Degradation and death by natural causes in working memory

Daryl Fougnie, Jordan W. Suchow, and George A. Alvarez

Nearly all complex activities require *working memory*, the purposeful storage of information over a short interval. The capacity of working memory can be quantified either by the number of objects that can be held (Cowan; 2001; Luck & Vogel, 1997; Miller, 1956; Zhang & Luck, 2008) or by an information limit more broadly (Bays & Husain, 2008; Fougnie, et al., 2010; Wilken & Ma, 2004). These approaches to quantifying the capacity of working memory treat it as static, but in reality, working memory unfolds over time, with maintenance and degradation pulling in opposite directions (Gold et al., 2005; Magnussen, 2000; Phillips, 1974; Sperling, 1960; Vergauwe, Barrouillet, Camos, 2009; Yang, 1999). To understand the limitations of working memory, then, it is not enough to consider only its static behavior — the dynamics are of primary concern. Considering working memory as a process affords the explanatory power needed to understand why the fidelity of our memories worsen over time.

One account of changes to working memory representations was given by the *Sudden Death* model of Zhang & Luck (2009). According to the model, memory failures are dramatic events, with individual memories dying suddenly, rather than gradually losing their fidelity until they are no longer of any use. Evidence for sudden death came from a technique that allows for separate estimation of the quality and existence of a memory. Zhang & Luck (2009) measured these properties of memory for colored circles, held in mind for 1 to 10 s. The errors at 10 s showed more random guesses than those at 1 s, which suggests that information was lost over time. However, estimates of the quality of remembered items barely changed. The authors concluded that memory failures are sudden and complete.

Here, we challenge the conclusion that memory representations do not experience gradual degradation over time. We show that the pattern of results thought to be the signature of sudden death is consistent another account: degradation and death by natural causes. Recent research has shown that there is variability in the quality of our working memories (Fougnie et al., 2012; van den Berg, et al., 2012), even within trials (Fougnie et al., 2012; Suchow, Fougnie & Alvarez, under review). It is thus important to consider which memories are most likely to die off, because it can have a dramatic effect on the measured fidelity of memory. The Sudden Death model, in contrast, assumes that memories are of equal quality and that it thus cannot matter which of them fail.

Consider the following analogy. Everyone on earth ages one year per year — it could not be any other way. But the average age of these individuals (excluding births) would not increase by a year. Why is this? Mortality removes individuals from the group selectively—older individuals are more likely to die off. The expected age of living individuals increases by less than a year each year.

These processes of variation plus selective removal may also be at play in memory degradation. Importantly, selective removal of poorly remembered items and memory degradation produce opposite effects on the measured quality of memory — degradation worsens quality, selective removal improves it. If low-quality memories are more likely to be forgotten, then memories can gradually decay yet leave no trace in the form of a decline in the average. In the extreme, where the lowest-quality memories are forgotten more rapidly than the gradual degradation of them all, we might even see an *increase* in the measured quality of memory over time.

When only small (or no) changes in memory quality are observed, one must be careful not to conclude that there was no degradation over time. Selective loss of low-quality memories may have masked decreases in quality. One can only measure memories that still exist. Suppose, for example, that a person remembers three items with precisions 1, 3, and 5 (though the unit of measure is arbitrary, higher values correspond to more precise memories). On average, our measurements would yield an estimated precision of 3. Now suppose the item with the lowest precision (1) is lost, dying suddenly, leaving only two items (3, 5) that contribute to our estimate of quality. Selective removal increases the average quality to 4. But selective removal can play out differently when paired with a decrease in quality. Suppose that, over time, each item lost unit of precision, falling to 0, 2, and 4, respectively. Now, the loss of the lowest quality item (0) leads to a measured average that is equal to the value before the decrease in quality (3). Each item degraded, but the average remained the same.

We propose a new model: Degradation and Death by Natural Causes (DDNC). We use the term *degradation* to refer to a loss of memory quality over time and are agnostic as to its source: decay, interference, or some other source. "Death by natural causes" refers to the idea that the process of degradation leads to the selective removal of low-precision memories, similar to the tendency of nature to remove the oldest individuals of a population at higher rates. The model proposes that information about items is lost gradually over time. Death, rather then being a special or sudden state, is merely the end result of having lost all information about an item. To preview our results, this model better predicts the changes in the quality and existence of memories over time than variants of the Sudden Death model.

Experiment 1 – Changes in memory over time

The goal of Experiment 1 is to observe how memory for a simple visual feature, color, changes over time. Participants were shown three colorful circles for a short duration and were instructed to remember the color of each item (Fougnie et al., 2012; Wilken & Ma, 2004; Zhang & Luck, 2008; 2009; Figure 1). After a retention interval of either 1 or 10 s, participants' memory was tested for a single, random item by highlighting a location. Participants were asked to select the color of the item that was at that location by choosing one of many colors presented on a circular color wheel. This estimation task (Wilken & Ma, 2004) provides a continuous measure of performance that is useful for differentiating different sources of error (Zhang & Luck, 2008).

Methods

Twelve participants (8 female, 4 male) between the ages of 18 and 25 (mean age 20), drawn form the Harvard community, participated for course credit or monetary compensation. Participants completed two 1.5-hour sessions on two separate days, no more than a week apart. Participants were awarded a $10 bonus for completing both sessions.
Participants were asked to remember the color of three circles (0.5° in radius) that were evenly spaced along an invisible circle (3.5° in radius) centered on a fixation cross. There were 180

colors evenly distributed in a circle cut out from the CIE $L^*a^*b^*$ color space (centered at L=54, a=18, b=-8, with a radius of 59).

The timeline of each trial (Figure 1) began with the brief presentation of colorful circles for 600 ms. A blank screen (except for the fixation cross) of 1 s or 10 s appeared after the stimulus. The response screen appeared after the retention interval and remained visible until a response was made. The screen consisted of a solid white circle (0.5° radius) at the (randomly selected) tested location and hollow circles at the untested locations. A circular color wheel (6° radius) surrounded the item display and had a black selection bar on the outside of the wheel. The position of the selection bar matched the angular position of the mouse with respect to fixation. When participants moved the mouse, the selection bar moved to indicate the currently selected color. In addition, the tested item's color would update to the chosen color once the mouse was moved. This was to provide an additional cue of the selected color. Participants were instructed to click the mouse when they had selected the color that matched the item at that location during stimulus presentation. The response screen remained until participants made a response. A 1 s interval separated trials. The two retention-interval conditions were randomly intermixed within blocks of trials. Participants completed 300 trials for each retention condition, 600 trials in total.

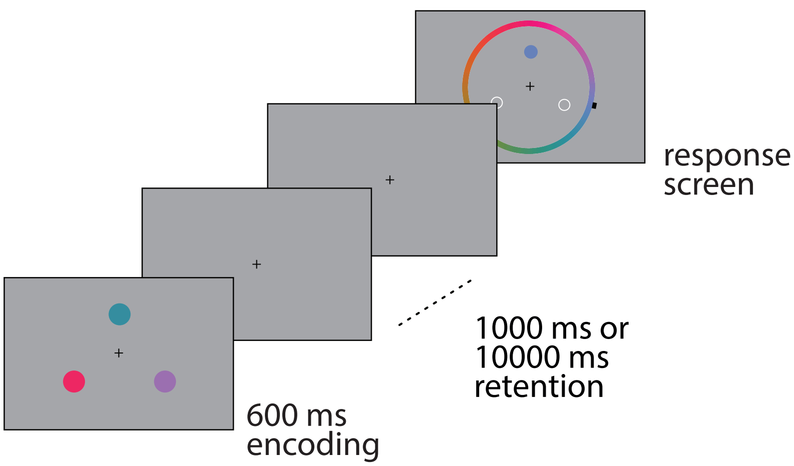

Figure 1: Trial timeline for the experiment. Participants were to remember the color of circles for either 1 s or 10 s.

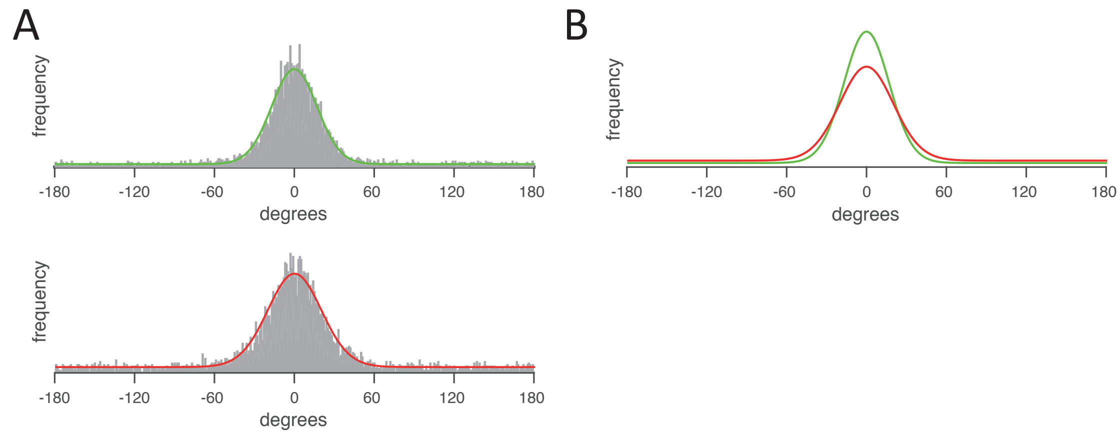

Figure 2: A: Frequency distributions of the error responses for all participants for the short (top) and long (bottom) retention intervals. The frequency distributions were fit with a weighted mixture of a circular normal and a uniform distribution (solid lines) separately for the short (top; green) and long (below; red) retention intervals. B: The fitted models for the short (green) and long (red) retention intervals are overlaid to highlight differences in shape of the distributions for the two conditions.

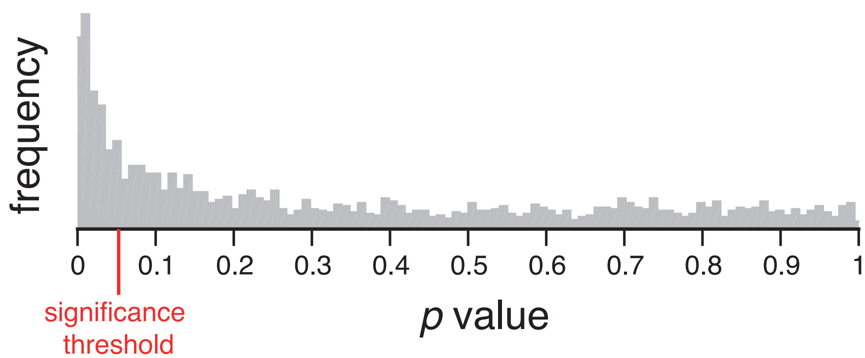

Figure 3: We performed a resampling analysis to determine the expected *p*-value distribution for differences in memory quality if our study had only 150 trials per condition, as in Zhang & Luck (2009), as opposed to the 300 trials per condition in our study. We performed 2000 bootstrap samples. For each resample, we selected 150 trials at random from each participant, separately for each retention interval. These data were fit by a mixture model to get quality estimates for each participant. For each resample, we compared the quality estimates for the retention intervals using a paired t-test. The gray bars show the frequency of *p* values for this resampling analysis. The *p* value of 0.10 found by Zhang & Luck (2009) (solid black line) falls within our resampled distribution, showing no evidence for large differences in data between the studies.

Results and Discussion

Changes in memory over time

Consistent with past work, performance in the memory task was worse at longer durations, with an average absolute error (offset between the response and the true value, in degrees) of 28.5° at 1 s and 36.1° at 10 s, $t(11) = 6.53$, $p < 0.001$ (paired sample *t*-test). While the increase in error demonstrates that memory worsens over time, it does not tell us whether memories become less accurate or die off. Fortunately, we can use an analytical technique known as mixture modeling to decompose the error distribution into two kinds of error (Zhang & Luck, 2008). In this approach, the error-frequency distributions of all trials in each retention condition (Figure 2) are used to fit a model with built-in assumptions about how errors are generated. The model used by Zhang & Luck (2009) assumed that participants either remember an item with a certain quality (in which case the responses accord with a circular normal distribution) or that an item is not stored (in which case the responses are guesses—uniform over the space of errors). Critically, different changes to memory will produce different expected shapes in the error distribution. A decrease in memory precision should produce a wider distribution peaked at the true value. An increase in guesses will uniformly increase all error values. In line with Zhang & Luck (2009), we modeled each error distributions as a weighted mixture of a uniform and a circular normal distribution (the model was fit separately for each participant and condition). This model had 3 parameters—the mixture parameter (weight of uniform distribution), a standard deviation parameter (width of the circular normal distribution), and a bias parameter μ (the offset of the circular normal distribution's mean relative to the true value). The μ parameter did not differ across conditions and was not considered further. Data were fit using the MemToolbox (Suchow, Brady, Fougnie, & Alvarez, 2013). Importantly, an increased guess rate is consistent with memory death, while a greater standard deviation would reflect changes in memory quality.

We found more random responses in the 10 s condition (20.0% at 1 s and 27.2% at 10 s; $t(11) = 3.86$, $p < .005$; Figure 2A), consistent with memories failing over time. We also found worse memory precision (increase in standard deviation estimate) for the 10 s condition (18.1° at 1 s and 21.7° at 10 s $t(11) = 2.76$, $p = .02$), suggesting a cost in memory quality over time. Similar results were found using alternative models of memory. Using a mixture model that allowed precision to vary across trials (Fougnie et al., 2012; see also van den Berg et al., 2012) we found higher random responses (18.7% vs. 25.5%; $p < 0.005$; Figure 2B) and worse precision (18.8° vs. 22.9°; $p < 0.01$) and no differences in the variability in precision (4.1° vs. 6.0°; $p < .21$). In addition, we also explored whether memory changes reflect a greater propensity to report the wrong item at the long retention interval (Bays, Catalo, & Husain, 2009). However, when we added a swap parameter to the model, we found a low swap rate, below 3%, and we observed differences in guess rate (19.7% vs. 25.5%, $t(11) = 3.13$, $p < .01$) and precision (18.1° vs. 22.0°, $t(11) = 2.53$, $p = .05$) between the short and long retention intervals.

Unlike Zhang & Luck (2009), we found significant changes *both* in guess rate and in precision over time. Importantly, our data resembles theirs. The difference in our conclusions about precision likely stems from the fact that our study had twice as many trials per condition. Indeed, if we down-sample the number of trials per participant to 150, half of our original number, we find that the p value reported by Zhang & Luck, 0.10, is not unexpected given our data. To show this we conducted a resample analysis in which we constructed 2000 resamples using a random 150 errors from each participant. We find that the p value reported by Zhang & Luck falls within the expected range of p values (Figure 3). Thus, there is no evidence to conclude that the findings by Zhang & Luck are inconsistent with our data—they may not have had sufficient power to detect a small change in precision. However, it is fair to conclude that that an increase in random responses is the *most drastic* change in memory responses, consistent with a model where memory death happens relatively suddenly.

Modeling the change in performance across retentions

In the above analysis, we followed precedent by separately fitting models across conditions and inferring the success or failure of a theoretical model by examining changes in the parameter estimates deemed most likely given observed data. A more principled and stronger test is to build a model that predicts how parameters change across conditions. The Sudden Death model predicts that, over time, memories will fail with no change in individual memory quality. Can such a model be fit simultaneously to data across multiple retention intervals?

The original Sudden Death model implicitly assumed that memories (within and across trials) are all the same quality. There is now considerable evidence for variability in memory quality (Fougnie et al., 2012; van den Berg et al., 2012). Fortunately, variability in memory quality can easily be incorporated into a Sudden Death model by starting from a theoretical model that is a weighted mixture of a uniform distribution and a mixture of Gaussian distributions (Fougnie et al., 2012). We extend the model to make predictions across retention intervals. To accomplish this, we include a free parameter to account for the increase in memory failures we expect between 1 and 10s. We leave this as a free parameter because the Sudden Death model is an attempt to explain why failures occur, not the number of failures in a given time interval. The model has four parameters—a guess rate (or mixture parameter), two parameters governing the distribution of memory quality, and a parameter accounting for the loss in memory over time.

The Sudden Death model of Zhang & Luck (2009) did not consider the question of which memories will fail, but we must consider it to predict how performance will change across retention intervals. We generated two versions of the Sudden Death model that differ in how they answer this question. In one version, Sudden Death–proportional (SD-P) (Figure 4), we assume that memories fail with probability proportional to the inverse of their precision (one over the variance in error), such that less informative memories are more prone to failure. As can be seen in Figure 6A, this model makes the counter-intuitive prediction that the average measurable memory quality will increase over time. This occurs because memories that are of lower quality are more likely to fail, removing them from the estimated circular normal distribution. We also considered a model (Sudden Death-random, SD-R) (Figure 4) where memories fail randomly, irrespective of their quality. This model predicts no change in the shape of the distribution of non-guess trials over time, only an increase in guess responses (Figure 6B). Both models predict that *individual memories* do not change in quality over time.

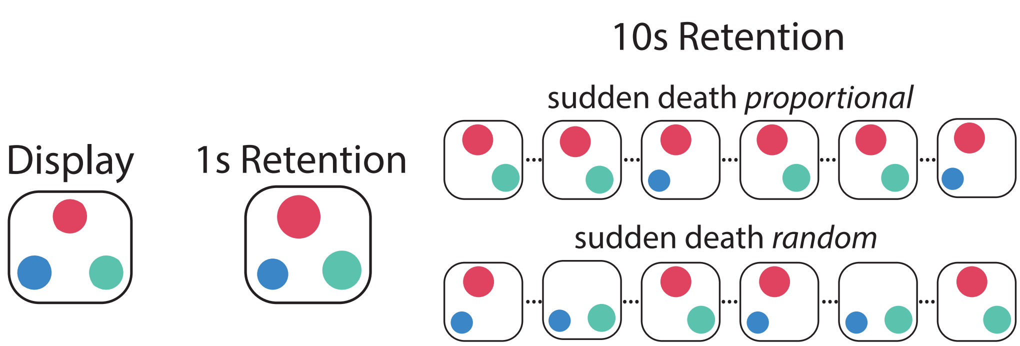

Figure 4: Schematics of Sudden Death models. A display of 3 colored items is encoded via three slots with variable precision (quality). Here we represent quality as the size of the color in memory (larger implies better precision). We have (arbitrarily) assigned red as the most precise and blue as the least precise. After one second, all items are in memory. By ten seconds, one item is lost via sudden death. Here we show six outcomes

of the memory state at 10 s. In the Sudden Death–*Proportional* model (top, SDP) the item that dies a majority of the time is the least precise item. In the Sudden Death–*Random* model (below, SDR) each item is equally likely to fail. Importantly, the precision of remaining items is higher in the SD–P model than the SD–R model (and is also higher than the memory state at 1s).

An alternative model

The Sudden Death model (Zhang & Luck, 2009) suggests that memory quality is constant over time, but that memories may suddenly fail. We propose an alternative model where information in memory is being lost over time leading to worse memory quality. Over time, the items with the lowest information may fail, perhaps because there is no longer any information left. Memories may terminate, not with a sudden death, but with a whimper.

Specifically, we assumed that items in memory were coded by a set of noisy samples (sample-based models; Bonnel & Miller, 1994; Palmer, 1990). This choice is agnostic to how flexibly information can be allocated to objects. Some models posit an infinite number samples while others assume only a limited number of independent slots of samples (slots; Zhang & Luck, 2008). Rather than specify the number of samples, we leave the number of samples and the information content of each sample as free parameters. These parameters jointly determine the maximum information capacity[1], but they are not the only factors influencing memory performance. Working memory requires maintaining perceptual information in the absence of bottom-up input. We propose that information is fallible and that not all samples will be available at test. To represent this, in our model each sample is given an equal and independent probability of surviving, a free parameter in the model (Figure 5). Thus, the content of memory is not solely determined by information capacity. Indeed, the information content of memory is set by a number of weighted coin flips where heads implies informational persistence of a sample. Memory quality over trials is related by a binomial distribution, which governs the probability of the number of samples that survive. This distribution predicts the expected distribution of error since the number of surviving samples is inversely proportional to response variance (Bonnel & Miller, 1994; Zhang & Luck, 2008). Note that our parameterization of the model involves expected trade-offs between parameters (e.g. a higher information content but greater failure rate bears some similarity to a state with lower information content but less failure of samples). Thus, the relationship between parameters becomes more important and each individual parameter value, considered in isolation, is less important.

To generate predictions across retention intervals, we added a fourth parameter to account for the degradation in memory. Over time, we expect a higher likelihood of sample failure. Thus, the changes over time should be reflected in an increased failure rate of samples. As with the Sudden Death model, we let this be a free parameter since we do not have an a priori prediction about the amount of degradation that should occur. Unlike the Sudden Death model, this account predicts that individual memories will get worse over time, because information is gradually lost. However, the end result is only a modest change in the width of the non-guess distribution (Figure 6C), given that low-quality memories at 1 s are highly likely to be removed from the pool of observable memories at 10 s, which acts as a slight countermeasure to the worsening quality of all memories. Importantly, this model suggests that 'guess' responses do not

[1] We make the simplifying assumption that participants divide the samples of information as equally as possible amongst the items to remember.

necessarily reflect a special state or process, but that they may be the end result of gradual memory degradation.

Model comparison

We fit each participant's data (including 1 and 10 s retention intervals) to the three models (Figure 7 shows the error distributions estimated for each condition by model) and compared them using Akaike's Information Criterion (corrected; AICc) a measure of goodness-of-fit that penalizes for the number of free parameters (Akaike, 1974; Burnham & Anderson, 2002). We found a clear victory for the DDNC model (AICc differences were 4.1 in favor of the DDNC model over SD-P and 2.9 over SD-R).

The data supports a model where memories worsen over time and where memory death reflects the end result of a degradation process rather then a sudden or distinct process. We rejected two versions of a Sudden death model, one in which failures were inversely proportional to quality and one in which failures were unrelated to quality. For the proportional version of the model we had to make an assumption about how quality and failure were related. Note, however, that any implementation with greater failure for memories with less information should predict some form of decrease in response variance in remembered items, and will therefore not be a successful model. Indeed, the Sudden Death model that assumes that failures are unrelated to quality was more successful in explaining the data then one that assumed that failures were proportional to uncertainty.

Our model assumed discrete samples that can be averaged to reduce uncertainty about items in memory. We chose this both because it is an influential and useful framework for thinking about the relationship between information and uncertainty (Bonnel & Miller, 1994; Luce, 1977; Zhang & Luck, 2008). However, some may prefer to think of the units of information as infinitely fine, or nearly so. Indeed, the discreteness in our model may be capturing the minimum information content for a memory to be useful as opposed to the discretization of memory above and beyond threshold. Thus, there would be large similarities between our model and a model with continuous information, information loss, and a threshold of information, below which, memories are inaccessible (Brady, Konkle, Gill, Oliva, & Alvarez, 2013).

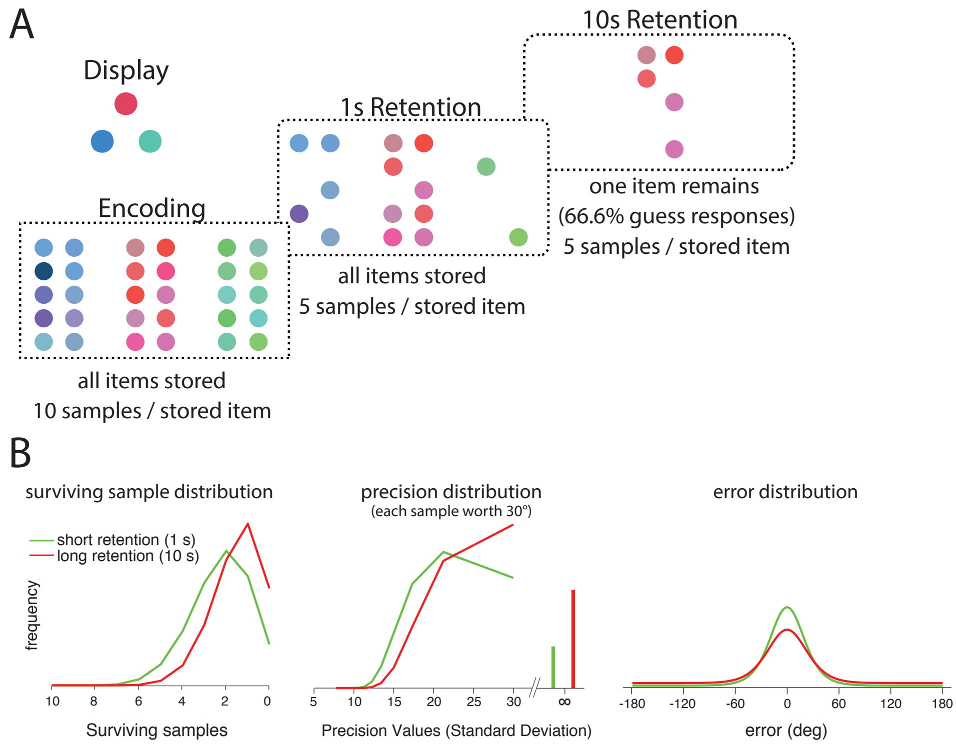

Figure 5: A: Schematic of Degradation and Death By Natural Causes model. A display of 3 colored items is encoded via 30 noisy samples (10 / item) (slight differences in sample color are meant to convey noise). Each sample is volatile, and has some probability of failing. By 1s half of the samples have failed, but the individual retains some information about each item. Variability in the quality of the items emerges as a result of how the probabilistic failure plays out. By 10s, perhaps only one item remains. The items of poorer quality may drop out, so even information if lost from each item, the average quality of the *remaining* items may not change. This example case is meant to demonstrate why quality estimates, taken in isolation, are not useful. The full characterization of the model (B) is a distribution of surviving samples (left) that produces a distribution of expected standard deviations when the information per sample is taken into account (B). This precision distribution yields an expected error distribution for each retention interval (C). The example model here has 10 samples and each sample has an uncertainty that would leave to a 30° circular standard deviation. The only change across retention intervals is in the probability of sample failure.

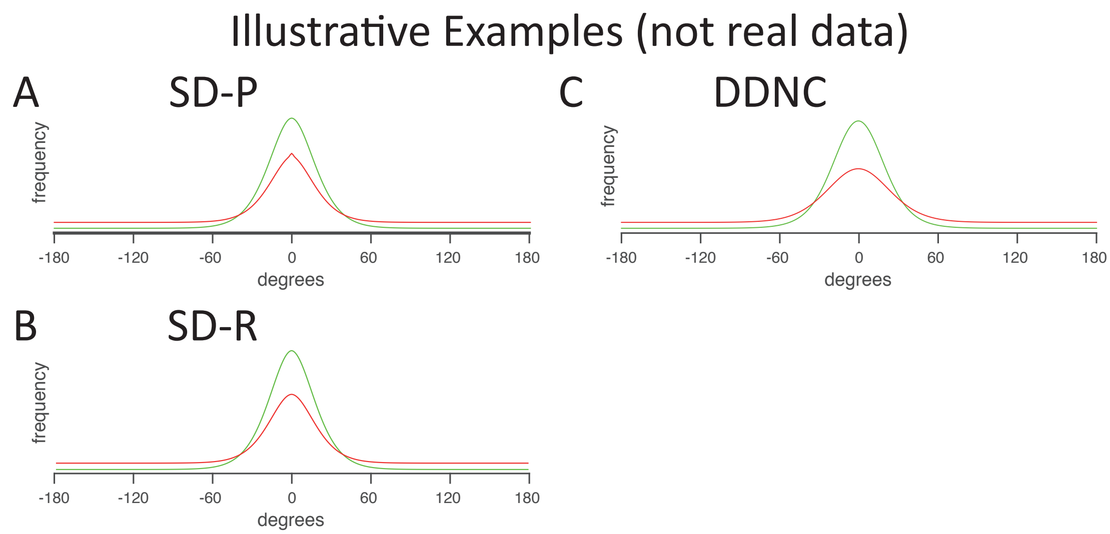

Figure 6: Example distributions for short (green) and long (red) distributions meant to convey the predicted changes to the shape of the error distributions for 3 models: Sudden Death–proportional (SD-P) (A), Sudden Death–random (SD-R) (B), and Degradation and Death by Natural Causes (DDNC) (C). To demonstrate effects for similar states, each model is given an equivalent state at short retention intervals and a delta parameter that is similar (expected change in memory over time). In the Sudden Death–proportional model the selective removal of low quality memories produces a distribution that is slightly more concentrated at 0 error. In the Sudden Death–random model the loss of memories does not change the shape of the distribution of remembered items. In the Death by Natural Causes model items more likely to be lost and remembered less precisely (red distribution is slightly wider) at long retentions. Note that the shown memory loss is more than one would expect for our experiment parameters, and is meant to illustrate the differences in model predictions by highlighting predictions with an exaggerated effect.

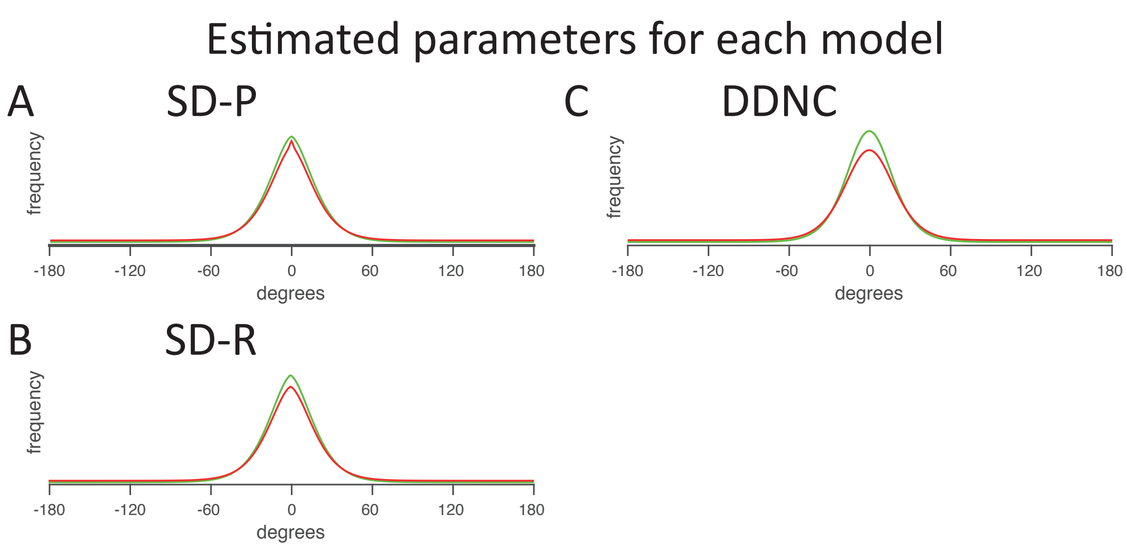

Figure 7: The model fitted distributions (using data from all participants) for short (green) and long (red) retention conditions for 3 models: Sudden Death–proportional (SD-r) (A), Sudden Death–random (SD-P) (B), and Degradation and Death by Natural Causes (DDNC) (C). Only the error distributions for the DDNC model (panel C) have more uncertainty during long retention (red), a characteristic found in the true data.

Results and Discussion

Despite our subjective impression that visual representations are rich and detailed, we represent a paucity of this information in mind (Noë, Pessoa, & Thompson, 2000; O'Regan, 1992; Simons & Chabris, 1999; Simons & Levin, 1997; Rensink, 2000, 2002). This is true even for objects defined by simple visual features, such as color or orientation (Luck & Vogel, 1997; Vogel et al., 2001). Research has focused on explaining these limits in terms of a capacity-limited store, with debates centered on understanding its nature and structure (Alvarez & Cavanagh, 2004; Bays & Husain, 2008; Bays, Wu, & Husain, 2011; Fougnie & Alvarez, 2011; Luck & Vogel, 1997; Wilken & Ma, 2004; Zhang & Luck, 2008). Though the storage metaphor for explaining working memory limitations has been productive, it impedes our understanding of how the processes of memory maintenance and degradation unfold over time. If limits to memory are can by explained by a limited store of information — analogous to a bucket that will only hold so much water — then why does memory performance drop as retention intervals in working memory tasks increase (Gold et al., 2005; Magnussen, 2000; Phillips, 1974; Sperling, 1960; Vergauwe, et al., 2009; Yang, 1999)?

Zhang and Luck (2009) tackled this question by analyzing the errors produced in a color estimation task across different retention intervals. One possibility is that individual memories gradually fade and become less precise over time. This possibility was ruled out because the distribution of errors did not grow wider. Instead, worse memory largely reflected an increase in responses spread across the space of possible errors—a pattern consistent with a decrease in memory existence and a corresponding increase in random guesses. The authors concluded that memories are stable up until they die suddenly. This model attempted to answer how memories change over time and provided a glimpse of how an information-limited approach could be extended to understand memory limitations at different time points.

Here, we show that the conclusion of the Sudden Death model is incorrect—over time, memories lose fidelity and become more imprecise. Memory failures may be the end result of this process of degradation. Thus, the end of our working memories is more akin to death by natural causes then to a sudden, un-precipitated event. Importantly, we are not arguing over the shape of the data. We similarly find large changes in the extent of what appear to be randomly distributed error values and more modest evidence for increases in the width of error distributions.

The central advance is how to interpret this data in the context of a full generative model. The Sudden Death model assumes that all memories are of equal quality, and therefore never considers which memories are dying. However, there is now considerable evidence for variability in the quality of individual memories both from modeling continuous-report data (Fougnie et al., 2012; van den Burg et al., 2012; see also Bae, et al., 2014, 2015) and from metacognitive judgments about internal uncertainty (Fougnie et al., 2012; Rademaker, Tredway, & Tong, 2012) even when stimulus factors are controlled (Suchow, Fougnie, & Alvarez, under review). Thus, it is imperative to consider which memories are failing. If memories with lower quality are more likely to fail, this could mask the ability to observe changes in memory quality over time, since we only measure the quality of memories that still exist. This issue, combined with the reduced sensitivity in the Zhang & Luck dataset (relative to ours), may have led to an incorrect conclusion that memories do not undergo changes in quality over time.

We performed model comparisons of versions of the Sudden Death model to our DDNC model. Our model starts from the assumption that information-quality is directly related to the amount of internal resource given to a stimulus (Bonnel & Miller, 1994; Luce, 1977; Shaw, 1978; Palmer, 1990). We allow this resource to be allocated in discrete, noisy samples that may be averaged to reduce uncertainty (Bonnel & Miller, 1994; Zhang & Luck, 2008). But rather than fixing the number of samples to be small (Zhang & Luck, 2008), or having information be infinitely divisible (Huang, 2010), the number of samples and their information capacity were free parameters in the model. Critically, our model also posits that information is fallible. Each sample was given an equal and independent probability of failing over time. The model holds that memory changes over time can be captured by increasing the probability of sample failure; over time more samples will have failed due to the volatility of information. Interestingly, we found that this model fit the error data remarkably well and that there was more evidence for it than for the Sudden Death models.

The idea that memories change over time is not new. Indeed, there is an extensive literature on whether changes in memory reflect interference or decay (e.g. Portrat, Barrouillet, & Camos, 2008; Cowan, 1999; Oberauer & Lewandowsky, 2008). When we use the term "degradation, we remain neutral on the source of changes in memory quality. However, in future work it will be important to understand more about why degradation occurs and what factors modulate it.

Here, we proposed a model where memory failures arise from a process of degradation that may lead, probabilistically, towards memory failure. This contrasts with the way evidence for random guesses is often theoretically interpreted—as evidence for some structural upper limit in the number of items that can be remembered. An implication of the fact that our model can account for the experimental data is to provide another plausible interpretation of the source of failures in working memory tasks. Rather then implying that these objects were never in memory, and never had any chance of being there, the lack of any (or very low) information could reflect a process whereby information is lost probabilistically, sometimes resulting in individual memories that contain no useful information by the time of test. A common metaphor for working memory is a storage space consisting of a set of slots or buckets in which information is placed—people make memory errors because only so much information can be placed in the buckets. A more appropriate metaphor would also consider whether the buckets might be leaky. This modification acknowledges that degradation plays a role in memory performance (rather then just storage limits) and affords a natural explanation of trial-to-trial variance in memory quality (even under conditions in which encoding and stimulus factors can be accounted for Suchow et al., submitted).

The end goal for research on working memory is not just to account for and measure the capacity of memory, but also to have a generative model of the process of memory—from stimulus encoding to response generation. Here, we've expanded our understanding in one critical way: by building a more complete model of how memories change over time. Having done so, we find evidence that contradicts fundamental assumptions about how the process of memory unfolds and challenges existing ideas about the nature of memory limits.

References

Akaike, H. A new look at the statistical model identification. (1974). *IEEE Transactions on Automatic Control*, 19, 716-723.

Alvarez, G. A., & Cavanagh, P. (2008). Visual short-term memory operates more efficiently on boundary features than it does on the surface features. *Perception & Psychophysics*, 70(2), 346–364.

Bae, G. Y., Olkkonen, M., Allred, S. R., & Flombaum, J. I. (2015). Why Some Colors Appear More Memorable Than Others: A Model Combining Categories and Particulars in Color Working Memory. *Journal of Experimental Psychology: General*, 144(4), 744-763

Bae, G. Y., Olkkonen, M., Allred, S. R., Wilson, C., & Flombaum, J. I. (2014). Stimulus-specific variability in color working. *Journal of Vision*, 14(4), 1-23.

Bays, P. M., Catalao, R. F. G. & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, 9, 1–11.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, 321(5890), 851–854.

Bays, P. M., Wu, E. Y., & Husain, M. (2011). Storage and binding of object features in visual working memory. *Neuropsychologia*, *49*, 1622–1631.

Bonnel, A.M., and Miller, J. (1994). Attentional effects on concurrent psychophysical discriminations: Investigations of a sample-size model. *Attention, Perception, & Psychophysics 55*, 162–179.

Brady, T. F., Konkle, T.F., Gill, J., Oliva, A. and Alvarez, G.A. (2013). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science*, 24(6), 981–990.

Burnham, K. P., & Anderson, D. R. (2004). Multi- model inference. *Sociological Methods & Research*, 33(2), 261–304.

Cowan, N. (1999). An embedded-processes model of working memory. In: *Models of Working Memory: Mechanisms of active maintenance and executive control* (Miyake A, Shah P, eds). New York: Cambridge University Press, 62–101.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114.

Fougnie, D. & Alvarez, G. A. (2011). Object features fail independently in working memory: Evidence for a probabilistic feature-store model. *Journal of Vision, 11,* 1–12.

Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision, 10*(12), 1–11.

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, 3, 1229.

Gold, J.M., Murray, R.F., Sekuler, A.B., Bennett, P.J., & Sekuler, R. (2005). Visual memory decay is deterministic. *Psychological Science*, 16(10), 769–774.

Huang, L. (2010). Visual working memory is better characterized as a distributed resource rather than discrete slots. Journal of Vision *10*, 1–8.

Luce, R. D. (1977). Thurstone's discriminal processes fifty years later. *Psychometrika*, 42, 461–489.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–281.

Magnussen, S. (2000). Low-level memory processes in vision. *Trends in Neurosciences*, *23*(6), 247–251.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63, 81–97.

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(2), 332–350.

Phillips, W. A. (1974). On the distinction between sensory storage and short- term visual memory. *Perception and Psychophysics*, 16, 283–290.

Noe, A., Pessoa, L., and Thompson, E. (2000). Beyond the grand illusion: What change blindness really teaches us about vision. *Visual Cognition, 7*(1-3), 93–106.

Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference?. *Psychological review*, *115*(3), 544.

O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception: The world as an outside memory. Special issue: Object perception and scene analysis. *Canadian Journal of Psychology, 46*(3), 461–488.

Rademaker, R. L., Tredway, C. H., & Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of Vision*, *12*(13), 21.

Rensink, R. A. (2000). Visual search for change: A probe into the nature of attentional processing. *Visual Cognition, 7*(1-3), 345–376.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology, 53*(1), 245–277.

Shaw, M. L. (1978). A capacity allocation model for reaction time. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 586–598.

Sperling, G.(1960). The information available in brief visual presentations. *Psychological Monographs*, 74.

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*(9), 1059–1074.

Simons, D. J., and Levin, D. (1997). Change blindness. *Trends in Cognitive Science, 1*, 261–267.

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*(10).

Suchow, J. W., Fougnie, D., & Alvarez, G. A. (under review). Looking inwards and back: isolating the contribution of realtime monitoring to judgements of confidence in memory.

van den Berg, R. & Shin H., C. W. C., George, R. Ma, W.J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences,* 109, 8780–8785.

Vergauwe, E., Barrouillet,P., & Camos, V. (2009). Visual and spatial working memory are not that dissociated after all: a time-based resource-sharing account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(4), 1012.

Vogel, E. K., Woodman, G. F., and Luck, S. J. (2001). Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27*(1), 92–114.

Wilken, P. A & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4, 1120–1135, doi:10.1167/4.12.

Yang, W. Lifetime of human visual sensory memory: Properties and neural substrate. *University of Pennsylvania Institute for Research in Cognitive Science Technical Report No. IRCS-99-03. (1999).*

Zhang, W., & Luck, S. (2008). Discrete fixed-resolution representations in visual working memory. *Nature, 453*(7192), 233–23.

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science*, 20(4), 423–428.