistic text). The authors make some interesting observations, but I think it is premature to claim that they “have developed a falsifiable causal theory of variation in keystroke dynamics” (p. 20), especially in the absence of direct experimental testing. (They might want to dial down such claims.) My impression is that this manuscript could be improved substantially by more empirical evidence and/or more sophisticated modeling. As it currently stands, I am on the fence about whether there is enough here to warrant publication as a stand-alone contribution. At the very least, I think my major comments would need to be addressed before this work might be publishable.**

AR: Thank you for your thoughtful feedback, we hope our revision is sufficient. We have dialed down our causal claims as suggested, and made substantial revisions throughout to address the comments.

Major Comments

RC: **1. Missing statistical support for the key empirical findings. Visual inspection of Figure 1A is consistent with the authors’ claims about first-letter and mid-word slowing, but it would be good to have more specific statistical support (i.e., statistical tests tailored to the effects of interest rather than all the pairwise comparisons in Figure 1B). For example, a statistical analysis comparing first-letter IK-SIs versus other-letter IK-SIs might suffice to indicate first-letter slowing. Mid-word slowing is a bit more complicated to analyze because it depends on word length, but there should be a way to do it. My overarching point is that the currently reported statistics do not seem to directly assess the specific trends of interest.**

AR: We agree it would be a good idea to report statistics beyond all of the pairwise comparisons, and thank the reviewer for the suggestion. We have removed the pairwise comparisons entirely. We include two new sections in the results about the first-letter slowing effect, and the mid-word slowing effect (we followed Ostry’s method to analyze mid-word slowing). In both cases, there is statistical support for a first-letter effect and a mid-word slowing effect.

RC: **2. How clear-cut is the support for the authors’ preferred model? The authors find that letter uncertainty (conditionalized on n-1 letter identity) does well at accounting for the IKSI pattern ($R^2 = .81$). However, they noted that letter uncertainty (unconditionalized) and position as predictors also works well - perhaps even better ($R^2 = .86$). Is there a reason why the first model should be preferred, beyond an argument based on parsimony? Framed another way, is the evidence strong enough to reject a dual-process interpretation involving a planning/buffering process?**

AR: This are excellent questions that we did not do enough to address in the original manuscript. We have now added a new section to the general discussion called “Planning versus learning about letter uncertainty”, where we discuss these issues in detail. In short, we do not reject a dual-process account, and we discuss relationships between the accounts.

RC: 3. Model development, assessment, and connections to the literature: I have a few comments related to the modeling effort:

AR: As an overview, we made numerous changes to the presentation of the model, and substantially elaborated on connections to other modeling efforts in the literature.

RC: (a) Considering the authors' findings concerning the importance of n-1 letter identity, I was puzzled as to why n-1 letter identity was not incorporated into their instance model simulations. It should be possible to develop a variant of their model in which either "bigram instances" are retrieved or bigram information provides context for retrieval.

AR: The short answer to this question is that we now report the model that includes n-1 letter identity. We agree this is what a reader would be expecting to see, and that is one reason we added it. It is also worth noting that the general question at hand is whether predictions for performance from the model are the same as predictions based on measures of uncertainty. We found that instance theory predictions for mean RT for a set of items, were perfectly explained by H for the set of items, as practice increased. In other words, instance theory predictions were linearly related to H by constant factors.

RC: (b) The model was assessed for relatively small amounts of practice, whereas the subjects who provided the typing data likely had extensive pre-study practice in standard typing. Considering the practice-related compression in simulated RTs in Figure 4 (i.e., the range in RT seems to become narrower with practice), I wonder whether the key trends (first-letter and mid-word slowing) would still occur after much more practice. For example, in the "500" practice condition, mid-word slowing already seems somewhat muted. What happens if the model is simulated for much longer periods of practice (e.g., 10,000 keystrokes)?

AR: We now extend practice to 10,000, and show that instance theory predictions become even closer to letter uncertainty as practice increases.

RC: (c) Beyond visual inspection of Figure 4, can the authors provide any sort of quantitative comparison between the simulation results and the empirical IKSI data?

AR: We argue that comparing model predictions to empirical IKSI data is redundant with our existing analysis relating H to IKSI. This is because predictions from the model are identical to H, and would show the same thing in relation to the IKSI data.

RC: (d) I think there should be more discussion about how the present modeling effort is connected with other models in the literature; specifically, Jamieson and Mewhort (2009) and Logan (2018), both of which are cited but not adequately discussed. Jamieson and Mewhort showed that an instance model using recent context could produce the Hick-Hyman law, which seems very much related to what the present authors wanted to demonstrate. Logan showed that an instance-based, context-driven retrieval process could account for various kinds of error data in typing. Even though he did not model keystroke timing, he discussed how his model might handle it.

AR: Thank you for this comment. We completely agree that we left out important avenues for discussion. We have expanded on these connections in new sections. We now credit Jamieson and Mewhort (2009) for first applying an instance model to the hick-hyman law. We motivate our model as an extension to their efforts. We also now add a simulation of the Hyman (1953) experiments to show similarities and differences between our approach and the MINERVA 2 approach. We added an extended discussion of the CRU model (Logan 2018) to the general discussion.

Minor Comments (chronological)

RC: 4. p. 4: “IKSIs” should appear in parentheses after the first occurrence of “interkeystroke intervals” because the abbreviation is used later.

AR: Fixed

RC: 5. p. 8, Design and Procedure: It is unclear why “the task took around 30 to 45 minutes” when subjects only had to type five short paragraphs, which likely would have taken about 10 to 15 minutes at typical typing speeds. (This was clarified only when I consulted the Behmer and Crump chapter, and discovered that subjects also typed other texts.)

AR: We now clarify that typists performed additional typing tasks, and followed the action editor’s advice to move most the methods to supplemental materials. Additionally, some of the extra time involved reading task instructions, reading and agreeing to the consent form, and filling out a demographic questionnaire.

RC: 6. p. 8, Data Analysis: Is it really necessary to list all the R packages used? I think it would be simpler to state that R was used to analyze the data, then direct the reader to the GitHub repository in the Author Note for more details.

AR: It is not necessary the reference have been removed.

RC: 7. Considering that this manuscript involves reanalyzing data from Behmer and Crump (2017), it might be useful to briefly summarize the data analyses and results from that chapter (perhaps in a footnote).

AR: We added a longer summary of the Behmer & Crump (2017) results in the general discussion, where we connect the issues in this paper to other lines of evidence showing support for an instance-based view of sequential action in typing.

RC: 8. p. 11: “Visual inspection of the graph shows that variation in letter uncertainty maps closely onto variation in mean IKSI...” I would not say that “maps closely” accurately describes the correspondence (or lack thereof) between Figures 1A and 2A, as indicated by Figure 2B.

AR: We have rewritten that section, and agree that maps closely is inaccurate. The revised section now states “To our knowledge this a novel analysis of how letter uncertainty in natural English varies by word length and position. We were interested in whether letters appearing the first position of words would have larger H values than letters in other positions; and, whether letters appearing the middle of words would have higher H values than letters appearing around the surrounding positions. Visual inspection of the graph shows letters in the first position have generally large H values, and letters in the middle positions have generally larger H values than letters surrounding the middle position. This suggests that natural variation of letter uncertainty across position and word in English may account for aspects of the first-letter and mid-word slowing phenomena in typing.”

RC: 9. p. 15: “causal claim” The correlational data in Figure 3 cannot be used to infer causality.

AR: We agree the correlational nature of the data cannot be used infer causality, and are aware of the correlational nature of the design. We also agree that our discussion of causality needs to be clarified. This section has now be revised to say “The correlational nature of this evidence prevents any causal conclusions about how variation in mean IKSI might be caused by variation in uncertainty. In the general discussion we discuss further steps for future work to establish causality by experiment. In addition to empirical work, it is also necessary provide a working process model that articulates how variation in letter uncertainty could

cause variation in mean IKSIs. In this section, we establish theoretical plausibility by showing that letter uncertainty influences on performance can be explained in terms of Logan's (1988) instance-based memory model of automatization."

RC: 10. p. 17: "This assumes traces for specific letters are stored and retrieved in a context-dependent fashion." This seems consistent with the context-based retrieval mechanism in Logan's (2018) model of typing.

AR: We agree, a citation is added. Additionally, we cite prior empirical work showing IKSIs for specific letters are context-dependent.

RC: 11. The conditionalized aspects of the results (i.e., that IKSI depends on letter position, word length, and n-1 letter identity) fit well with the Logan-Crump two-loop theory of typing, in which the critical unit for typing is the word rather than individual letters. Letter position, word length, and n-1 letter identity make sense as useful predictors of typing performance for word units but not for single-letter units, so it might be worth mentioning this connection to the two-loop theory.

RC: 12. (a) Figure 1: It is unclear what the error bars represent in Panel A. (b) Figures 2 and 3: It is unclear what the shaded area around the regression line represents in Panel B of each figure.

AR: We have changed all of the figures. When error bars are drawn the figure captions explain what they are (e.g., 95% confidence intervals). The shaded area for the regression line was a continuous 95% confidence interval, but we have removed them.

RC: 13. Figures: It is difficult to distinguish between nine levels of grayscale shading. The authors might want to consider using different symbols (circle, square, triangle, diamond, etc.) to make it easier for the reader to distinguish between conditions.

AR: We agree the original figures were difficult to interpret. The new figures now present mean IKSIs and H values as a function of word length and letter position in different facets, all with the same level of shading. We think this makes it easier to see the patterns in the data.

RC: 14. Grammatical nitpicks: (a) All three occurrences of "it's" in the manuscript are grammatically incorrect (should be "its" without an apostrophe). (b) Reaction times or retrieval times can be short or long, not fast or slow. The underlying process can be fast or slow.

AR: Thank you for these nitpicks. They have all been fixed.

4. Reviewer #2

RC: This is a splendid paper; it should be published. That said, I have a few comments that, I trust, the authors will find useful.

AR: Thank for that assessment and for your helpful comments.

RC: IKSL not defined as an abbreviation before it is used.

AR: fixed

RC: Tulving (1963, Amer J. Psychol., 76, 143-146) discussed how to adjust Miller's estimates of information per letter. I don't know whether his arguments affect yours, but I thought I'd note Tulving's paper for

your consideration.

AR: Thank you for the reference.

RC: The long list of R packages should be in a footnote or appendix; it distracts the reader.

AR: That list is now deleted

RC: We need a more complete account of the model. How close is it, for example, to the Minerva (Hintzman's multi-trace memory model) or to Jamieson's multi-trace model for the Hick-Hyman law (QJEP, 2009, 62 (9), 1757 - 1783). I need a description in addition to a pointer to code (which I may not be able to read, depending on the programming language).

AR: In the revision we have attempted to provide a clearer description of the model. We did not engage in close model comparison between the incarnations of MINERVA and the instance theory of automatization. The code for the model is available in the supplementary materials, and in the repository for this manuscript. The model is written in R.

5. Reviewer #3

Reviewer #3: The authors provide a convincing demonstration that time between keystrokes during typing is closely related to the uncertainty associated with the typed letter, given the letter's position, the word's length, and the identity of the preceding letter. This result is consistent with Logan's instance theory of skill acquisition, which in turn links to information theory.

RC: This is a very interesting proposal and the argument is convincing. I have only a few minor suggestions for improving the presentation.

AR: Thank you for your time and effort reviewing this work, and for your helpful comments.

RC: 1. p. 13, bottom. The H calculation for the n-1 distributions is based on the mean of all 26 possible distributions. But taking the simple, unweighted mean implies that all 26 letters are equally likely at a given position in a word with a given length. The authors have shown, however, that these likelihoods vary, so it might be better to generate weighted averages for H with each letter weighted by its context-specific probability.

AR: Thank you for pointing this out. We have recalculated H using weighted averages as suggested. The general pattern of H (conditionalized by n-1 letter, position and word length) remains the same, but the amount of variance in mean IKSI is reduced to .7.

We also added a new regression that included letter position (first vs. other), as we had originally done with the analysis for letter uncertainty that did not include n-1 letter identity. The major finding is that adding in a factor coding first vs. other position explains additional variance. We have modified our major conclusions in the general discussion. We note that the evidence suggests that letter uncertainty does explain a large portion of the variance, but that first-letter slowing appears to have some unique variance, likely to due to a planning process.

RC: 2. p. 21, second paragraph. The authors mention that it is possible that the first trace entered into memory is one that has a short retrieval time. This implies that in the model, each trace has a fixed RT associated with it. I do not think that is how the instance theory is modeled. I believe each trace

has an independently determined RT each time a retrieval attempt is made. So a short RT after much practice comes from having many chances for at least one of the traces to have a short RT on a given retrieval attempt. But the RT for a particular trace may vary across retrievals. Michael Masson

AR: First, we removed sections of the general discussion that discuss the possibility of single-trial learning. We thought this aspect of the discussion was more tangential than necessary to the main points. Second, our implementation of the model used independently determined RTs each time a retrieval was made.