Decision contamination in the wild: Sequential dependencies in Yelp review ratings

David W. Vinson, Rick Dale
Cognitive and Information Sciences
University of California, Merced
[dvinson][rdale]@ucmerced.edu

Michael N. Jones
Departments of Psychology, Cognitive Science, and Informatics
Indiana University, Bloomington
jonesmn@indiana.edu

Abstract

Current judgments are systematically biased by prior judgments. Such biases occur in ways that seem to reflect the cognitive system's ability to adapt to the statistical regularities within the environment. These cognitive *sequential dependencies* have been shown to occur under carefully controlled laboratory settings as well as more recent studies designed to determine if such effects occur in real world scenarios. In this study we use these well-known findings to guide our analysis of over 2.2 million business review ratings. We explore how both within-reviewer and within-business (between reviewer) ratings are influenced by previous ratings. Our findings, albeit exploratory, suggest that current ratings are influenced in systematic ways by prior ratings. This work is couched within a broader program that aims to determine the validity of laboratory findings using large naturally occurring behavioral data.

Keywords: Sequential dependency; Online reviews; Large natural data; Decision making

Introduction

Humans are surprisingly bad at rating the absolute magnitude of their internal cognitive states. Regardless of the task, judgments of the *absolute* magnitude of a stimulus, experience, or feeling, are inherently contaminated by *relative* information from the sequence of judgments prior to the current one. Although we tend to believe that our judgment reflects the absolute value of the current experience, a good deal of the judgment is in fact determined by the relative difference between the current experience and experiences from previous trials (Laming, 1984; Stewart, Brown, & Chater, 2005). This pattern is complicated by the fact that decisions are also influenced by other factors, such as stimulus, response, and feedback (see Donkin, Rae, Heathcote, & Brown, 2015, for a review).

These cognitive *sequential dependencies* (SDs) occur whenever behavior on a trial is influenced by behavior on preceding trials. Far from rare, SDs are ubiquitous in cognition, contaminating absolute judgments from low-level perception all the way up to high-level moral judgments. We see the effect of previous trials on RT, accuracy, the type of errors produced, and interpretation of ambiguous stimuli. SDs seem to affect all levels of the cognitive system, including motor control (Dixon, McAnsh, & Read, 2012), spatial memory (Freyd & Fink, 1984), face perception (Hsu & Yang, 2013; Liberman, Fischer & Whitney, 2014), selective attention (Kristjansson, 2006), decision making (Jesteadt, Luce, & Green, 1977), and language processing (Bock & Griffin, 2000).

SDs have primarily been studied in the laboratory or at least with well-controlled experimental stimuli. They are more difficult to study in real-world scenarios because of the very large number of trials that would be required to identify their effects. In stimulus identification, for example, the immediately preceding (n-1) and non-adjacently preceding (n-2…7) items exert opposing forces on identification of the stimulus presented on trial n (Lockhead, 2004). To observe this pattern in a reasonable amount of time in the lab, carefully designed stimulus sequences are needed.

In this paper, we explore SDs in a real-world situation by mining a large natural database of online review ratings from Yelp, Inc. This is one of many freely available structured databases that can be explored. Here we use the dataset to determine if current review ratings are contaminated by previous reported experiences. In what follows we first review SD trends observed in standard laboratory tasks.

SDs in the Laboratory

Assimilation occurs whenever the judgment of stimulus n moves closer on the measurement scale to the judgment of stimulus n-k than it otherwise would have been. *Contrast* is the opposite effect, when the judgment of stimulus n moves further away on the measurement scale from the judgment of stimulus n-k. In this sense, assimilation can be thought of as a pulling force from the preceding stimulus, while contrast can be thought of as a pushing force (Zotov, Jones, & Mewhort, 2011).

Much of the early work on SDs was psychophysical in nature and involved rating unidimensional stimuli such as the loudness of a tone or length of a line (Garner, 1953; Holland & Lockhead, 1968). Identifying the *absolute magnitude* of these stimuli (e.g., line length) has been well studied: Errors when identifying stimulus n assimilate towards the stimulus on trial n-1[1]. Participants are more likely to estimate the absolute value of some stimuli more similar to their most proximal previous estimate. Oddly, *categorization* of the same stimuli shows the opposite effect from the most recent response—when placing stimuli into categories, classification of stimulus n shows contrast from stimulus n-1 (Stewart, Brown, & Chater, 2002; Ward & Lockhead, 1971).

The contrast effect (push) of trial n-1 on the category rating of trial n is not limited to low-level perception, but is seen across levels of cognition. As a striking high-level demonstration, consider Parducci's (1968) example of classifying the event of "poisoning a neighbor's barking dog that was bothering you" on a moral judgment scale from 1-10 scale (where 10 is "extremely evil"). This terrible statement was rated as more evil by subjects if it was preceded by a mild judgment ("stealing a towel from a hotel") than if it was preceded by a nastier judgment ("using guns on striking workers")—a contrast effect when classifying moral judgments. Similar patterns of SDs have been seen in a variety of laboratory tasks designed to tap real-world scenarios, including brake initiation latencies in driving behavior (Doshi, Tran, Wilder, Mozer, & Trivedi, 2012), jury evidence interpretation (Furnham, 1986), and clinical assessments (Mumma & Wilson, 2006). In addition, SDs seem to be immune to practice—they are seen even in overlearned and expert behaviors.

At first glance, SDs appear to be an irrational bias in decision making (or perhaps in event memory), and have been traditionally viewed as the natural by-product of low-level brain dynamics such as residual neural activation. However, more recent theoretical perspectives suggest that SDs may be a rational property of any cognitive system. These accounts characterize SDs in terms of an individual's adaptation to the statistical regularities of a nonstationary environment with related stimulus bundles (Qian & Aslin, 2014; Wilder et al., 2010; Yu & Cohen, 2009).

Our interest is to mine Yelp, guided by knowledge from laboratory studies, to look for these naturally occurring contaminations that may affect how a business is currently rated and can expect to be rated in the future. Future business demand is largely influenced by online reviews (Cantallops, Silva, 2014; Mudambi & Schuff, 2010) affecting a business's revenue between 5-9% with this number increasing by 50% for businesses with more than 50 reviews (Luca, 2011). Computational models that explain how SDs emerge from the decision making process are now being developed, at least for low-level perceptual tasks (e.g., Mozer, et. al, 2010). These models have great promise in that they may be reversed and then applied to rating data to "decontaminate" the SD pollution in the rating, essentially producing a more accurate estimation of the individual's absolute experience of a business by removing the pollution from the relative information. This has an obvious benefit to both the service quality Yelp aims to provide, as well as a more accurate assessment of the business in question.

In Yelp, reviewers rate their experience with a business on a scale of 1 to 5 stars. Because both the rating and rating scale are most similar to categorization tasks studied in the laboratory (i.e., what is the best label to classify the exemplar, experience with the business, on a scale of 1-5 stars), our predictions are loosely drawn from SDs in categorization. Businesses typically specialize in specific services (such as a restaurants that serve American cuisine) while reviewers typically do not provide more than one review per business—similar to many individuals rating the same moral statement, opposed to one individual rating different statements. In particular, we expect that within reviewers we will see a *contrast* effect from ratings across businesses: The rating of a business will be artificially inflated if previous ratings from this reviewer were lower than if they were higher. Secondly, and more tentatively, we expect that businesses may act like categories themselves—a rating of a business is likely to *assimilate* towards preceding ratings. In addition, while little known work investigates the effects of SDs on increasing temporal distances, we anticipate that the effects of stimulus distance will be similar to temporal distance (cf. Ward, 1973). In this sense, our predictions of Yelp review ratings are a simple extension of both the perceptual work of Zotov et al. (2011), and the moral judgments of Parducci (1968).

Natural datasets are wrought with noise. Yet, where they lack structure they make for in sheer size. We do not anticipate that SDs will play such a substantial role as to alter the usefulness of user or business ratings on its face. Instead we expect to fine *echoes* of cognitive influence detectible in large datasets of naturally occurring behavior. We consider this work a guided exploration, in an effort to bridge laboratory findings with relevant and functional natural behavior.

Method

We used the most recent release of the Yelp Inc., dataset, part of Yelp's Dataset Challenge[2]. The dataset consists of just over 2.2 million reviews spanning 12 years from 2004-2016, with ratings between one (negative) and five (positive) stars, from approximately 552,000 reviewers on roughly 77,000 businesses. Reviews were provided from nine cities across four different countries (United States, Canada, Scotland and Germany). Interestingly, star ratings follow a J-shaped distribution (**fig. 1**, *top*) with mostly four and five star ratings a dip in two star ratings and roughly and equal number of one and three star ratings. In addition, the number of reviews increased steadily over Yelp's lifetime (**Fig. 1**, *center*).

[1] Interestingly, the same absolute judgment that assimilates to the most proximal past judgment *contrasts* from stimuli n-2…5.

[2] Further information on how to access the dataset for free as part of Yelp's dataset challenge can be found here at

[2] Further information on how to access the dataset for free as part of Yelp's dataset challenge can be found here at http://www.yelp.com/dataset_challenge

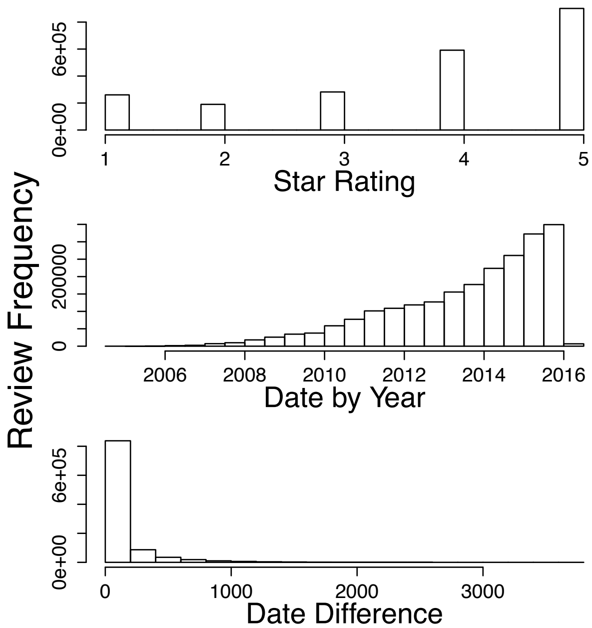

Figure 1: Frequency of reviews by star rating (*top*), frequency of reviews by year (*center*), frequency of reviews at different temporal distances in days (*bottom*).

We tested whether previous reviews influence the current review both within reviewers and within businesses. For this reason, we predict reviews from the same reviewer will be pushed *away* from previous reviews showing a contrast effect (cf. Zotov et al., 2011). Alternatively, there may be an assimilation effect for reviews within businesses. When successive stimuli are presented from the same category, the representation of that category is pulled toward the exemplar of the previous trial. Similarly, successive reviews on the same businesses, a type of category, may be pulled toward ratings from previous reviews. To be sure, this prediction is more exploratory as the nature of most laboratory studies on SDs have not focused on the influence of previous judgments from other individuals on the assessment of the same category. We anticipate that these effects will dissipate the farther away the previous review is from the current review.

Measures We first determined how far the current review rating was from its mean:

$$R_x - M(R_{T\text{-}x})$$

Where R_x is the current rating and $M(R_{T\text{-}x})$ is the average rating by reviewer or business with the current value x removed to account for possible inflation within our statistical models. This allows us to determine whether the current review is systematically biased away from the average response relative to the value of the preceding n-k review(s). To assess how distance is related to this deviation measure, we use *Review Distance* (k), and *Date Difference*. *Review Distance* is a lag measure of the number of reviews (k) between the current review and previous review, while *Date Difference* is a time measure of the number of days between reviews. Reviews that are farther displaced both in time and in the number of reviews may show dependence on previous review ratings.

Results

We first determined whether one's current review was related to one's previous review rating at k-distances. **Fig. 2** presents the mean and standard error bars for deviation of the current review rating from the mean (y-axis) by the previous star ratings (x-axis) at seven different Review Distances (k) within reviewers. The figure reveals a *contrast* effect that dissipates the farther away the previous review is from the current review. At n-1 (the immediately preceding review) for example, a 1-star rating resulted in an artificial increase in the subsequent rating from the overall mean rating. The opposite would be the case if the n-1 rating was 5 stars—the subsequent rating would be a lower star rating on average than it otherwise should have been. In this sense, the data are very much consistent with Parducci's (1968) "dog poisoning" example in that the current rating is systematically biased in the opposite direction from the previous rating.

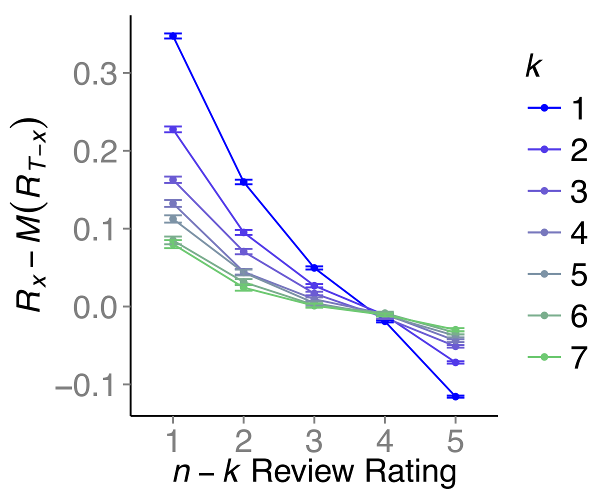

Figure 2: Within-reviewer contrast between previous and current review ratings at k Review Distances

To assess the visual impression quantitatively, we use a linear model to predict current review ratings by n-k ratings for each of seven different values of k. That is, we treated each value of k as distinct. The results, presented in **Table 1,** reveal that as the value of k increases, or the current review is farther displaced from the previous review, the contrast effect dissipates. However, due to size of our dataset, all results show a significant negative relationship, accounting for ~2% of variance at the closest Review Distance (k = 1).

Table 1: Regression model by Reviewer

k	99.9% (CI_b)	F (df)	R^2_{adj}
1	(-.11, -.10)	$2.8x10^4$ (*1, 1.7x10*^{6})	.016
2	(-.07, -.06)	8482 (*1, 1.4x10*^{6})	.01
3	(-.05, -.05)	3681 (*1, 1.4x10*^{6})	.003
4	(-.04, -.03)	1997 (*1, 1.1x10*^{6})	.002
5	(-.04, -.03)	1333 (*1, 1.0x10*^{6})	.001
6	(-.03, -.02)	756 (*1, 9.4x10*^{5})	.001
7	(-.03, -.02)	562 (*1, 8.7x10*^{5})	.001

Turning now toward within-business reviews, **Fig. 3** presents the mean and standard error bars for the deviation of the current review rating from the mean (y-axis) by the previous star rating (x-axis) at different Review Distances (k). This figure suggests an *assimilation* effect that dissipates the farther away the previous review is from the current review.

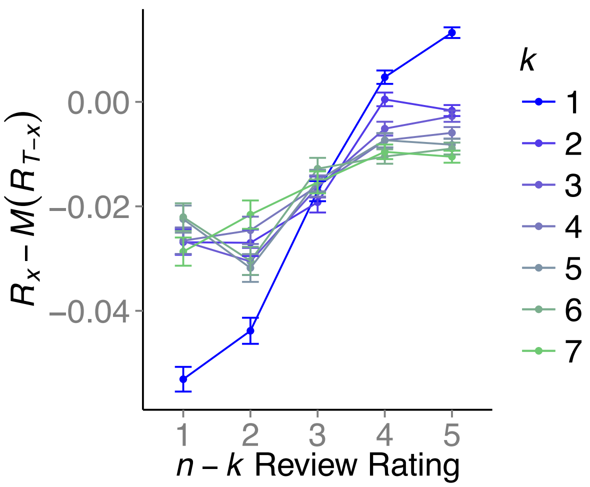

Figure 3: Within-business assimilation between previous and current review ratings at k Review Distances

The results of seven linear regression analyses on the within-business data are presented in **Table 2**. The linear regression model shows a significant negative relationship between previous and current star ratings. As the value of k increases the model accounts for less of the variance in current star ratings. All results show a significant negative relationship, though accounting for less than .1% of variance at the closest Review Distance (k = 1). Hence, this within-business assimilation effect is considerably weaker than the within-reviewer contrast.

Table 2: Regression model by Business

k	99.9% (CI_b)	F (df)	R^2_{adj}
1	(.02, .02)	1179 (*1, 2.1x10*^{6})	<.001
2	(.01, .01)	203 (*1, 2.1x10*^{6})	<.001
3	(.01, .01)	175 (*1, 2.0x10*^{6})	<.001
4	(.004, .007)	121 (*1, 1.9x10*^{6})	<.001
5	(.003, .007)	75 (*1, 1.9x10*^{6})	<.001
6	(.002, .006)	56 (*1, 1.8x10*^{6})	<.001
7	(.002, .006)	54 (*1, 1.8x10*^{6})	<.001

Next, we briefly explore whether there is a similar effect for reviewers and businesses by *date*. Yelp provides the date (in hours) for each review. However, reviewers occasionally provide multiple reviews within the same time frame (e.g., hours, days). To clearly differentiate between previous and current reviews, we first took the average review rating per day, by reviewer and business, and rounded this to the nearest star rating (1-5). The distribution of the number of days occurring between successive reviews is shown in **Fig. 1** (*bottom*). Temporal distances between successive reviews for both within-reviewer and –business were log-normally distributed and thus log transformed for all subsequent analyses (i.e., there were significantly more reviews that occurred closer to one another in time, than across time). We call this *Date Difference*—the number of days between reviews—and use it in subsequent analyses below.

Using a simple linear regression model, we first centered and squared Date Difference. There was a significant interaction between Date Difference and lagged star rating (t = -63), such that as the time between a reviewer's previous star rating increased, an observed contrast effect became more extreme, $F(3,1.0x10^6)$ = $1.65x10^4$, R^2_{adj} = .047.

Turning to within-business effects of temporal distance on current review ratings, a linear regression analysis revealed a significant interaction between Date Difference and lag star rating (t = -53), $F(3,1.9x10^6)$ = 1061, R^2_{adj} = .002. This within business effect, albeit weak, is also a between reviewer effect—a study yet to be tested in a controlled laboratory environment. That is, businesses are often not reviewed sequentially by the same reviewer, if ever. However, the effects of this interaction are less clear, showing a slight assimilation effect, that reverses to contrast only at the longest temporal intervals, requiring a more sophisticated series of analysis, discussed below, prior to further interpretation.

Discussion

This study was a guided exploration into the influence that previous business ratings might have on current ratings. We tested the presence of sequential dependencies in business reviews both within reviewer and business, finding that there are significant, albeit subtle, sequential patterns.

Prior research guides interpretation of these findings. Past work shows that individuals are likely to provide contrasting evaluations when asked to rate different stimuli on a similar rating scale (Parducci, 1968). In addition, with longer temporal intervals the effects of previous responses on current ones tends to dissipate (Doshi et al., 2014). Our predictions are loosely drawn from prior work on SDs in categorization tasks (Zotov, et al., 2011) as well as moral judgments (Parducci, 1968), such that within-reviewer ratings may contrast with previous ratings while within-business ratings may show effects of assimilation, exploring the effects of longer stimulus intervals (n = 2, 3, etc.) on current evaluations is a relatively new approach. In this respect, the current study stands as an initial exploration into the effects of SDs *in the wild.*

We found that a reviewer's current rating deviates from their mean rating in contrast with previous ratigns. If a reviewer's previous rating was positive, their current rating is more likely to be less positive than average. This effect dissipates the farther the previous review is from the current review, an effect that replicates previous findings that show contrast when making sequential moral judgments on the same scale (Parducci, 1968). In addition, this effect was observed across time, such that successive ratings that were displaced across different temporal distances were more likely to contrast with previous ratings. Findings from this analysis suggest that the observed contrast effect may be stable across time. This warrants further exploration

We found that ratings given to the same business were more likely to assimilate to the previous rating, an effect that flattened at greater review distances. The weakness of this effect is most likely due to the nature of the dataset, such that many reviewers provide reviews to a single business, so that any effect is naturally between reviewers. Just as a participant's report of an experience is not independent of their prior experience, in social circumstances there may be a sequential dependence *across* persons. We speculate (very tentatively) that such an effect, if true, would have interesting implications for how we ought to conceptualize our own judgments as entirely independent of others' judgments.

In addition, we found a very slight assimilation effect within businesses between previous and current review ratings over time. However, this effect reversed at longer temporal distances. While initially this effect appears to be consistent Review Distance findings, the reversal is puzzling. Such an effect suggests, perhaps, that evaluations of our seemingly independent experiences are represented relative to others' previous ratings. As such, further exploration as well as experimentation is necessary to more fully understand how sequential dependencies influence natural behavior.

One possible aim for future studies is to control for the business's current rating—normalizing the reviewer's rating using the average rating for the business around the time at which the reviewer's rating is made. This would provide a more absolute difference measure, adjusted to the business's average which could be used to determine if observed SD effects can be explained by the business's current rating. In addition, one could generate an artificial baseline dataset with a fixed number of ratings per reviewer. However the type of distribution we assume when generating such data may artificially inflate our findings if it does not reflect a natural distribution we see here.

Note that the Yelp distribution is not normal, exhibiting a J-shape or bimodal distribution at 1-star and 4/5-stars with a mean of 3.75. Recent studies suggest that a J-shape bimodal distribution, unique to review data, may be the result of an underreporting bias (Hu, Zhang, & Pavlou, 2009), such that reviewers are more likely not to provide reviews when the average business rating is similar to their own experience. Determining an appropriate baseline measure will be dependent, in this case, on how we interpret the cause of the J-shape distribution. For instance, this may be dependent on the type of reviewer that is considered. Critics are more likely to have a unimodal distribution whereas non-critic reviewers tend to produce a J-shape distribution (Dellarocas & Narayan, 2006). One speculative hypothesis is that when one has a choice to write a review (e.g., non-critics), they are susceptible to influences of SDs of other's reviews, resulting in a bimodal shape compared to those who may have less of a choice to write a review (e.g., critics).

Computational models that explain how sequential dependencies emerge from the decision making process can help decontaminate current evaluations so as to obtain a more accurate measure of one's experience (e.g., Mozer, et. al, 2010). Such models, though currently only developed for low-level perceptual tasks, might be fruitfully applied to areas such as online rating systems shown to impact a business's future success (Luca, 2011). Our current work is a first step toward uncovering contamination effects that may be a rational property of the cognitive system (Qian & Aslin, 2014; Wilder et al., 2010; Yu & Cohen, 2009) within naturally occurring behavior. Developing tools that can adjust for such effects might help to provide ratings that reflect the consumer's true experience.

Conclusion

Data sets such as the Yelp, Inc. dataset are incredibly noisy. Reviewers sometimes don't review for various reasons and businesses change their names and their products adapting in real time to the demands of consumer behavior. In SD experiments the stimulus is often held constant, but this may not be the case in the real world. Restaurants go out of business, while other change drastically over time. Moreover, Yelp reviews occur over a much larger time course than sequential dependency experiments, and Yelp reviewers do not see their previous review ratings at the time when they make a new rating. If there are trends in business quality or reviewer performance over time, or adjustments to Yelp's user interface (Yelp, too, must adapt to its customers), the ability to discover echoes of cognitive effects *in the wild* may be affected.

The current work targets a broader goal of validating well-known findings from carefully controlled laboratory studies in large, unconstrained, natural and noisy behavioral data. Our aim was to determine if we can use well-known cognitive findings from controlled laboratory experiments, to sift through that noise in an effort to understand one's true experience and how that experience is affected by cognitive biases. To this end, our exploratory analysis found that current judgments, such as business review ratings, are in some way dependent on previous judgments. Reviewer's current ratings tend to be displaced from their average rating in a direction that contrasts with their previous ratings. While a business's current review rating tends to assimilate with its previous rating. Our findings, at times unpredicted and surprising, provide new avenues for future research while validating the efficacy of previous well-established laboratory findings *in the wild*.

Acknowledgments

This work was in part funded by an IBM PhD fellowship awarded to David W. Vinson for the 2015-16 academic year.

References

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning?. *Journal of Experimental Psychology: General*, *129*(2), 177.

Cantallops, S., & Salvi, F. (2014). New consumer behavior: A review of research on eWOM and hotels. *International Journal of Hospitality Management*, *36*, 41-51.

Dellarocas, C., & Narayan, R. (2006). A statistical measure of a population's propensity to engage in post-purchase online word-of-mouth. *Statistical Science*, *21*(2), 277-285.

Dixon, P, McAnsh, S., & Read, L. (2012). Repetition effects in grasping. *Canadian Journal of Experimental Psychology,* 66.

Donkin, C., Heathcote, B. R. A., & Brown, S. D. (2015). Why is accurately labelling simple magnitudes so hard? A past, present and future look at simple perceptual judgment. *The Oxford Handbook of Computational and Mathematical Psychology*.

Doshi, A., Tran, C., Wilder, M. H., Mozer, M. C., & Trivedi, M. M. (2012). Sequential dependencies in driving. *Cognitive science, 36*(5), 948-963.

Furnham, A. (1986). The robustness of the recency effect: Studies using legal evidence. *Journal of General Psychology, 113,* 351-357.

Freyd, J. J., & Finke, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 126-132.

Garner, W. R. (1953). An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology*, *46*(5), 373.

Holland, M. K., & Lockhead, G. R. (1968). Sequential effects in absolute judgments of loudness. *Perception & Psychophysics*, *3*(6), 409-414.

Hsu, S. M., & Yang, L. X. (2013). Sequential effects in facial expression categorization. *Emotion*, *13*(3), 573.

Hu, N., Zhang, J., & Pavlou, P. A. (2009). Overcoming the J-shaped distribution of product reviews. *Communications of the ACM*, *52*(10), 144-147.

Laming, D. (1984). The relativity of 'absolute' judgements. *British Journal of Mathematical and Statistical Psychology*, *37*(2), 152-183.

Liberman, A., Fischer, J., & Whitney, D. (2014). Serial dependence in the perception of faces. *Current Biology, 24*(21), 2569-2574.

Lockhead, G. R. (2004). Absolute judgments are relative: A reinterpretation of some psychophysical ideas. *Review of General Psychology*, *8*(4), 265.

Luca, M. (2011). Reviews, reputation, and revenue: The case of Yelp.com. *Harvard Business School*

Kristjánsson, Á. (2006). Simultaneous priming along multiple feature dimensions in a visual search task. *Vision research*, *46*(16), 2554-2570.

Mozer, M. C., Pashler, H., Wilder, M., Lindsey, R. V., Jones, M. C., & Jones, M. N. (2010). Decontaminating human judgments by removing sequential dependencies. In J. Laffterty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, & A. Culota (Eds.), *In Advances in Neural Information Processing Systems 23* (pp. 1705-1713). La Jolla, CA: NIPS Foundation.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on Amazon. com. *MIS quarterly*, *34*(1), 185-200.

Parducci, A. (1968). The relativism of absolute judgments. *Scientific American*.

Qian, T., & Aslin, R. N. (2014). Learning bundles of stimuli renders stimulus order as a cue, not a confound. *Proceedings of the National Academy of Sciences*, *111*(40), 14400-14405.

Stewart, N., Brown, G. D., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological review*, *112*(4), 881.

Ward, L. M. (1973). Repeated magnitude estimations with a variable standard: Sequential effects and other properties. *Perception & Psychophysics*, *13*(2), 193-200.

Ward, L. M., & Lockhead, G. R. (1971). Response system processes in absolute judgment. *Perception & Psychophysics*, *9*(1), 73-78.

Wilder, M., Jones, M., & Mozer, M. C. (2010). Sequential effects reflect parallel learning of multiple environmental regularities. In Y. Bengio, D. Schuurmans, J. Lafferty, C.K.I. Williams, & A. Culotta (Eds.), *Advances in Neural Information Processing Systems 22* (pp. 2053-2061). La Jolla, CA: NIPS Foundation.

Yu, A. J., & Cohen, J. D. (2009). Sequential effects: superstition or rational behavior?. In *Advances in neural information processing systems* (pp. 1873-1880).

Zotov, V., Jones, M. N., & Mewhort, D. (2011). Contrast and assimilation in categorization and exemplar production. *Attention, Perception, & Psychophysics 73*, 621-639.