Response to Editor and Reviewers

1) Reviewer 2 and I had questions about the generality of the conclusions that can be drawn from a single experiment in which the decision is fairly difficult. Would the PLBA model account for the data in the same way if the task were easier? Or, as Reviewer 2 asks, if the motion change at the switch point was in terms of motion coherence (or perhaps speed) rather than direction? Put differently, non-stationary information can be defined in many ways. You empirically tested and fit only one of those possibilities, suggesting that your conclusions should be similarly restricted.

We agree that conclusions based on the experiment are restricted, and have tried to be clearer about this. However, we have also pointed out that the PLBA provides a framework that could in principle account for tasks that differed in difficulty from ours or that used different types of change (such as Reviewer 2's example where the magnitude but not the sign of the stimulus change), as such differences would be accommodated by changes in the size and sign of rate estimates. We believe these restrictions are also made clearer by the changes elaborating the theoretical contribution of our work (see response to the next two points).

2) Both reviewers point to other papers and models that deserve discussion. Specifically, Reviewer 1 suggests that a piecewise drift-diffusion model might reach similar conclusions about your data and asks for some discussion. Reviewer 2 makes a similar point about alternative accounts in the form of inhibition and leaky accumulators.

These two concerns, taken together, lead to a bigger question: what is the theoretical advance provided by this work? To my eye, much of the focus was on describing the empirical results in terms of the new PLBA model. Do the alternatives accounts proposed by reviewers reach the same basic account of these data, or are there key differences?

We believe the key difference is that the PLBA, for the first time, provides a practical framework that allows theorists to explore a variety of plausible potential mechanisms that could be at work when decisions are made about changing information, including possibly delayed changes in the rates and thresholds as well as carryover effects that make the presentation of the changed information non-veridical. Although these mechanisms could in principle be instantiated in the piecewise (linear) drift-diffusion model (PLDDM) or LCA they have not, and with good reason; their addition makes these models very difficult to work with, to the degree that a thorough exploration of their behavior, let alone fitting to data, is extremely difficult. This, in turn, makes it difficult to know whether the alternative accounts proposed by the reviewers provide the same basic account of the data. We have now been explicit on this point. For the PLDDM we have made some explorations of the explicit suggestions made by Reviewer 1 by fitting the PLBA to data simulated from the PLDDM (see response below).

The current research needs to be better situated in the literature, without expanding the text too much (look to eliminate the redundancies). The primary criterion for publication in Cognitive Psychology is that a clear theoretical advance is provided. It seems likely that you can make that case in a revision, but it isn't there now.

We have tried to be clearer on the theoretical advance, which we believe has two main aspects.
1) Examination of five plausible potential mechanisms that could be at work when decisions are made about changing information: immediate or delayed changes in the rates or thresholds, and carryover effects causing a non-veridical representation of the changed information.
2) The provision of a tractable framework (the PLBA) to thoroughly explore these mechanisms in data from experimental paradigms through Bayesian model selection applied to hierarchical models.
These latter point is emphasized in the title and abstract of the paper and at the beginning of the

general discussion.

3) Finally, we come to the modeling itself. Both reviewers want some additional reassurance that the model is doing a good job fitting, but not over-fitting, the data. Reviewer 1 appreciates your attention to the issue of parameter identifiability, but asks that you go one step further and run a parameter recovery analysis with data simulated from the best-fitting PLBA variant.

>>

An extensive parameter recovery study had been run and is reported in a separate paper, which we provide a link to (see response to Reviewer 1 for details).

Reviewer 2 speculates about exactly how the model fits the RT quantile data in stationary trials, and asks to see those data and predictions together.

>>

We have provided a new figure showing that our model fits the patterns of RT quartile data for both stationary and switch trials. (see our response to Reviewer 2 for details).

He also suggests a variant of your null model that involves a random resampling of the drift rates at the switch point. These all seem like good suggestions.

>>

We included the additional null model; see response to Reviewer 2 for details. In brief, this model was rejected.

Reviewer #1

The article is well written, addresses a timely issue, and uses rigorous and thoughtful methodology. Many of my initial concerns upon reading the manuscript (e.g., that the dot motion task might behave differently than stationary perceptual stimuli that change) were either addressed outright or at least mentioned in the discussion. I think this will make a meaningful contribution to the literature with some minor changes and improvements. Comments are below:

We thank the reviewer for their interest, enthusiasm, and comments. Below we detail the additions and adjustments that have been made in response to these comments.

-- in the discussion of conflict models with time-varying drift rates, the work of Hubner et al. should be cited in addition to the White et. al paper

Done

-- the model section is very thorough and complete, but might be hard to digest for many readers. Specifically, the labeling conventions are difficult to keep track of. I had to keep going back to keep track of the numerous labels (C1, C2, RL, LR, uv1,v2,w1,w2, etc.). This is perhaps a necessary evil for this complex modeling endeavor, but the average reader will likely gloss over these details rather than take the considerable effort needed to disentangle the different variables. I am not sure if there is a more effective system for this problem, but I would just friendly remind the authors that no reader of this work will be as intimately familiar with these terms as the authors are, and hope they can attempt to make this section as reader-friendly as possible.

We have tried to make things more reader friendly by:

1) Replacing C1 and C2 by "pre-switch correct" and "post-switch correct", but still define the terms so they can be used as compact labels in Figure 1 (but have also defined them in the caption).
2) Completely removing the use of LR and RL (or L and R)

3) We have retained use of v1,v2,w1,w2, etc. in equations, and on the few occasions they appear in text we have given their verbal definition.

.

-- it would be interesting to get the authors' take on what, if anything, would change if these data were modeled with a piecewise drift-diffusion model instead of the LBA. For example, the noise in accumulation inherent in the DDM might account for some of the delay in the effect of stimulus change on drift rate. That is, the stochastic accumulation in the DDM could suggest that the drift rate changes more quickly after stimulus change than the LBA shows, but has a slower effect because of the noisy accumulation (that is not present for the LBA).

Also, how would the relative, rather than independent, accumulation for the two responses in the DDM affect the interpretations? My intuition is that a DDM account would naturally predict a slower effect of later stimulus evidence because 1) the process would be nearer to one boundary pre-switch and thus have to travel further to get to the opposite boundary after the switch, and 2) the noise in accumulation would slow that process (relative to the ballistic accumulation in the PLBA). Some discussion of this would be nice.

>>

Unfortunately we were unable to fit the piecewise (linear) DDM (PLDDM) to our data with the extensions required to address this question (i.e., allowing for a delay in the rate change, and also possibly an asymmetry in the rate before and after) because of issues with mathematical tractability. Instead, we simulated data from a PLDDM with no delay and symmetric rates and fit the PLBA to see whether it estimated a non-zero delay and/or rate asymmetry. We found that the PLBA indicated on a minor delay, and if anything a rate asymmetry slightly in the opposite direction to what we observed in our data. Hence, it does not appear that the effects intuited to the reviewer are problematic for our conclusions.

-- Parameter identifiability is a serious concern with these complicated decision models, and the authors did a fine job addressing it through the parameter correlation analysis. I think this could be taken one step further by simulating data from the best-fitting PLBA and performing a parameter recovery to assess whether the model can accurately recover the ground truth. As it stands, we have no "ground truth" to test the model in the current manuscript.

We absolutely agree. Complex models will always have issues with parameter correlations and being underdetermined by the data. It was not done in this manuscript, but we have done extensive parameter recovery for the PLBA. It turns out that there is quite a lot to discuss on this front however. This would substantially extend the length of the paper and detract from the results since significantly more methods discussion would be required. As such, a second paper is being published where the fitting method used here is discussed in great detail. We now cite this paper. For the reviewers information a copy of the most up to date version can be found at ????. The bottom line is that we are confident that parameter recovery is not problematic for the present paper.

Reviewer #2: I found the paper appealing in many ways, especially regarding the technicalities of the analyses and the Bayesian model fitting. Below I outline some concerns, which I believe could be addressed by the authors in a revision.

Major concerns

1) The authors report that accuracy improves as a function of RT quantile in stationary trials. In other words that more delayed responses result in higher performance in the stationary trials (c.f. Figure 2). Is this pattern captured by the LBA and (by extension) by the PLBA model? One would assume that conditioned on longer RT, the drift rate of the stronger accumulator would have smaller (than the mean of the sampling distribution) values. As a result, in trials that ended with a slow response, the difference between the two drifts will be smaller and more frequently reversed in sign, resulting in lower accuracy. Can the model capture the accuracy improvement as a function of RT? Towards addressing this point it would be helpful to show the predictions of the LBA and/ or the fits of the PLBA model on the quantile data of Fig. 2.

In stationary trials, there is an initial improvement in accuracy very early in the trial. In Figure 2a, the difference in accuracy between q1 and q2 is significant. However, there is no significant change in accuracy after q2. Rather, there is a decrease in accuracy late in the trial. This can be seen in Figure 2b. We've added new analyses on page 13 clarifying the relationship between accuracy and RT quartiles for stationary trials.

As discussed in Brown and Heathcote (2008) the LBA can produce both slow errors (when errors are caused by rate variability, the pattern described by the reviewer) and fast errors (when errors are caused by start-point noise dominate). In our data, we see both fast and slow errors. We have added new panels to Figure 4, that show the PLBA model predictions side by side with the quartile data. Panel 4e shows the quartiles for the stationary trials (note that this is different than Figure 2 where stationary trials were grouped based on block switch time to aid comparison with switch trials). The figure shows a small increase in accuracy from q1 to q2, but also a drop in accuracy from q2 to q4. The PLBA does a good job in capturing the drop in accuracy from q2-q4, but has difficulty capturing the small increase from q1 to q2. We've pointed out this weakness of the model fit on page 22. Panel 4f, compares the PLBA and quartile data for switch trials after the switch time (similar to Figure 2c). As the figure shows, the PLBA does a very good job of capturing the increase in accuracy in the last quartile of RT. Overall, we feel that PLBA does a good job capturing the relationships between accuracy and RT quartiles that we see in our data.

2) Additionally, the non-linear effect in Fig. 2c (accuracy difference between stationary and switch trials for the longer quantile only) could be explained without a delay in encoding. A long trial will be associated with lower drifts in the "before switch" interval and also lower accumulated differences between the two alternatives making the choice more prone to a reversal in the "post-switch" period where the new drifts are resampled afresh. I suggest an extension of the null model, where there is no rate encoding delay or threshold changes but just resampling of the drifts at the switch point. It would be useful to know whether and how this parsimonious model fails before considering more complex extensions. I think the null model currently used is a "straw man" involving no changes in the drift rates.

This is a nice idea. We have fit the suggested variant of the null model (which we called Model 00) to data where drifts are redrawn after the change of information, but from the pre-switch distributions. It does perform slightly better than the current null model (based on DIC), which is just the LBA without any drift updating. However all models except model 2s are selected over this model variant and there is still a significant gap between this version of the null and the best performing models. A brief description of this variant has been added to the description of Model 0. It is now referred to as Model 00 and the DIC value for this model is reported in the comparison table.

Secondary concerns/ suggestions

1) The motivation structure and the instructions that subjects followed are unclear. The initial training associated feedback with the dominant direction of motion (higher accumulated evidence). Accordingly, if subjects aimed at answering on the basis of the total evidence up to their response, then responding according to the "switch" direction would be very rarely correct (unless the subject waited for very long). Thus preparing a response on the basis of early evidence seems to be normatively justified, as opposed to observing and accumulating evidence with equal vigilance across the whole trial. Additionally, subjects having the impression that they should respond on the basis of total evidence should use a detected change in evidence as a response cue (so as to avoid a reversal in the total evidence). Please provide more details about the instructions as well as whether there were participants that seemed to respond on the basis of the current motion direction (rather than based on the average direction across the whole trial).

Participants only received feedback in the first 20 trials and instructions at no point suggested that they should pick the average direction of motion across the trial; they were simply told to decide if the dots were moving mostly to the left or right. We have added a new paragraph (spanning p.10-11) clarifying instructions. Although we cannot rule out that some participants adopted a strategy to close off sampling from the display early for whatever reason (instructions or otherwise) we do not think this fits with the pattern of the data. If this were the case then it is not clear why a difference would emerge for q4, if sampling were closed off early then there is no chance for contrary evidence to build up. Further, if such an effect were related to instructions one would expect a relationship to awareness of a change, but we did not find that.

2) Understandably the PLBA model has superior mathematical/ computational tractability at the expense of some simplifying assumptions. One particular assumption, that the drifts are encoded explicitly, as if subjects are omniscient, appears too strong especially given the relatively poor performance in the "change detection" task (and the lack of correlation with detection performance and performance in the switch trials).

We disagree that having drift rates change as a (perhaps delayed) function of the stimulus is in some way omniscient. All models we are aware of assume that rates are a function of the objective stimulus and that is all we have assumed (e.g., higher rate for the accumulator matching than mismatching the stimulus, which changes at some point after the stimulus changes, but perhaps at a delay and perhaps not in a veridical manner due to carry-over effects, such as the depletion mechanism we suggested). The poor performance is a separate issue; that emerges from noise in the system. Again this is assumed by all evidence accumulation models.

Drift-rate is unlikely to be represented explicitly. Rather the instantaneous motion direction is encoded in MT neurons and read-out "bottom-up" by evidence accumulation neurons (which of course would introduce within-trial noise that the LBA framework avoids). If explicit drift representation is not plausible then conclusions cannot be drawn safely at the "process", mechanistic level.

We think the statement “of course” here is far too strong. The perceptual system is well known to have low pass filter characteristics, which must smooth any within-trial noise in the stimulus to some degree. Further, MT neurons also do not represent instantaneous motion direction in any strong sense (to our knowledge); they have finite time constants that are sufficiently large that their response is far from instantaneous. The output from MT and other motion sensitive areas may also be further smoothed and spatially aggregated before in provided input to

decision circuits (after all the required decision is about global motion not motion of individual dots). Finally, the argument against an explicit encoding of rates seems to deny the widely held view that visual short-term memory (VSTM) mechanism could provide the appropriate stationary input for evidence accumulation (e.g., Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. Psychological Review, 116, 283–317.). Recounting these points seems discursive to the point of the present paper, as they are really a criticism of the LBA and any model that assumes input from a VSTM, but we would be happy to do so if the reviewer and editor deem it appropriate.

The conclusion that contradicting evidence takes time to be encoded (and that is encoded more strongly) can have a mechanistic/ neural interpretation or just a descriptive one. I would feel more comfortable with the latter given that the drift rate encoding assumption is itself an abstraction. However, in some parts of the manuscript PLBA is presented as offering mechanistic insights; and in other parts alternative neural/ computational mechanisms (inhibition, leak, adaptation) that could mimic MLBA are discussed. Please clarify the scope of the PLBA model.

As discussed above we believe a mechanistic interpretation is quite defensible and so have not changed the paper in this regard. In scientific theories there is always a level of abstraction (e.g., LCA claims neural plausibility but operates with levels of leakage that are not realistic for individual neurons, as pointed out by Wang and colleagues, and so clearly is an abstraction at an unspecified neural ensemble level) and science is replete with rigorous and useful mathematical formalizations that operate at different levels (e.g., diffusion of gas molecules at the micro level and the Gas Law at a more macro level). We believe the scope of the LBA and PLBA is clear and appropriate for exactly what we applied it to in this paper; a fine-grained account of response choices and response time. Again we would be happy to add these statements to the paper if required but believe these points are of such a general and well known nature that to do so would risk being discursive.

3) As mentioned above the authors refer to potential neural mechanisms (inhibition, leak, adaptation) that could explain their results. In fact the basic result agrees with a previous finding and corresponding mechanistic explanation, which is currently not discussed. In Tsetsos, Gao, McClelland, and Usher (2012) a subset of trials in Experiments 2A-B involved non-stationary evidence using an interrogation protocol: during the first half of the trial the dots moved towards one direction that switched midway to the other direction. Note that there were individual differences in the task but one participant showed a pattern whereby the first half of the trial determined responses in short trials but the second half in long trials. This transition from primacy to recency as trial duration increases is illustrated in Fig. 5 together with an explanation using the non-linear LCA with balanced leak and inhibition. During the first half, the dominant alternative is likely to suppress the weak alternative to zero activity. However, as evidence switches the activity of the previously weak alternative will start increasing. The rate of increase will be small in the beginning due to inhibition from the previously dominant alternative (whose activation was at ceiling the moment of switch). However, with the passage of time the currently strong alternative will start suppressing its rival and recover a stronger rate of accumulation. I see this regime of LCA compatible with the findings in the manuscript: switch in the evidence is slowly encoded (due to inhibition from the rival that is already high) with the accumulation rate increasing with the passage of time. I think discussing this previous result in light of the current findings would add strength to the paper.

We had already cited this paper and discussed the idea raised by the reviewer. Specifically we stated “Delay is most consistent with a competitive mechanism, whereby accumulated evidence

in the decision process suppresses new contradictory information, so it takes an extended period of time before that information has an impact." We have now added to this statement a citation to Tsetsos et al. (2012).

4) Relatedly to the above one could run the best-fitting PLBA variant with certain parameterizations for hypothetical "interrogation" trials that involve a midway switch, assuming that decisions would be determined by the balance of accumulated evidence at the interrogation moment. This could reveal a similar pattern as in Tsetsos et. al (2012) (Figure 5), with a primacy in short trials (due to t_rate) and recency in long trials (where drift asymmetry washes out the influence of t_rate). This demonstration would provide an alternative explanation to the duration/ primacy-recency interaction extending the explanatory power of PLBA.

We are a little unsure what is being said here as the term "interrogation" is never used in Tsetsos et al. (2012) but we will assume that means "a midway switch", which is what was done in our experiment. Assuming that, we are unsure of why we need to run a simulation. Our data show a pattern of primacy in short trials and recency in long trials (i.e., the change we found with q1 .. q4 division) and our model fits the data, so we already know that it can produce this pattern.

Instead, in the discussion where we added the citation to Tsetsos et al. (2012) we note how LCA with the right level of inhibition dominance can produce a pattern of transition of primacy to recency dominance and point to their Figure 5.

5) Please provide the False Alarm rates in the change detection task. It is conceivable that in the presence of exogenous (dots motion) and endogenous (attention) fluctuations people would mistakenly think that a switch happened. Also, was the timing of false alarms centered on the expected switch time (did people anticipate a switch?)?

The false alarm rate for the change detection task has been added on page 14. The average false alarm rate was 42% and false alarms occurred on average 850 ms after the switch time for the change detection task. So, people did not anticipate a switch.

6) The model currently assumes that the drifts change when a switch in the direction of motion takes place. However, one can imagine non-stationary trials where the motion coherence increases in the same direction instead of flipping. The current experiment/ analyses do not offer information about these cases and it is hard to predict whether the conclusions of the paper would generalize there. Throughout the paper non-stationary information is equated to categorically changing information (as in the switch trials). It would be useful to clarify that the paper focuses on this specific type of non-stationary information.

There is nothing in principle that would stop the PLBA from with dealing a change in motion coherence in the same direction, as this would just be a magnitude rather than sign change in drift rates. In response to the editor's comments we added material on the theoretical advances provided by our paper, and in that context we have clarified that the PLBA is a general modelling framework that can be applied to a wide range of change tasks, enabling tests of other paradigms such as the suggested magnitude-change case in future work.

7) In the best-fitting PLBA variant, the rate delay could be compensated by the drift-difference asymmetry. Does t_rate and asymmetry size correlate positively? The

correlations of Table 2 could be performed as partial correlations (including the t_rate / asymmetry correlation). Also, given t_rate, asymmetry size and each participant's threshold one could derive how much (on average) the drift-rate change delay is under or over-compensated by the asymmetry.

Response: The delay and asymmetry size do not correlate at all. The correlation level is 0.011 with a p-value of 0.95. This is not surprising though. The delay shows a relatively low level if any individual variation and does not correlate with any performance measure in Table 2. Thus there is no evidence of any compensatory relationship between the rate delay and asymmetry size. We now report this pattern in the General Discussion.

Minor typos

8) Figure 6. Row 7 is labeled " t_delay" instead of t_rate.

Response: Corrected

9) P.11, 2nd paragraph: Alternately-> alternatively.

Response: Corrected.

10) P. 24, 1st row: affect-> effect

Response: We could not find this typo.