Dwelling on the bad: Negative arguments and stimuli are given more weight in both cumulative and non-cumulative tasks

Devin M. Burns[1], Emmanuel M. Pothos[2], and Lee C. White[3]

[1]Department of Psychological Science, Missouri University of Science and Technology

[2]Department of Psychology, City University of London

[3]Organisational Behaviour, INSEAD, Singapore

Author Note

Correspondence concerning this article should be sent to Devin M. Burns, at burnsde@mst.edu

Abstract

Studies of order-effects have often been siloed into those focused on question-order effects, which examine pairs of purportedly independent items, and information-order effects, which ask participants to combine multiple pieces of information. We present data from both types of tasks demonstrating a previously unreported asymmetry, where negative stimuli have a stronger effect on subsequent positive stimuli than vice versa. Data are reanalyzed from three previously published studies of order effects, as well as two novel experiments; we observed consistent results across a variety of tasks and stimuli. These results are discussed in the context of both traditional models like Hogarth and Einhorn's belief-adjustment model and more recent attempts to use quantum probability theory to model order effects.

Dwelling on the bad: Negative arguments and stimuli are given more weight in both cumulative and non-cumulative tasks

Order effects are a fascinating subset of behavioural biases. Consider, for example, the well-known study conducted as part of an actual Gallup poll with more than 1,000 respondents (Moore, 2002). Participants were asked about whether Clinton was honest and, in a separate question, whether Gore was honest. Presented like this, the percentage ‘yes’ responses for the Clinton question was not very high – only about 50%. But, when different participants answered the same questions in the reverse order, the percentage ‘yes’ for Clinton rose to 57%. This is a shocking change in purported public opinion from what may seem like a fairly innocuous manipulation – changing the order of two questions, presented independently from each other (that is, participants were not told to consider the two questions together or given any other instructions to indicate that they should inform their answer to one question from the other). The importance of such question order effects is underwritten by the fact that Gallup polls can be an instrument to shape public opinion, especially for individuals who are undecided.

A closely related phenomenon (often formally treated in the same way) is information order effects, where multiple stimuli or pieces of evidence are presented to participants who then make a final judgment combining them. For example, Bergus et al. (1998) asked medical professionals to evaluate the probability of a disease given some information in a particular order: a lab test and a physical exam. Different participants were asked to evaluate the probability of the disease, on the basis of exactly the same information, but presented in the reverse order. Since the information is identical, the rational expectation is that final probabilities should be the same – but instead Bergus et al. found large differences! McKenzie et al. (2002) reported a similar result in the context of a jury decision making task. The important difference between this class of phenomena and those that concern question order is the cumulative nature of information processing in the former, with participants explicitly asked to combine multiple stimuli into a single judgment. Studies of question order are ostensibly non-cumulative, in that participants are presumed to answer each question independently of the other (though they often fail to do so). Of course, it is possible that the relevant cognitive biases are similar across the two cases.

Order effects have been unsurprising to social psychologists. The idea that an earlier question activates unique thoughts and perspectives, which influence the way subsequent questions are approached (Schwarz, 2007), offers a reasonable starting point to understand question order effects. Perhaps the form of this influence is in terms of increased availability of some information (Goldstein & Gigerenzer, 2002). For example, in the Gallup poll question order effect, answering the Gore question first and then the Clinton one might remind participants that Gore was known for his integrity and so, since Clinton and Gore worked closely with each other, Clinton is probably not so bad himself (of course, the converse reasoning applies to Gore too). The difficulty with this idea is that it is too general: beyond simply enabling us to anticipate question order effects, there is little else to guide us regarding their direction or size.

Hogarth and Einhorn (1992) presented one of the most complete quantitative frameworks for understanding information order effects, focusing on cumulative tasks. They considered a fairly large body of evidence and showed how their Belief-Adjustment Model was capable of predicting the direction and size of order effects under various conditions. The key mechanism in their model that produces order effects is that the amount that a belief or opinion is shifted by a piece of information is proportional to its distance from their current belief. For example, if a participant starts out with a fairly positive impression of something, an argument supporting it would not move them very much, but if they then see an argument against they could be shifted downwards substantially. If these same two arguments were presented in the reverse order, however, the negative argument would move them downwards first, allowing the positive argument to have a larger impact, thus producing a more positive final impression compared to the first order. This example demonstrates the common pattern of recency effects, where later stimuli are weighted more heavily than earlier stimuli, but their model also allows for primacy effects in other situations.

While Hogarth and Einhorn (1992) considered a wide variety of experimental conditions, all of their tasks were cumulative, with participants attempting to combine multiple pieces of information, so their model does not describe the related phenomenon of question order effects. A promising new framework that can describe both phenomena uses the probability rules from quantum mechanics to model decision making processes. Quantum theory is a set of rules for probabilistic inference, just like the more common Bayesian theory. One key difference between the two is that 'measurements' in quantum theory can change the relevant states in specific ways. For example, let us say you are trying to decide whether Clinton is honest vs. not (a binary question). The relevant state here is your mental state with all the information you have about Clinton's honesty. Prior to a decision, you might be uncertain as to whether you want to characterize Clinton as honest vs. not. On making a decision, quantum theory requires the mental state to change in a way which corresponds to the decision. Put differently, say that prior to the decision you think Clinton is honest with 40% probability and you decide that Clinton is indeed dishonest. Then, the mental state has to change so that, post decision, your belief that Clinton is dishonest is 100%. Therefore, quantum theory incorporates a fairly specific constraint for how question order effects arise: earlier questions change the mental state in the way required by quantum theory; therefore, later questions would be impacted in a corresponding way. A strength of this family of models is that the same logic can explain the presence of order effects in both cumulative and non-cumulative tasks.

Various researchers have employed quantum-like cognitive models to understand order effects (for a general overview see Pothos & Busemeyer, 2022). Wang et al. (2014) used quantum theory to derive a constraint concerning the responses in question order effects, including the aforementioned Clinton/Gore example. They called this constraint the Quantum Question (QQ) equality and they reported consistency of their prediction with empirical data, across a large number of questionnaires. Trueblood and Busemeyer (2011) modeled information order effects in cumulative tasks, including a medical decision making example (Bergus et al., 1998) and a jury decision task (McKenzie et al., 2002). They showed that a quantum model provided a better fit to the data than Hogarth and Einhorn's belief-adjustment model.

Order effects have clearly had a central role in the development of the relevant decision literature. Very broadly speaking, order effects show that earlier questions impact later ones. But an important subtlety is that this impact might be moderated depending on the history of previous judgments. For example, previous work shows that the size of such order effects depends on whether or not intermediate evaluations are provided during a series of stimuli. White et al. (2014) presented a pair of oppositely valenced stimuli such that the second one would always be evaluated, while the first stimulus was evaluated in only half of the trials. The second stimulus was supposed to be evaluated on its own (this was not a cumulative task), but the first stimulus showed carry-over effects such that the negative stimulus was rated more positively when following a positive stimulus compared to when it was rated on its own. Crucially, this effect was diminished when participants provided a rating to the first stimulus in a pair compared to when they didn't. White et al. (2020) referred to this effect as an evaluation bias, and showed it extended to more naturalistic organizational surveys as well. Burns and Hohnemann (2023) found similar results (which they referred to as a measurement effect) using a cumulative task that explicitly asked participants to combine the first and second stimuli to make a final judgment. Again, when responses were provided to the first stimulus, it had less impact on the final rating. Interestingly, the reduced salience of the first stimulus when intermediate ratings were provided was consistent across these two importantly different tasks: one where participants are trying to include the first stimulus in their final response, and the other where they aren't!

In reviewing the data from these previous experiments, we found an intriguing asymmetry in the data that has not been previously reported: negative stimuli appear to have a stronger impact on subsequent positive stimuli than vice versa. Most previous treatments of order effects have focused on comparing final responses for a stimulus sequence when shown in one order versus another, for example positive-negative compared to negative-positive. While this comparison tells us the overall size of order effect that has occurred, it cannot distinguish between multiple possibilities for where these differences emerge: for example, it could be the case that seeing a positive stimulus before a negative stimulus has zero effect on the response, and the measured order effect is produced entirely by the effect of seeing a negative stimulus before a positive. In order to disambiguate these possibilities, we need to compare these data points to a neutral case, where the stimuli in question are rated on their own. When we reanalyzed our data in this manner we found consistent evidence that negative information "sticks" more than positive and that this effect is independent of the evaluation bias we observed in our previous results.

While the evaluation bias was predicted *a priori* based on quantum models of cognition, there is nothing in those models that would predict this asymmetry. While different instantiations of quantum models are capable of showing various order effects, we have encountered no models or previous claims where the weight given to the first argument depended on whether it was positively or negatively valenced. Hogarth and Einhorn's belief adjustment model (1992) is capable of producing such an effect, but only if the negative stimuli were further away from the participant's initial opinion than the positive stimuli. In the opposite case, their model would predict that positive information would have a larger impact.

Somewhat similarly, Russo (2015) reviews a history of work on "information distortion": an effect similar to confirmation bias, where the interpretation of evidence is biased in the direction of a current leading response alternative. This proposal is like a predecisional form of cognitive dissonance (Festinger, 1957) or coherence effects, whereby preferences develop to align better with the actual choices (Ariely & Norton, 2008; Holyoak & Simon, 1999; Sharot et al., 2010). This could produce an advantage for negative stimuli if the participants started out with a preference against the decision in question, but the idea of information distortion would just as easily favor positive stimuli.

Perhaps more pertinently, a number of investigators have argued that belief formation and updating is asymmetrical between good and bad news. Apparently, we tend to pay less attention to negative information (Eil & Rao, 2011; Sharot et al., 2012), especially when the information concerns ourselves – for example, evidence that we are not as attractive as we have assumed (Koszegi, 2006). Such work is consistent with findings that negative stimuli create a withdrawal tendency (Robison et al., 2004). Regarding information order effects, such findings do suggest a putative role of emotion in information order effects. Specifically, if it turns out that there is a general bias to discount or avoid negative information in the environment, one might expect that a stimulus reflecting a negative emotion early on would have less of an impact on a subsequent stimulus reflecting a positive emotion and vice versa. Even though such a prediction extrapolates quite liberally the previous work of Eil and Rao (2011) and others, it is instructive in that it illustrates the line of reasoning which could help extract predictions for question order effects, from knowledge about the relative impact of negative and positive information.

In contrast, there is a significant body of work that supports a “negativity bias” (e.g. Cacioppo et al., 2014; Joseph et al., 2020) which refers to the idea that negative stimuli are more attention grabbing (Aue & Okon-Singer, 2015). This bias has been invoked to explain, for example, why we seem to have a preference for negative news stories. The Model of Evaluative Space (Cacioppo & Berntson, 1994) proposes that the affect system has separate positive and negative sub-systems, with distinct functions, which when combined generate an overall motivation for approach or avoidance (Norris, 2021). When inputs are equal, the negativity sub-system’s output is stronger than the positivity sub-system. It has been argued that the negativity bias is an evolutionary adaptation which provides an advantage as it emphasizes survival motivation through the avoidance of aversive stimuli (Rozin & Royzman, 2001). There is a fair amount of evidence for the existence of a negativity bias (e.g. Molins et al., 2022; Norris, 2021). It may appear that such work contrasts with that from Eil and Rao (2011) and others, though the apparent discrepancy is perhaps resolved in terms of the personal relevance of the negative information: whether it concerns ourselves or relates to personal planning vs. not. Of course, quantifying personal relevance is not straightforward and, moreover, individual differences are likely to complicate the story (Cacioppo et al., 2014).

In the present work, we focus on experiments with no personal relevance. We are interested in negative and positive stimuli, for which corresponding feelings would arise incidentally. For example, when processing a negatively valenced stimulus, e.g. a sad face from an unknown individual, we might experience fleeting feelings of sadness ourselves. Accordingly, we expect the negativity bias (Cacioppo et al., 2014) might be more relevant to establishing expectations for how stimulus emotional valence relates to information order effects: one might expect that negative stimuli will have a stronger effect on subsequent stimuli compared to positive stimuli. Overall, the present work concerns exactly this issue: is it the case that we can generalize ideas such as the negativity bias, to a broader expectation that 'negative information sticks'?

There are several reasons why such a prediction is not trivial. First, all the work regarding attention towards negative stimuli (or lack thereof) does not necessarily imply that a perception of a negative stimulus would impact on how a subsequent, ostensibly unrelated stimulus is approached, especially for non-cumulative tasks. That is, whatever processing bias operates when dealing with a negative stimulus, there is no necessity (from existing theory) that there will be a corresponding impact later on. This goes back to one of the key reasons why question order effects are so puzzling. For example, in Moore's (2002) Gallup poll, the two questions about Clinton and Gore were presented independently – it is not at all obvious that a bias for one question would affect the way we approach the other one. Corresponding explanations had to invoke social psychology ideas (e.g., Schwarz, 2007) or ideas from probabilistic modeling (e.g., White et al., 2014) to explain the apparent connectedness between the responses to the different questions, instead of models assuming that the relevant information is integrated towards a particular conclusion.

Second, in the case of sequences of questions specifically, responding to questions might diffuse any possible emotional impact. For example, in relation to emotional induction procedures, it has been found that emotion ratings can reduce the effectiveness of the procedures (Keltner et al., 1993). It is therefore possible that apparent asymmetries in the processing of positive-negative vs. negative-positive information might be due to the evaluation bias/ measurement effect from White et al. (2014) and Burns and Hohnemann (2023). Therefore, it makes sense to employ a paradigm which allows the study of order effects, in the context of evaluation biases. The paradigm employed by White et al. (2014) and Burns and Hohnemann (2023) is suitable. In this paradigm, unbeknownst to participants, stimuli were organized in pairs. In all cases, the second stimulus would be rated, while the first stimulus would be sometimes rated, sometimes not.

It should be clear from the above that there are somewhat conflicting predictions from different theoretical perspectives concerning the differential impact of positive and negative stimuli. Additional work is needed to resolve the question concerning the role of valence in order effects, which is the intention of this current work.

Reanalyses of Archival Data

Previously published data from Burns and Hohnemann (2023) and White et al. (2014) provide useful tests for the differential impact of positive and negative stimuli. Both papers were focused on the evaluation bias: the finding that when multiple stimuli are presented in series, asking participants to provide an intermediate response to the first stimulus reduces its contribution to subsequent evaluations. For example, if a negative argument is presented before a positive argument, the final judgment to both will be more positive when participants were asked to provide an intermediate response to the negative argument. Intriguingly, this effect was similar for both a cumulative task that required participants to combine the two (or more) stimuli and a non-cumulative task in which they were only asked to respond to the second stimulus independently of the first. Neither of these papers examined whether positive and negative stimuli were impacting participants to a similar degree, so here we reanalyze that data to answer this question. All data and analyses considered in this paper can be found in one place at https://osf.io/me5xu/, but individual experiments each have their own OSF pages with timestamped pre-registrations (where applicable).

Reanalysis 1

Experiment 1 from Burns and Hohnemann (2023) presented 299 MTurk participants with pairs of arguments, a positive argument for making a particular behavioural change and a negative argument against the change. Participants were asked to rate how likely an average American was to make that change in response to the arguments. There were weak and strong versions of positive (P) and negative (N) arguments for four different decision examples, so 16 total arguments were presented in 8 pairs (in either PN or NP order). The primary manipulation was that while participants always provided a final rating after considering both arguments, in half of the trials participants also provided an intermediate response after reading the first argument in the pair, giving us double- and single-rating conditions. The previous results focused on model fitting with respect to the measurement effect: the main finding was that the weight given to the first argument in the final ratings decreased when participants provided this intermediate rating. Here we reanalyze the data with respect to differences between the positive and negative arguments. This data and analysis code can be found at osf.io/j3gy8.

Results

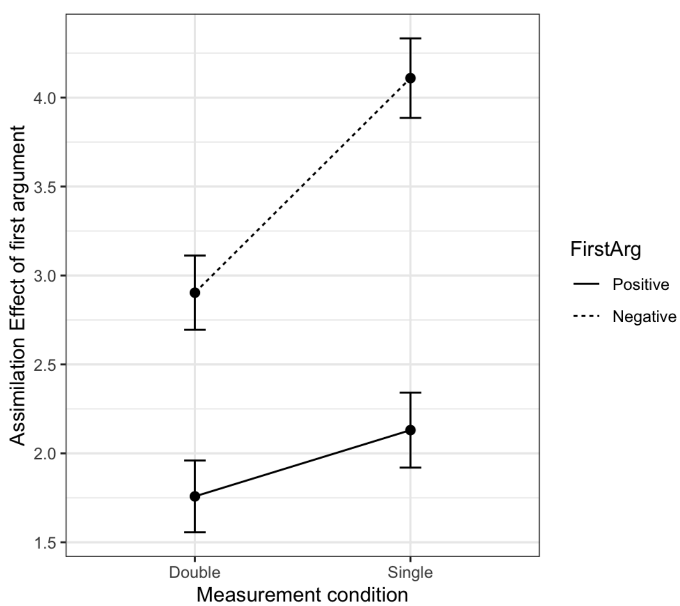
In order to examine differences between positive and negative arguments, it helps to transform the response data in a way to represent the impact that the first argument is having on the final response. Because we used contrasting arguments which naturally pull responses in opposite directions, we frame the analysis in terms of the strength of the *assimilative* effect of the first stimulus: the degree to which the first argument pulls the final response towards it. Instead of assuming that the stimuli we used were all of equal salience, we can quantify the effect of each individual argument using the intermediate responses that participants provided in the double measurement condition after reading the argument on its own. We then compute the difference between this baseline and the conditions when this argument was presented as the second in a pair of arguments (either in the single or double measurement conditions). This measures the effect that the first stimulus had on the final combined judgment, by comparing it when that second stimulus was rated on its own.

Because negative initial arguments should pull the final response down while positive initial arguments pull it up, we change the sign of the difference depending on argument order: for PN pairs, the assimilative effect was computed by subtracting intermediate responses for the negative argument (i.e. when it appeared as the first argument in different trials) from the final responses (i.e. when this argument was preceded by a positive argument) – if there is assimilation, this would be positive. For the opposite order, final responses to the NP pair were subtracted from intermediate responses to the positive argument. Thus, positive values of the difference imply the first argument is having an assimilation effect, regardless of order.

A factorial ANOVA tested whether the assimilative effect of the first stimulus was different depending on which argument came first (N or P), the measurement condition (single or double), and a potential interaction between the two. There was a main effect of argument order, with trials where the negative argument came first showing significantly higher assimilation, $F(1, 2388)=63.96$, $p<.001$, $\eta^2=.03$. The main effect of measurement condition was smaller, but also significant, with double-measurement trials showing increased assimilation, as shown in previous analyses, $F(1, 2388)=16.56$, $p<.001$, $\eta^2=.01$. Finally, there was also a significant interaction $F(1, 2388)=4.55$, $p=.03$, $\eta^2=.002$, such that the gap between negative and positive was larger in the single-measurement condition, as shown in Figure 1.

Figure 1

Assimilation Effect of First Argument on Final Ratings



A potential alternate explanation for this pattern (pointed out by a helpful reviewer) is that despite our attempts to balance the stimuli, perhaps the negative arguments were just stronger than the positive ones, in which case they may have larger assimilative effects for that reason alone. Our first inclination was to compare the average intermediate response to negative stimuli, -2.28, to that for positive stimuli, which was 4.76. While this provides some evidence against the possibility that the negative arguments were stronger, this does not necessarily imply that the negative stimuli were indeed weaker, because we cannot assume that participants were starting from an initial neutral point of zero. Rather than attempting to definitively show equal salience across stimuli, we can instead present an alternate analysis that affirms the "negative information sticks" hypothesis without relying on balanced stimuli.

The important difference between these two potential explanations is that under our hypothesis, when negative and positive arguments are combined, the resulting decision gives greater weight to the negative stimulus *than would be expected based on its individual salience*. We can expect to find this effect, even without precise quantification of individual stimulus salience, by comparing responses averaged across all stimuli for the intermediate rating compared to the final rating. Even if the negative stimuli were stronger, this alone should not cause these two averages to differ: every stimulus appears equally often in the first and second positions, so stronger negative stimuli should equally affect each average. In contrast, if negative information sticks more than positive when arguments are combined, this will produce a lower average for final responses compared to the intermediate response average. This latter pattern is what we find, with final responses being on average .59 lower in the double measurement condition compared to intermediate responses, $t(2376)=2.68$, $p=.007$, $d=.11$. The drop was even larger for the single measurement condition, a full 1.00 point, $t(2383)=4.50$, $p<.001$, $d=.18$. Note that our previous analysis focused on the assimilation metric (Figure 1) is a more specific illustration of this difference in averages: if positive assimilation effects matched those for negative assimilation, these response averages would be equal.

Discussion

This analysis shows that in this experimental context, negative arguments had a larger impact on the final combined judgments than the positive arguments. In short, negative information sticks. The recency effects that we previously reported for this data can now be described in more detail: the first argument has the least impact (creating strong recency) when it is a positive argument that participants issue an intermediate response to. Recency decreases when the first argument is negative and when intermediate responses are withheld.

Our comparison of averaged responses provides support that this effect is not just due to higher salience of negative arguments overall, but rather that these arguments receive disproportionate weight when combined with positive arguments.

Reanalysis 2

The second experiment from Burns and Hohnemann (2023) used the same set of 16 arguments, but presented them in blocks of four, instead of pairs. Again, the critical manipulation was whether participants responded incrementally after each argument (quadruple-measurement) or only after reading all four arguments (single-measurement). Previous results from this experiment also showed that stimuli had less effect on the final combined rating when participants provided these intermediate ratings, but here we examine asymmetric effects of argument valence. This data and analysis code can be found at osf.io/j3gy8.

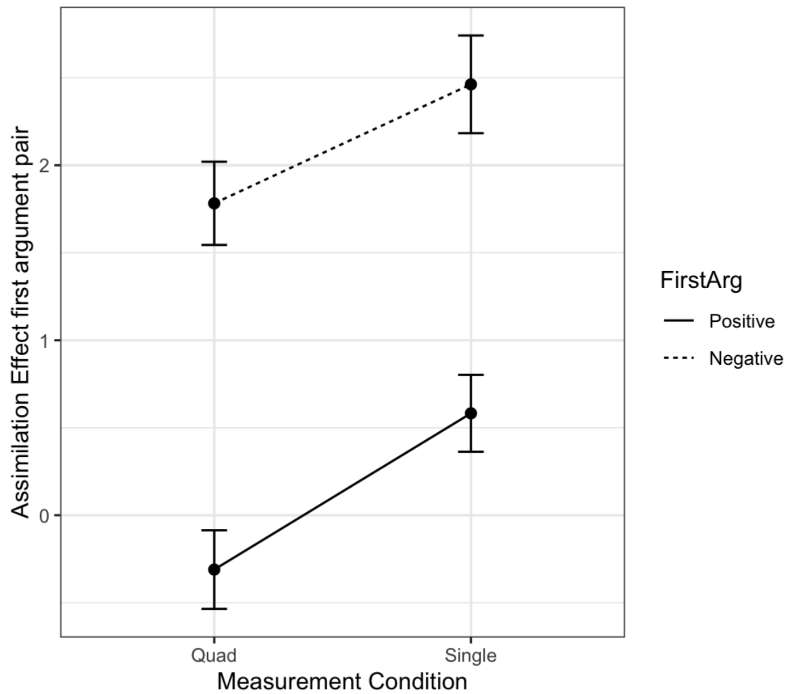
Results

Because this experiment used sequences of four arguments (two positive and two negative), picking the appropriate comparison to test whether negative information sticks is more complicated. The most obvious difference between negative and positive information can be seen by focusing on the subset of trials when the arguments were blocked by valence, so that two positive arguments were followed by two negative arguments or vice versa. Assimilation effects were computed similarly to Experiment 1, except this time for pairs of arguments. We did this by averaging intermediate responses to each unique pair of arguments in the quadruple measurement condition (after seeing the first two arguments) as a baseline measure of reactions to that argument pair (when presented in that order). We then computed the assimilative effect as the difference between this baseline and the final responses when this pair was presented as the third and fourth arguments (either in the single or quadruple measurement conditions). We again changed the sign of the difference depending on order, such that positive values always mean assimilation.

A factorial ANOVA was used to test whether the amount of assimilation was different depending on which pair of arguments came first (PPNN vs. NNPP), the measurement condition (single or quadruple), and a potential interaction between the two, with results shown in Figure 2. There was a significant main effect of argument valence, with trials where the negative arguments came first showing significantly higher assimilation, $F(1, 1672)=68.48$, $p<.001$, $\eta^2=.04$. The main effect of measurement condition was smaller, but also significant, with single measurement trials showing increased assimilation, $F(1, 1672)=10.74$, $p=.001$, $\eta^2=.01$. This time there was no significant interaction, $F(1, 1672)=0.20$, $p=.66$, $\eta^2<.001$.

Figure 2

Assimilation Effect of First Argument Pair on Final Ratings (for Blocked Stimuli)

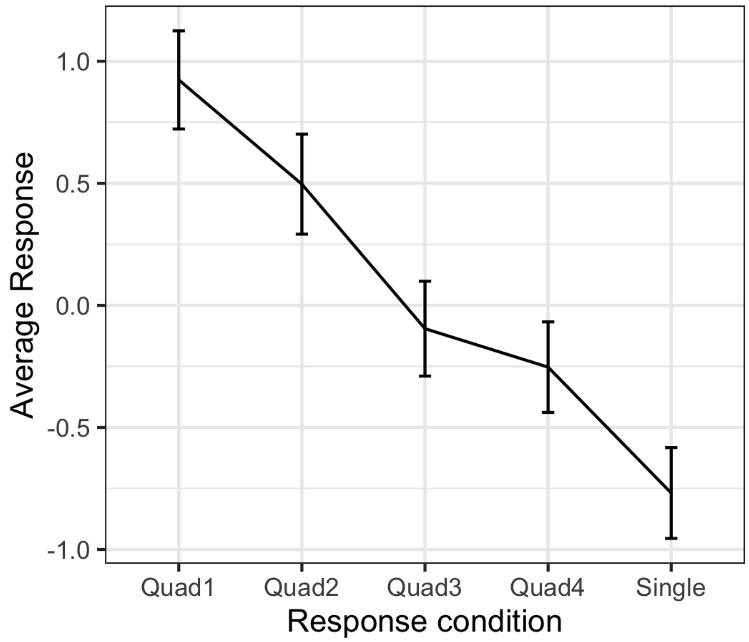


For the half of the trials where argument valence was interleaved instead of blocked, we can also use data from the first two arguments as a replication of the result from Experiment 1 (but only for the quadruple measurement condition, since there were no trials where participants responded after the second stimulus but not after the first). The main effect of argument valence was again significant and of similar size, with negative arguments causing 1.34 more points of assimilation, $F(1, 416)=10.03$, $p=.002$, $\eta^2=.02$.

As with the data from the first reanalysis, we can also examine average responses after, 1, 2, 3, or all 4 arguments in order to control for potential asymmetries in argument strength. As can be seen in Figure 3, response averages decrease by 0.41 points for each additional stimulus, $F(1, 4188)=56.47$, $p<.001$, $\eta^2=.01$.

Figure 3

Response averages get lower with more stimuli (the single measurement condition is also after seeing all four stimuli)



Discussion

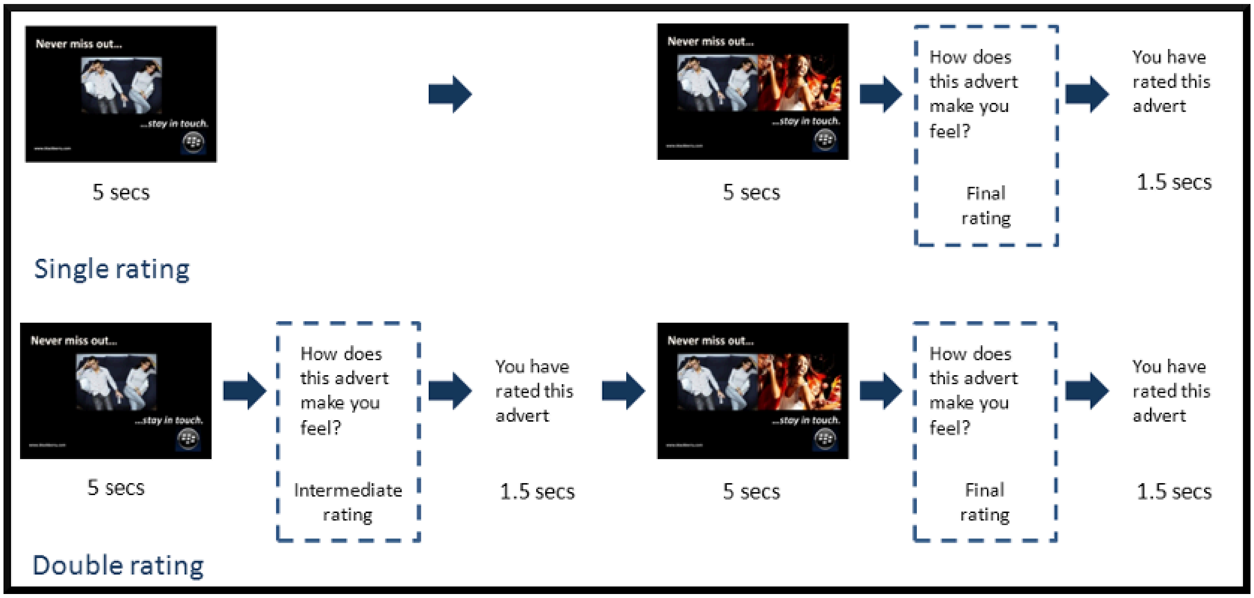
Two different analyses using longer sequences of stimuli showed the same effects as seen in the first experiment: negative stimuli had a larger impact on ratings than positive stimuli, that is, negative information sticks. In fact, assimilation effects were close to zero when positive arguments came first, indicating that the negative arguments were rated similarly whether they were presented on their own or preceded by the pair of positive arguments. There was no significant interaction with measurement condition in these data. We also showed that response averages drop consistently as more stimuli are included, indicating that greater weight is being given to the negative stimuli.

Reanalysis 3

In Experiment 1 from White, Pothos, and Busemeyer (2014), 54 participants rated six pairs of adverts which had positive and negative valence. The first advert in the pair contained a single image of positive (or negative) valence and then the image in this first advert was paired with another one of opposite valence, as in Figure 4. All advert pairs were designed so that having two images of opposite valence together would make sense. The six adverts corresponding to the study design were randomly presented together with 24 filler adverts.

Figure 4

Design of Experiment 1 from White et al. (2014)



Note. Permission to reproduce will be obtained after acceptance.

Participants were asked to rate how different adverts made them feel. Participants always provided a final rating after considering both adverts (that is, a rating for the combined impression of the two images in a pair). Additionally, in half of the trials participants also provided an intermediate rating after viewing the first advert. The primary manipulation in the study corresponded to whether participants provided both an intermediate and a final rating vs. just a final rating, giving us single- and double-rating conditions. As noted above, the analyses in the original work focused on the evaluation bias, with the main finding being that in the double rating condition, the rating of the second advert was more negative in the PN condition and more positive in the NP condition. Here we reanalyze the data with respect to differences between the positive and negative arguments, as in Reanalyses 1 and 2.

Results

In order to examine differences between positive and negative arguments, the data was transformed using the same method as that used for Reanalysis 1:

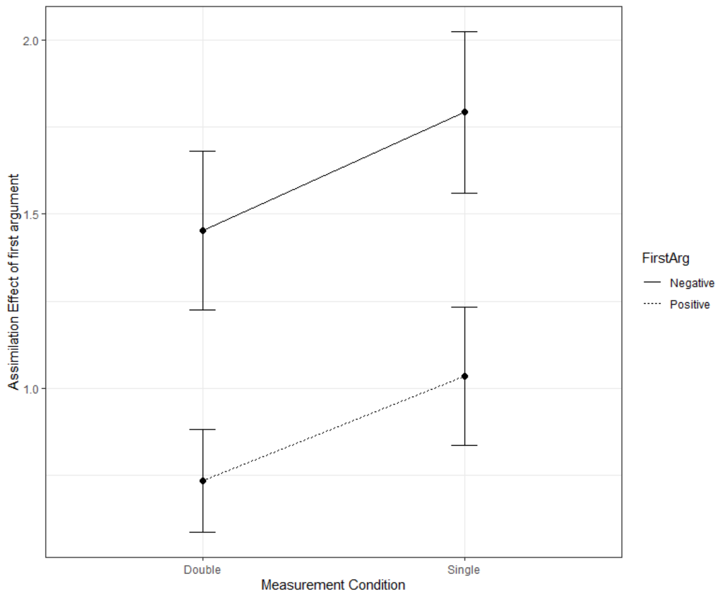
For PN pairs, Assimilation effect = rating for N stimulus when presented second minus rating for N stimulus when first

For NP pairs, Assimilation effect = rating for P stimulus when presented first minus P stimulus when second.

A factorial ANOVA tested whether the assimilation effect of the first stimulus was different depending on which argument came first (N or P, argument order), the measurement condition (single or double), and a potential interaction between the two. There was a significant main effect of argument order, with trials where the negative argument came first showing higher assimilation ($F(1, 196)=17.44$, $p<.001$, $\eta^2=.08$) as shown in Figure 5. There was no main effect of measurement condition nor was the interaction significant.

Figure 5

Assimilation effect in White et al. (2014) Experiment 1.



As for Reanalysis 1, to check whether negative stimuli were simply stronger in salience, we conducted the same comparison of responses averaged across all stimuli for the intermediate rating compared to the final rating for the single and double rated conditions. If the negative stimuli were stronger, we would still expect these two averages to be similar. In contrast, if negative information sticks more than positive when arguments are combined, this will produce a lower average for final responses compared to the intermediate response average. This second outcome is what we find, with final responses being on average .36 lower in the double measurement condition compared to intermediate responses, $t(49)=3.77$, $p<.001$, $d=.46$. The drop was similar for the single measurement condition, on average .38 lower in the single measurement condition compared to intermediate responses, $t(49)=2.97$, $p<.001$, $d=.38$.

Discussion

This analysis shows that in this experimental context, negative adverts had a larger impact on the final combined affective evaluations of the mixed adverts, than positive adverts.

This is the same pattern of results seen in the first two reanalyses, despite the change in format and task. As in the other analyses, comparison of the averaged responses suggests that the effect is not just due to more salient negative stimuli overall, but rather that these stimuli receive disproportionate weight when combined with positive arguments.

Reanalysis 4

Experiment 2 from White, Pothos, and Busemeyer (2014) used the same design shown in Figure 4, except that the stimuli were amended so that each advert always contained one single image of positive or negative valence, regardless of whether the advert was the first or second in a pair. Importantly, contrary to the three studies discussed so far, this change meant that this was a non-cumulative task where participants responded to each advert one at a time. While the two stimuli in a pair both were for the same product (smartphone or insurance), there were no instructions to integrate reactions across stimuli, just respond to "How does this advert make you feel?"

Twenty participants rated six pairs of PN adverts and six pairs of NP adverts. The primary manipulation was the same as that described in Reanalysis 3 and the result in the original study was also the same, with the main finding being that in the double rating condition, the rating of the second advert was more negative in the PN condition and more positive in the NP condition. Here we reanalyze the data with respect to differences between the positive and negative arguments regarding the assimilation effect.

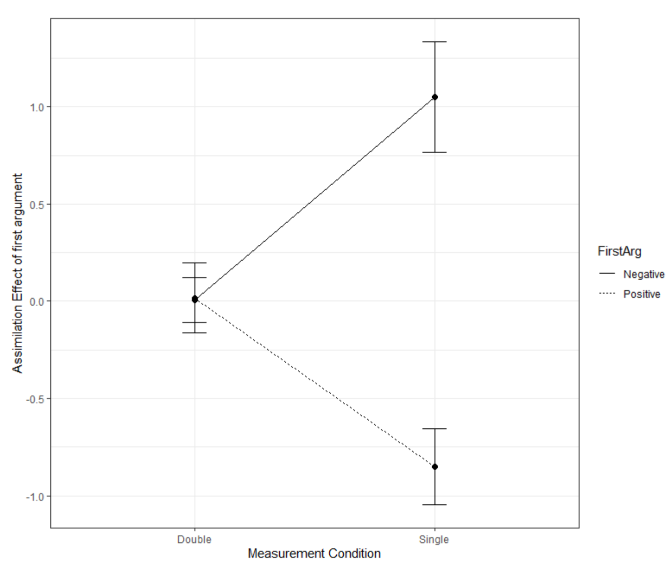
Results

In order to examine differences between positive and negative stimuli, a dependent variable corresponding to the assimilation effect was computed using the same method as that used for the previous reanalysis.

A factorial ANOVA tested whether the assimilative effect of the first stimulus was different depending on which advert came first (N or P), the measurement condition (single or double), and a potential interaction between the two. There was a significant main effect of advert order, with trials where the negative advert was presented first showing significantly higher assimilation (F(1, 76)=28.72, p<.001, η^2=.21). There was no main effect of measurement condition, but there was a significant interaction (F(1, 76)=29.22, p<.001, η^2=.22), such that the gap between negative and positive first adverts on the assimilation effect was large in the single-measurement condition, but non-existent in the double-measurement condition, as shown in Figure 6.

Figure 6

Assimilation effect in White et al. (2014) Experiment 2.



We conducted the same check on whether negative stimuli were simply stronger in salience, by comparing responses averaged across all stimuli for the intermediate rating compared to the final rating for the single and double rated conditions. As before, if the negative stimuli were more salient, we would still expect these two averages to be similar. In contrast, if negative information sticks more than positive when arguments are combined, this will produce a lower average for final responses compared to the intermediate response average. In this analysis we found no difference between responses in the double measurement condition and intermediate condition ($t(19)=-0.05$, $p=.95$) and similarly for the single measurement condition ($t(19)=1.76$, $p=.095$).

Discussion

Even with a non-cumulative task, where participants are not required to base their evaluation of a second stimulus on the evaluation of the preceding stimulus, there is evidence that the valence of an initial stimulus has a higher impact on a later one when this valence is negative compared positive; that is, negative information sticks. However, these results were different from the previous ones in several regards: there was no assimilation for either positive or negative stimuli in the double measurement condition, and a reverse assimilation (contrast) effect for positive stimuli in the single measurement condition. It may be worth noting that this experiment had only 20 participants, and therefore is more likely to return anomalous results than our other, more highly-powered experiments.

Experiments

Experiment 1

Our final reanalysis, Experiment 2 from White et al. (2014), had several potentially important differences from the previous studies: our first three all used a cumulative design where participants were purposefully combining the two pieces of evidence, which should naturally produce larger assimilative effects of the first stimulus on the second compared to the non-cumulative design if this final study. Additionally, our first two experiments focused on the evaluation of arguments for specific behaviours, while the latter concerned the affective impression of advertisements. To more fully explore the relevant conditions, we created a new experiment using arguments similar to Burns and Hohnemann (2023), but in a non-cumulative design where participants only responded to single arguments. We also instructed different sets of participants to focus on more logical or emotional reactions, to see if this manipulation might moderate the assimilative effect we observed in the previous work.

All materials, data, code, and pre-registration are available through the Open Science Foundation at https://osf.io/594sc/

Participants

We recruited 200 participants from Mechanical Turk. Of these, 10 had attention scores less than three out of four and were excluded, leaving 190 participants for analysis. These were evenly split between instruction conditions, with 95 focusing on logical appraisals and 95 on emotional reactions.

Procedure

In contrast to Burns and Hohnemann (2023), there was no manipulation of argument strength in this experiment: for each example there was just one argument for making a green behaviour change and one argument against. These arguments were largely similar to the ones used in the previous experiments, but modified to equate their strength.

Participants were presented with eight different pairs of arguments. Each pair had one positive and one negative argument and we balanced whether the negative or positive argument was presented first. Participants were instructed to respond to the arguments one at a time, rather than weigh them against each other. We used page breaks in the (online) survey to help ensure participants focused on a single argument at a time. To provide an example, participants were told "Joe's community is considering changing their approach to trash. He is presented with the following two arguments:" The next page had the argument that "This behaviour will lead to a large reduction in landfill use and greenhouse gas release." In the double measurement condition, participants would be asked at that point to rate how Joe would feel about this argument using a Likert scale from -10 ("definitely negative") to +10 ("definitely positive).

In the single measurement condition, participants would proceed to the next argument without providing a response. The second argument in this example was that "This behaviour is unpleasant for most residents and will take them extra time every day." Regardless of measurement condition, participants were asked the same question, rating Joe's reaction to this second argument.

Participants were presented with a random ordering of four such pairs using the single measurement condition and four different pairs with double measurements. They then responded to all eight pairs again in a new random order, with measurement condition swapped so that every participant responded to every pair of arguments once in each measurement condition. This allowed us to quantify evaluation biases on a within-participant basis, to help ensure that any such effects would not arise as averaging biases. Additionally, including the manipulation of observing some information without a decision, provides a test of the negative information sticks idea, in relation to the possibility that decisions attenuate emotional impact (Keltner et al., 1993).

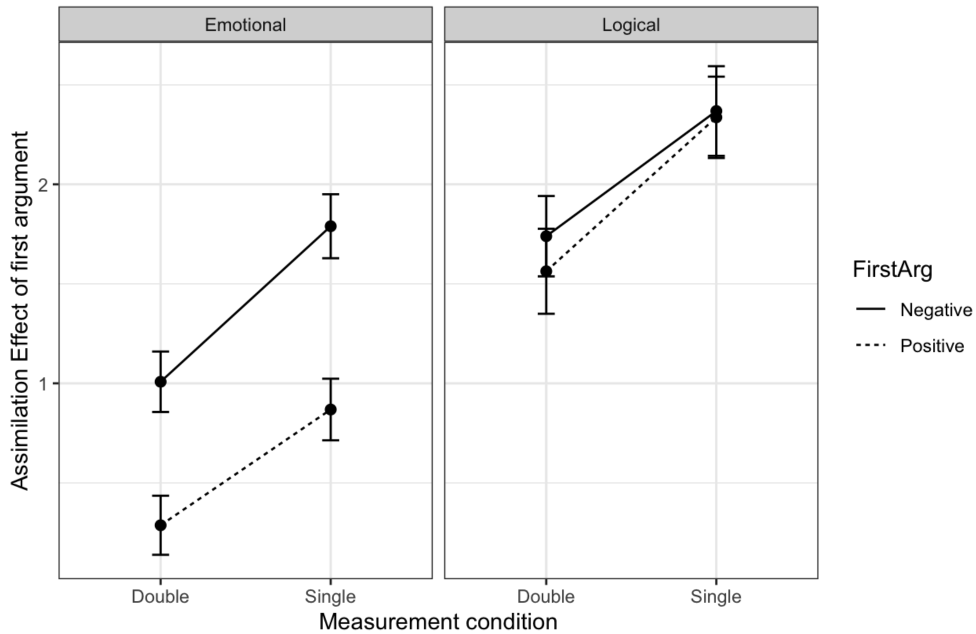
Finally, there was a between-participant manipulation of response focus, with half of the participants asked to focus on emotional reactions concerning how Joe might feel about the arguments. The other half were asked to focus on logical reaction and assess the likelihood that Joe would adopt the new behaviour.

Results

Assimilation effects were computed as before to quantify the effect the first stimulus in a pair had on the second. As seen in Figure 7, a factorial ANOVA showed a main effect of argument order, with negative arguments having stronger assimilation effects on subsequent positive arguments than vice versa, $F(1, 755)=7.08$, $p=.008$, $\eta^2=.01$. There was also a significant effect of measurement condition, with less assimilation when participants issued an intermediate response to the first argument in the double measurement condition $F(1, 755)=15.83$, $p<.001$, $\eta^2=.02$. Finally, there was a significant effect of participant instructions, with higher assimilation overall in the logical response version, $F(1, 755)=34.03$, $p<.001$, $\eta^2=.04$). There was also an interaction between instructions and valence, $F(1, 755)=4.26$, $p=.04$, $\eta^2=.006$), such that there was little to no difference between negative and positive stimuli for participants issuing responses based on logical reactions.

Figure 7

Assimilation Effects Separated by Response Condition



Comparisons of response averages showed a small but insignificant difference between intermediate and final responses in the emotion-focused condition, *t*(1496)=1.49, *p*=.13, *d*=.08. There was almost zero difference in the logical reaction condition, *t*(1484)=0.05, *p*=.96, *d*<.01.

Discussion

First, it is interesting to note that overall assimilation effects of the first stimulus are still within the general range of those found in our reanalysis of Burns and Hohnemann (2023), despite changing the task to a non-cumulative one. The original experiment had asked participants to combine the two arguments into a single judgment, which should result in the first stimulus having an assimilative effect on the final response to the pair. In contrast, in the present experiment we explicitly told participants to only rate the second argument on its own, discouraging any impact of the first argument. The previous assimilation effects were larger, between 1.5 and 4 points on our 21-point scale, compared to the .5 to 2.5 range (same scale) seen here, but given the difference in tasks, we think it is interesting that there is a comparable effect at all. The effect of measurement (single vs double rating condition) was similar in the current and previous experiment, with single measurement trials in the original experiment yielding .80 points (η^2=.01) more assimilation than double measurement, and .75 points (η^2=.02) in the new data. This consistency is also surprising, given that the previous study compared evaluation biases between participants, while in this study participants were responding to the exact same pairs of arguments only a few minutes later. Such effects appear to be robust and important.

The novel finding we are focused on here is the difference in assimilation effects between positive and negative stimuli. The data from our previous experiment showed that negative stimuli produced an assimilation effect on final ratings 1.5 points (η^2=.04) stronger than positive stimuli. The new data shows that participants who were asked to focus on emotional responses showed a very similar effect, with a difference of 1 point (η^2=.04). However, the participants who focused on the logical evaluation of the arguments showed no effect of argument valence! Note that the Burns and Hohnemann (2023) experiment asked participants to predict how Joe would respond to the arguments, without explicitly focusing their attention on logical or emotional reactions. Interestingly, this may suggest that in the absence of more specific instruments, the default responses were more to the emotional side. Equally suggestive, these results may imply that participants are capable of reducing or eliminating the negative information sticks bias by attempting to focus on logical reactions!

We should also note that the evaluation bias was consistent for both response instructions and both positive and negative arguments, suggesting that this effect is separable from the negative information sticks effect, which was only present for the affective version of the task.

Experiment 2

In the first new experiment, although participants were instructed to respond to the second argument on its own, it is possible that the task structure of having a pair of arguments both relevant to a single decision scenario encouraged participants to weigh the positive and negative arguments against each other, making assimilation effects more likely. To address this concern, our second experiment also used a non-cumulative task, but with stimuli which had no clear pairing. Instead, participants were asked to rate the tone of 24 different emails.

Unbeknownst to the participants, the order of the emails was carefully structured such that every other email was ambiguous in tone, and could be interpreted either positively or negatively without additional context. These ambiguous emails were preceded by emails that were either positive, negative, or neutral to see if we would find the same pattern of effects we have seen in our other experiments: assimilation from preceding stimuli on subsequent ratings, which decreases when ratings were provided to the previous emails (double measurement condition) compared to when they were not (single measurement), and which is larger when the previous stimulus was negative instead of positive.

All materials, data, code, and preregistration are available through the Open Science Foundation at osf.io/bdftj/.

Participants

A total of 240 participants were recruited from Amazon Mechanical Turk. Participants were required to have their MTurk Masters qualification (not related to educational attainment), live in the United States, and have an approval rating greater than 85% with more than 50 approved tasks. The average completion time was 12 minutes and participants were compensated with $1.33 for their time. All participants completed an informed consent document prior to participation. Preregistered exclusion criteria dropped seven participants for scoring less than five out of six on attention check questions, leaving 233 participants who were 58.8% male (n = 137), 40.8% female (n = 95), and .04% declining to respond (n = 1).

Procedure

Participants read a total of twelve email pairs, each pair consisting of a priming email with a clear tone and an ambiguous email that could be interpreted positively or negatively (e.g. ``Good afternoon, please meet me in my office at the end of the day to discuss your recent work."). This structure was concealed from participants: they just read 24 emails one at a time. There were four primes of each potential valence: positive (e.g. ``I have been very impressed with your work since you have joined our organization - it has been exceptional! A few of the higher ups are going to dinner tomorrow night and they asked me to bring along my best employee. Would you like to join us?"), neutral (e.g. ``The annual state of the organization address is scheduled for Friday at 3 pm in the auditorium. The address will be recorded and made digitally available by Monday morning for those who cannot attend. -Office of the Executive"), and negative (e.g. ``I have heard many complaints about your performance and am shocked at your audacity to work this way! Please come to my office at 3 pm on Monday."). The gender of the person sending the email, as indicated by the header and signature line, was balanced across tone and held consistent within a pair. Neutral prime emails were sent from an institutional address (e.g. parking management) and were therefore genderless. Measurement condition was also manipulated within participants: in double measurement trials participants were first asked to rate the tone of the prime email and then rate the tone of the ambiguous email, while in single measurement trials no rating was given for the primes. All responses were measured using a single question, "How would you rate the tone of this email?" with a scale from -3 (extremely negative) to 3 (extremely positive).

The conditions were presented in the same pseudo-random order for all participants, though whether a given ambiguous email was preceded by a positive, neutral, or negative prime was balanced across three versions of the experiment (between participants). Within each version, the four pairs of emails of a given prime tone were assigned to each potential combination of author gender and measurement condition for different groups of participants.

Results

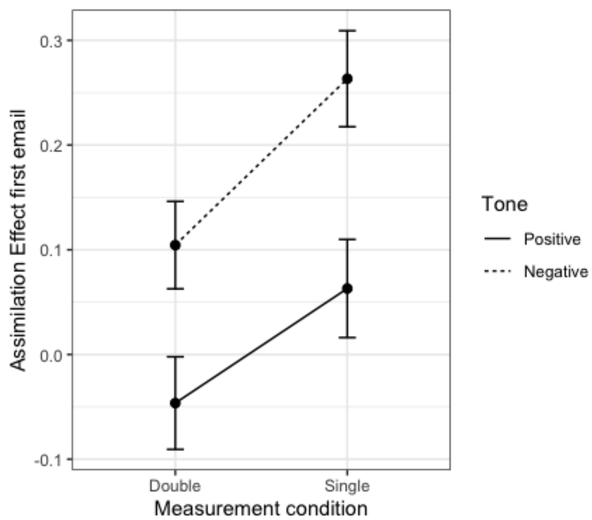
First of all, we analyzed responses to the priming emails. Negative primes were rated on average as m=-2.58 (sd=.63), positive primes 2.76 (sd=.56), and neutral primes 0.21 (sd=.62), indicating that our oppositely valenced primes were of similar strengths.

We computed the outcome variable of assimilation effect similar to our previous experiments, but using trials with neutral primes as a baseline. For each ambiguous email, we computed the average response when it was preceded by a neutral email, and subtracted this from responses when it was preceded by a positive email. We did the same for negative primes, but flipped the sign, subtracting the negative responses from the neutral average. This meant that across all primes, positive scores indicate assimilative effects.

As seen in Figure 8, a factorial ANOVA showed a main effect of prime valence, with negative emails having stronger assimilation effects, $F(1,1860)=18.14$, $p<.001$, $\eta^2=.01$. There was also a significant effect of measurement condition, with less assimilation when participants issued a response to prime email in the double measurement condition $F(1,1860)=10.59$, $p=.001$, $\eta^2=.006$. There was no significant interaction effect, $F(1,1860)=0.36$, $p=.55$, $\eta^2<.001$.

Figure 8

Assimilation Effects of Priming Email on Ratings of the Ambiguous Email



Discussion

Although effect sizes were somewhat smaller for this experiment than the others presented here, the three primary effects of interest remained consistent. Even when participants were blind to the fact that the stimuli were arranged in pairs for the purpose of analysis and even though they were instructed to respond to a single stimulus on its own, we still found assimilation effects such that the ambiguous emails tended to be rated as having a tone more similar to the email that preceded them. This assimilation effect was reliably smaller when participants were asked to rate the previous email compared to conditions when they merely read it and then moved on to the next email (Keltner et al., 1993). As regards the main focus of this study, negative emails produced larger assimilation than their positive counterparts.

Conclusions

Across a variety of experimental methodologies, we have found surprisingly consistent results that when negative stimuli were followed by positive stimuli, they had a greater impact on subsequent judgements compared to the effect of positive stimuli followed by negative. Amongst the many previous examinations of order effects in the literature, we have not encountered previous work focusing on this asymmetry depending on information valence. Reanalyses 1, 2, and 3 showed this effect in cumulative tasks using sequences of two or four arguments. Reanalysis 4, along with two novel experiments presented, showed a similar effect for non-cumulative tasks, when participants were asked to only respond to the most recent stimulus (though experiment 1 did not find this effect for the logical-reaction condition).

Overall, we have presented convergent evidence across several studies that negative information sticks. Our results are likely related to other reports which offer evidence that negative information is prioritized in cognitive processing (Aue & Okon-Singer, 2015; Cacioppo et al., 2014), but are importantly different in that we find negative evidence being given higher weight than would be expected based on its salience. The implications for formal models for question order effects (and evaluation biases), such as the ones from Hogarth and Einhorn (1992) and White et al. (2014), are that such models are, at best, incomplete. Although the belief-adjustment model has free-parameters allowing for a differential impact from negative vs. positive information, Hogarth and Einhorn did not report any hypotheses about which information would be more salient under which conditions. It is intriguing to ask whether it might be possible to identify ways to understand the effect of negative information sticking within such models, but it is unclear how this might be possible (note, invariably, there are efforts to understand similar puzzling differences in mechanistic terms, so there might be an avenue concerning positive vs. negative information too, e.g., Leek & Pothos, 2001).

The consideration of the evaluation bias in this work was primarily motivated from work such as that of Keltner et al. (1993), showing that a judgment can diffuse the emotional impact from a stimulus. Using a paradigm similar to the one we adopted here, previous experiments (Burns & Hohnemann, 2023; White et al., 2014) have found a similar consistency of evaluation biases across both types of tasks, with the impact of the first stimulus reliably diminished when participants provide an intermediate rating to it instead of immediately proceeding to the second stimulus. The consistency of the evaluation bias across so many diverse studies attests to its robustness. Note, the specific evaluation bias prediction in the initial White et al. (2014) was motivated by quantum theory. But, as just noted, the corresponding quantum model assumes that positive and negative stimuli would impact participants in similar ways. Our new analyses here show that in many of these datasets, this assumption does not hold, and negative stimuli have reliably stronger impacts across both single and double measurement conditions.

The only exception to this trend was for the participants in our first experiment who were asked to focus on the logical evaluation of the arguments. These participants showed no difference in the amount of assimilation produced by positive or negative arguments. Notably, evaluation biases were consistent across these different instructions, and equal for positive and negative stimuli. This suggests that the presence of evaluation biases is separable from these novel stimulus valence effects, and both impact the degree to which previous stimuli have a carry-over effect on later responses. Further investigation is warranted to verify whether similar instructions can reliably diminish these effects, especially within different experimental paradigms.

Do the present results conflict with those from Eil and Rao (2011) and Sharot et al. (2012), that participants selectively ignore negative information in favor of positive information? One critical difference between these studies and the present one is personal relevance. For the present results (and the ones from Burns and Hohnemann, 2023; White et al., 2014) personal relevance was between very limited and non-existent. Even when the judgments concerned action that the participants could imagine taking (either directly or as if in the shoes of a hypothetical participant), the negative information had no implications regarding effort or cost for participants. By contrast, in Eil and Rao (2011) and Sharot et al. (2012), the negative information was directly personally relevant. There are ideas here, perhaps relating to ego threat (Baumeister & Boden, 1998), which can account for why negativity and personal relevance interact like this, but it remains a matter of speculation as to what are the specific relevant processes.

A related avenue for future work is to clarify the way the negative information sticks effect is realized. For example, it could be that negative information (that is not personally relevant) is more memorable or more attention grabbing. For example, there is evidence that participants have better memory for specific visual details of negatively coded objects compared to neutral (Kensinger et al., 2006) – threat relevance theory postulates that humans have evolved to prioritize attention towards threatening stimuli (Ohman & Mineka, 2001). Such a difference in memorability could conceivably allow negative evidence to have a greater impact on subsequent responses even if the negative stimuli were not rated as stronger on their own. We think it is this feature of the negative information sticks bias which makes it interesting, its robustness across a range of manipulations. Relatedly, a fruitful avenue for future research would be to better understand the conditions under which one would expect to see more focus on negative information compared to positive information, beyond the use of judgments (Keltner et al., 1993) and manipulating personal relevance. Future studies could examine individual differences in this regard, and how they may relate to differences in emotion regulation strategies (Kobylińska & Kusev, 2019).

References

Ariely, D., & Norton, M.I. (2008). How actions create – not just reveal – preferences. *Trends in Cognitive Sciences, 12*, 13-16.

Aue T., Okon-Singer H. (2015). Expectancy biases in fear and anxiety and their link to biases in attention. *Clinical Psychology Review, 42,* 83-95.

Baumeister, R. F., & Boden, J. M. (1998). Aggression and the self: High self-esteem, low self-control, and ego threat. In *Human aggression* (pp. 111-137). Academic Press.

Bergus, G. R., Chapman, G. B., Levy, B. T., Ely, J. W., & Oppliger, R. A. (1998). Clinical diagnosis and the order of information. *Medical Decision Making: An International Journal of the Society for Medical Decision Making*, *18*(4), 412–417.

Burns, D. M., & Hohnemann, C. (2023). Measurement effects in decision‐making. *Journal of Behavioral Decision Making*, *36*(3).

Cacioppo, J., Cacioppo, S., & Gollan, J. (2014). The negativity bias: Conceptualization, quantification, and individual differences. Behavioral and Brain Sciences, 37(3), 309-310.

Eil, D. & Rao, J. M. (2011). The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself. *American Economic Journal: Microeconomics, 3*, 114–38.

Festinger, L. (1957). *A Theory of Cognitive Dissonance*. Stanford Univ. Press, Stanford.

Fredrickson, B. L., & Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of personality and social psychology*, *65*(1), 45.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75–90.

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

Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General, 128*(1), 3-31.

Keltner, D., Ellsworth, P.C., & Edwards, K. (1993). Beyond simple pessimism: Effect of sadness and anger on social perception. *Journal of Personality and Social Psychology, 64*(5), 740–752

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2006). Memory for specific visual details can be enhanced by negative arousing content. *Journal of Memory and Language*, 54, 99–112.

Kobylińska, D., & Kusev, P. (2019). Flexible emotion regulation: How situational demands and individual differences influence the effectiveness of regulatory strategies. *Frontiers in psychology*, *10*, 72.

Koszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economics Association, 4,* 673–707.

Leek, E. C. & Pothos, E. M. (2001). Open peer commentary. What is specific about category-specificity? Fractionating patterns of impairments and the spurious living/nonliving dichotomy. *Behavioral and Brain Sciences, 24*, 487-488.

Lieder, F., Griffiths, T. L., & Hsu, M. (2018). Overrepresentation of extreme events in decision making reflects rational use of cognitive resources. *Psychological Review, 125*(1), 1-32.

McKenzie, C. R. M., Lee, S. M., & Chen, K. K. (2002). When negative evidence increases confidence: change in belief after hearing two sides of a dispute. *Journal of Behavioral Decision Making*, *15*(1), 1–18.

Moore. (2002). Measuring New Types of Question-Order Effects: Additive and Subtractive.

Public Opinion Quarterly, *66*(1), 80–91.

Ohman, A. & Mineka, S. (2001). Fears, phobias, and preparedness: toward an evolved module of fear and fear of learning. *Psychological Review, 108*, 483-522.

Pothos, E. M. & Busemeyer, J. M. (2022). Quantum cognition. *Annual Review of Psychology*, 73, 749-778.

Robinson, D. T., Smith-Lovin, L., & Rogalin, C. (2004). Physiological measures of theoretical concepts: Some ideas for linking deflection and emotion to physical responses during interaction. *Advances in Group Processes, 21*, 77–115.

Ross, M. & Wilson, A. E. (2003). Autobiographical memory and conceptions of self: getting better all the time. *Current Directions in Psychological Science, 12*, 66-69.

Russo, J. E. (2015). The predecisional distortion of information. In E. A. Wilhelms & V. F. Reyna (Eds.), *Neuroeconomics, judgment, and decision making* (pp. 91–110). Psychology Press

Schwarz, N. (2007). Attitude Construction: Evaluation in Context. *Social Cognition*, *25*(5), 638–656.

Seta, J. J., Crisson, J. E., Seta, C. E., & Wang, M. A. (1989). Task performance and perceptions of anxiety: Averaging and summation in an evaluative setting. *Journal of Personality and Social Psychology*, *56*(3), 387.

Seta, J. J., Haire, A., & Seta, C. E. (2008). Averaging and summation: Positivity and choice as a function of the number and affective intensity of life events. *Journal of Experimental Social Psychology*, *44*(2), 173-186.

Seta, J. J., Haire, A., & Seta, C. E. (2008). Choices and affective reactions to negative life events: An averaging/summation analysis. *Judgment and Decision Making*, *3*(5), 425-434.

Shafir, E., Simonson, I., & Tversky, A. (1994). Reason-based choice. In *Reasoning and decision making*, P. N. Johnson-Laird & E. Shafir (Eds), Oxford: Blackwell.

Sharot, T., Velasquez, C. M., & Dolan, R. J. (2010). Do decisions shape preference?: Evidence from blind choice. *Psychological Science, 21*(9) 1231–1235.

Sharot, T., Kanai, R., Marston, D., Korn, C. W., Rees, G., & Dolan, R. J. (2012). Selectively altering belief formation in the human brain. *Proceedings of the National Academy of Sciences, 109*, 17058-17062.

Trueblood, J. S., & Busemeyer, J. R. (2011). A quantum probability account of order effects in inference. *Cognitive Science*, *35*(8), 1518–1552.

Wang, Z., Solloway, T., Shiffrin, R. M., & Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(26), 9431–9436.

White, L. C., Pothos, E. M., & Busemeyer, J. R. (2014). Sometimes it does hurt to ask: the constructive role of articulating impressions. *Cognition*, *133*(1), 48–64.

White, L. C., Pothos, E. M., & Jarrett, M. (2020). The cost of asking: How evaluations bias subsequent judgments. *Decisions*, *7*(4), 259–286.