# Incidental Memory for Naturalistic Scenes: Exposure, Semantics, and Encoding

**Moreno I. Coco (micoco@psicologia.ulisboa.pt)**

Faculdade de Psicologia, Universidade de Lisboa

Alameda da Universidade, Lisboa, 1649-013, Portugal

**Nicholas D. Duran (nicholas.duran@asu.edu)**

School of Social and Behavioral Sciences, Arizona State University

Glendale, AZ 85036, USA

## Abstract

Visual memory for naturalistic scenes is mediated by: amount of exposure, semantic content, and type of encoding. These factors might interactively contribute to scene memorability. Thus, we tracked computer-mouse movements during an encoding phase where participants verified the congruency of sentence and scene pairs, which varied in plausibility. The presentation time of the scenes was also manipulated. Subsequently, in an unexpected recognition phase, participants had to indicate whether they remembered scenes (old and new). Recognition improved when correct verifications were made during encoding especially: when the scene was implausible, the stimuli pair congruent, and for longer presentation times. When comparing the trajectories between encoding and recognition, we found greater hesitancy during encoding, especially for implausible scenes in incongruent pairs, decreasing as presentation time increased. These results provide novel insights into the factors modulating the memorability of naturalistic scenes.

**Keywords: visual memory; action-dynamics; presentation time; semantic plausibility; active encoding.**

## Introduction

The visual system needs as little as 20 milliseconds to extract semantic information from a complex scene for subsequent decision-making (Thorpe, Fize, & Marlot, 1996). Besides showing rapid processing, the visual system is also extremely efficient at retaining in memory information about naturalistic scenes, both for short- and long-term storage (Brady, Konkle, Gill, Oliva, & Alvarez, 2013). For example, this has been shown in Potter (1976)'s seminal study, where short-term recognition of rapidly presented pictures (e.g., 50 ms) was highly accurate, improving to almost ceiling performance as the presentation time of the pictures increased. More recently, Konkle, Brady, Alvarez, and Oliva (2010) have found that recognition of naturalistic scenes can be equally impressive for long-term memory. In their study, participants were asked to view, for a later recognition test, a long sequence of naturalistic scenes (2912 items), displayed one by one, and drawn from different contexts. Even when up to 64 different scenes were drawn from the same context, recognition accuracy remained well above chance.

However, not all visual scenes are equally memorable; and the semantic information of the objects contained within a scene has been found to be a key predictor of its recognition (Isola, Xiao, Torralba, & Oliva, 2011). Importantly, objects can be more or less plausible within a certain scene context,

where a scene depicting 'a boy eating a SANDWICH' is certainly more plausible than 'a boy eating a BRICK'. Plausibility has been shown to have widespread impact on behavioral (Coco, Malcolm, & Keller, 2014) and neurological processes (Mudrik, Lamy, & Deouell, 2010)). With respect to visual memory, the plausibility of a scene is known to influence its encoding and subsequent recognition. For example, Davenport and Potter (2004) found that when an object in the foreground of a scene (e.g., a priest) is implausible with the overall background (e.g., a soccer field versus a church), the ability to later recognize the object significantly declined.

Even so, the memorability of scenes does not depend on their plausibility content alone, but also on whether the information attended to is actively processed during encoding. One piece of evidence is Makovski, Jiang, and Swallow (2013)'s study that shows recall for faces improves if participants have to judge the sex of the character depicted during the encoding phase, compared to when such judgment was not required.

From this brief review, we can isolate three key factors responsible for mediating the representational strength of complex scenes in visual memory. The first factor is presentation time, with longer exposures implying more accurate recognitions than short exposures. The second factor is the plausibility of visual scenes, with implausible scenes implying less accurate recognitions than plausible scenes. And finally, the type of encoding condition under which visual scenes have been viewed, with active encoding implying better recognition than passive encoding.

These three factors are not mutually exclusive, and thus it is unlikely that they will bear independent consequences on visual memory. In fact, longer presentation time of an implausible visual scene might, for example, aid its recognition; or conversely, more efficient encoding might result in better recognition, independent of presentation time. Thus, in this study, we examine the possible interdependencies between such factors, building on attempts that have only recently been explored.

## The Present Study

The overarching goal of this study is to investigate how presentation time, plausibility of information, and active encoding contribute to memory formation for naturalistic scenes and on their subsequent recognition. We do so using an action

dynamics framework that tracks the semi-continuous movement of the arm during reaching decisions (Spivey & Dale, 2006). In such a framework, participants are typically presented with two alternative choices at the top of a computer screen, and response trajectories are monitored en-route to the response. Analysis of these micro-behaviors have been shown to correlate with underlying cognitive mechanisms in a variety of domains spanning deception (Duran, Dale, & Mc-Namara, 2010) to memory (Papesh & Goldinger, 2012).

Participants completed two consecutive tasks: a verification (encoding) task and an unexpected memory (recognition) task, i.e., participants were not informed that their memory would be tested. To preview the experiment procedure detailed in section *Apparatus and Procedure*, during the encoding phase, participants first read a sentence, then viewed a scene, and finally had to judge whether the content of the pair of stimuli matched or not (by moving their computer mouse to a response button). During the recognition phase, participants saw a scene that was either present or not during encoding and had to judge whether they remembered it (again by moving their computer mouse). These tasks allow us to investigate two primary theoretical questions: (1) Is recognition for rapidly presented scenes, which are incidentally memorized, predicted by responses at encoding? (2) Is plausibility of information playing a role on recognition, and if so, how do responses compare between encoding and recognition?

## Method

The experimental design crossed *Plausibility* (plausible, implausible) of information depicted in the sentence and visual scene with the *Congruency* (congruent, incongruent) between each pair (i.e., did the two match in content or not?) (schematically described in Figure 1). Both factors were within-participants. We also manipulated the *Presentation Time* of the visual scene (33, 100, 250, 500) between-participants (twelve participants for each presentation time).

## Participants

Forty-eight students at the University of Lisbon, all native speakers of Portuguese, participated in the study for course credit. The experiment was granted by the Ethics Committee of the Department of Psychology, in accordance with the University's Ethics Code of Practice.

## Materials

We selected 56 sentence-scene pairs from a previous study based on a similar paradigm (Coco & Duran, under review) [1], such that there was no difference in the lexical frequency of the target word/object between plausible ($M = 822.72$) and implausible ($M = 909.29$) conditions, $t(209) = 0.61$, $p = 0.5$; but that, at the same time, plausible scenes were rated as more likely ($M = 5.20$, $SD = 0.4$) than implausible scenes ($M = 1.82$, $SD = 0.43$), $t(220) = 60.16$, $p < .00001$, on a Likert



Figure 1: Experimental design with a full set of crossed pairs of stimuli: Plausibility (Plausible and Implausible) and Congruency (Congruent, Incongruent). The sentence is read self-paced, then a scene is presented for either 33, 100, 250, or 500ms. The target word (e.g., *hamburger* vs. *brick*) was always positioned at the end of the sentence.

scale [2]. This ensures that effects observed on the responses are not contaminated by lexical properties of the stimuli, but are genuinely triggered by their plausibility. Sentences were written in Portuguese, and had a fixed length of either seven or eight words to minimize variability between items. The sentences were also checked for grammaticality by two independent native-speaking annotators, who also ensured that the target object depicted in the scene was recognized as the target word used in the sentence.

The target object was pasted into the scene using the free software GIMP. The size of each scene was fixed at 550 x 550 pixels. Each scene was presented in two plausibility conditions (plausible: a boy eating an hamburger; implausible: a boy eating a brick). We crossed plausibility with congruency by pairing each scene in the plausible conditions with two different sentences, as with Figure 1. We had a total of 224 unique sentences (112 scenes paired with 2 different sentences.

## Apparatus and Procedure

The experiment was designed using Adobe Flash 13.0, which allows sampling at 60 Hz. The stimuli were presented on a 21" plasma screen at a resolution of $1024 \times 768$ pixels. Participants sat between 60 and 70 cm from the computer screen. Calibration of the mouse position was ensured by forcing participants to click on a black target circle (36 pixels across) located precisely at the bottom-center of the screen at the start of the trial, and throughout its different phases.

During the encoding phase, participants first read a sentence, using a word-by-word self-presentation method, by clicking on the calibration button located at the bottom of the screen. After the last word was read, a visual scene was displayed for either 33, 100, 250 or 500 ms. The scene then

---

[1]Approximately half of the scenes included were taken from the dataset of Mudrik et al. (2010)

[2]Plausibility judgments were collected on a sample of sixty-four participants, refer to Coco and Duran (under review)
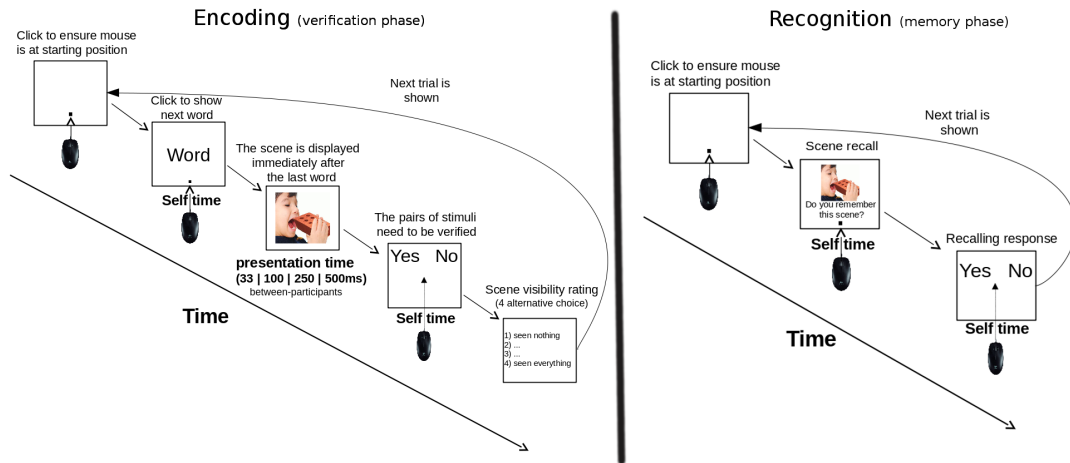
Figure 2: An example of a trial run for the *encoding* (left) and *recognition* (right) tasks.

disappeared and the response buttons (YES, NO), counterbalanced (left/right) between participants, were displayed at the top of the screen. Once the participant clicked on a response, on a separate screen, they were asked to rate (four possible choices) how clearly they saw the scene. As expected, scene visibility significantly increased as a function of the stimulus presentation time, $\beta = 0.001, t(2710) = 43.72, p < .0001$, hence assessing the validity of this manipulation (refer to Figure 2 (left panel) for an example of encoding trial).

After the encoding phase, an unexpected recognition task was presented, where participants were asked whether they remembered scenes from a pool of 56 (28 Old, 7 for each experimental condition; and 28 New, 14 plausible and 14 implausible) presented one at time. To make these responses, a scene was presented that participants could view for as long as they liked. When they were ready to respond, they clicked on the calibration button, and response buttons (YES, NO) appeared at the top of the screen (identical to the encoding trial, refer to Figure 2, right panel, for an example of the recognition trial).

Each participant completed a total of 112 randomized trials (56 in encoding and 56 in recognition). The experiment took about 30 to 45 minutes to complete.

**Analysis**

In order to investigate the relationship between encoding and recognition, we focus our analysis on the *old* items, where we have responses for the same items in both encoding and recognition. From these, we analyzed a total of 1,313 unique trials. We removed 2% (31 trials) of the data due to encoding verification times that were greater than 4 standard deviations from the mean, or due to machine error.

**Accuracy** Our first research question addresses whether recognition of naturalistic scenes could be predicted by responses given during encoding. We are especially interested in whether correct verification of sentence-scene pair congruency (whether it was indeed congruent or not), bears any consequence on the recognition of the visual scene.

Our first analysis therefore examines whether correct responses given during recognition, a binomial variable (0, incorrect; 1, correct), were predicted by accuracy at encoding, also a binomial variable (0,1). We also examined whether recognition accuracy is mediated by encoding, scene plausibility, and its presentation time.

We expect that accurate encoding should enhance scene recall, particularly when the scenes were plausible, or congruent with an associated sentence. Consistent information is, in fact, known to be processed more efficiently (Davenport & Potter, 2004). However, we extend previous literature by showing that memorability of scenes is enhanced, even when not explicitly probed by the task, i.e, incidental memory. In line with previous literature, we also predict that longer presentation times will enhance recognition, with the greatest gains again for scenes previously processed as plausible and congruent.

To conduct these analysis, we employed linear mixed-effects models based on the R statistical package lme4 (Bates, Maechler, & Bolker, 2011), examining recognition accuracy as a function of a) encoding accuracy, b) whether scenes were congruent or not with an associated sentence, c) the plausibility of the scene, and c) presentation time.

We construct full linear-mixed effects model, i.e., predictors are entered as main effects as well as in interaction, with maximal-random structure (MLME, (Barr, Levy, Scheepers, & Tily, 2013)), where each random variable of the design (e.g., Participants), is introduced as intercept, and as uncorrelated slope on the predictors of interest (e.g., Plausibility). The random variables of our design are Participants (48), treated as a between design variable, and Scenes (112, as we have 56 scenes in two conditions of Plausibility).

Moreover, we controlled for possible effects of *encoding order* (the order in which visual scenes were presented) and *recognition order* (accounting for fatigue) on recognition accuracy. This was done by residualising recognition accu-

racy on these two co-variates in a simple linear regression model (logistic), *RecognitionAccuracy* ∼ *EncodingOrder* + *RecognitionOrder*, using the R syntax, and taking the residuals obtained as the new DV for inferential analysis [3].

We report tables with coefficients of the predictor-terms that were significant at $p < .05$, rather than all terms. We also report their standard error, the $t - value$, and derive $p$-values, as calculated from F-test based on Satterthwaite approximation to the effective degrees of freedom.

**Response dynamics** For our second research question, we examined 967 trials ($\approx 71\%$) from the larger old-items subset, where accurate responses were given both during encoding (overall accuracy = $\approx 80\%$) and recognition (overall accuracy = $\approx 90\%$), and compared the arm movement response dynamics between the encoding and recognition phases. We focus on two dynamical measures: (a) *latency* (the time taken to move outside an initial region of 100 pixels around the calibration button), which represents the initial hesitancy to commit to a decision, and (b) *x-flips* (the number of directional changes on the x-axis), which indicates changes of mind as the decision unfolds. Response dynamics were calculated using the R-package mousetrack (Coco & Duran, 2015).

We expect latency of the movement to be longer during the encoding response compared to the recognition response, as encoding involves a comparison between two stimuli (i.e., a sentence and a scene), whereas recognition involves recalling only one stimulus (i.e., the scene)

Moreover, for incongruent and implausible pairs, we also expect greater latency for encoding, corroborating with our previous results Coco and Duran (under review). However, by increasing presentation times, we expect the latency to start the movement to reduce. We also expect *x-flips* to corroborate these predictions, where a greater number of x-flips will be seen during encoding compared to recognition, particularly for those items that were judged as incongruent and implausible.

We applied similar linear mixed-effects models as above for each dynamical measure, but with a fixed-effect for Task (Encoding vs. Recognition) instead of Encoding Accuracy, as well as, the other experimental variables of interests, Presentation Time, Plausibility and Congruency.

## Results and Discussion

### Accuracy

In Figure 3, we illustrate how recognition accuracy relates to encoding accuracy during the experimental conditions of Congruency and Plausibility [4]. We find that when the visual scene was congruent with a sentence in the encoding, and the

---

[3]Note, we opted to residualize these two co-variates, rather than include them together with the experimental variables of interest in the lme model, to avoid rather complex models with interactions up to the fifth order

[4]Note, the dependent measure used in the LME is a trial-by-trial residualized response of recognition accuracy. Means over subjects are only taken to better visualize the trend as continuous, i.e., at a trial-by-trial level we only observe 0s or 1s.
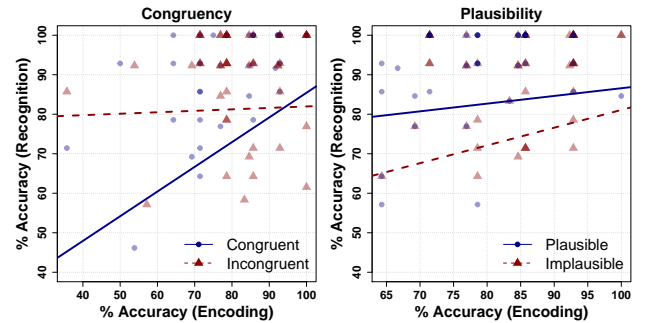


Figure 3: Scatter plot of recognition accuracy as a function of encoding accuracy. Each point is the mean accuracy of a subject. On the left panel, we compare the two conditions of Congruency; on the right panel, we compare the two conditions of Plausibility. The lines represents the estimate of a generalized linear model fit to the data.

Table 1: Mixed-effect maximal model analysis of recognition accuracy, with centered and contrast-coded fixed effects for Encoding Accuracy (EA), Congruency (Incongruent: $-0.5$, Congruent: $0.5$), Plausibility (Implausible: $-0.5$, Plausible: $0.5$), and Presentation Time (a continuous variable with four values: 33, 100, 250, 500). Random intercepts and slopes on Participant (48) and Scene (112).

| Predictor | β | SE | t | p |
|---|---|---|---|---|
| Intercept | 0.003 | 0.014 | 0.175 | 0.8 |
| Presentation | 0.143 | 0.036 | 3.937 | 0.0001 |
| EA:Congruency | 0.134 | 0.042 | 3.206 | 0.001 |
| EA:Plausibility | -0.128 | 0.043 | -2.996 | 0.004 |
| EA:Congruency:Presentation | -0.228 | 0.114 | -2.011 | 0.04 |
| Congruency:Plausibility:Presentation | 0.162 | 0.082 | 1.965 | 0.05 |

response at encoding was also correct, the recognition of the scene is facilitated (two-way interaction EA:Congruency, refer to Table 1 for coefficients and their significance.) Implausible scenes are better recognized if they were correctly verified during encoding (two-way interaction EA:Plausibility). Presentation Time also plays an important role. In line with previous literature, we find that with increasing presentation times, recognition accuracy significantly improves. A benefit is found for visual scenes from incongruent pairs, especially when the verification was correct (three-way interaction Encoding Accuracy:Congruency:Presentation Time). Finally, plausible visual scenes in congruent pairs are also more likely recognized for increasing presentation times (three-ways interaction Congruency:Plausibility:Presentation Time.

These results highlight an intriguing dependency between the type of encoding of visual scenes during a verification task and their recognition. In particular, if correct verifications were given during encoding, then visual scenes were more likely to be remembered. However, this was only the case if the verification was with congruent stimuli. This indicates that correctly accepting as congruent a pair of stimuli might strengthen their memorability, and therefore enhance their recognition. Crucially, a similar effect is found on plausibility of stimuli. Recognition for implausible scenes improves if they were correctly verified. Beside confirming the classic presentation time (SOA) effect, our results go beyond it by showing that its effect is modulated by congruency of
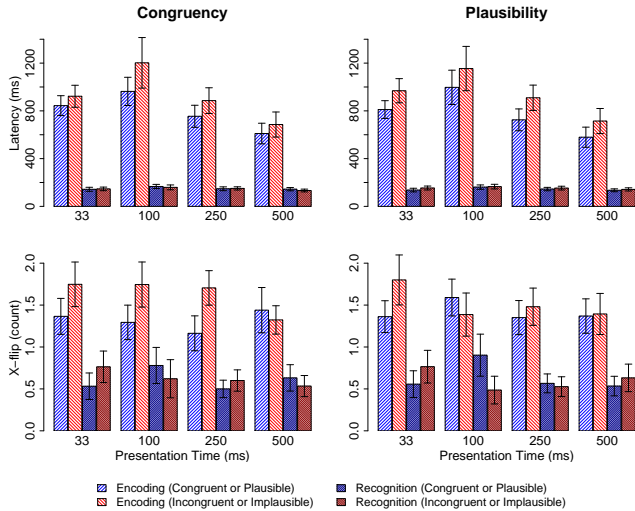
**Congruency** · **Plausibility**

Figure 4: Bar-plot, mean, and 95% confidence intervals for *latency* and *x-flips*. Compares Encoding (low density bars) and Recognition (high density bars) items in terms of initial Congruency (congruent - blue, incongruent - red) and Plausibility (plausible - blue, implausible - red), left and right panels, respectively.

Table 2: LME estimates values for the dynamical measures: *latency* and *x-flips*. Centered and contrast-coded fixed effects include Task (Encoding: −0.5, Recognition: 0.5), Congruency (Incongruent: −0.5, Congruent: 0.5), Plausibility (Implausible: −0.5, Plausibile: 0.5), and Presentation Time (a continuous variable with four values: 33, 100, 250, 500). Random intercepts and slopes on Participant (48) and Scene (112).

| | Latency | | | |
|---|---|---|---|---|
| Predictors | β | SE | t | p |
| Intercept | 526.854 | 27.864 | 18.908 | 0.0001 |
| Task | -750.008 | 53.558 | -14.004 | 0.0001 |
| Congruency | -49.529 | 21.574 | -2.296 | 0.02 |
| Plausibility | -80.251 | 26.247 | -3.058 | 0.002 |
| Presentation | -180.802 | 72.521 | -2.493 | 0.01 |
| Task:Congruency | 109.675 | 35.948 | 3.051 | 0.002 |
| Task:Plausibility | 140.527 | 50.516 | 2.782 | 0.005 |
| Task:Presentation | 328.152 | 139.962 | 2.345 | 0.01 |
| | X-flips | | | |
| | β | SE | t | p |
| Intercept | 1.038 | 0.049 | 20.979 | 0.0001 |
| Task | -0.849 | 0.068 | -12.501 | 0.0001 |
| Congruency | -0.171 | 0.062 | -2.761 | 0.009 |
| Task:Congruency | 0.296 | 0.095 | 3.124 | 0.002 |
| Congruency:Presentation | 0.327 | 0.157 | 2.089 | 0.04 |

the stimuli pair, and their plausibility.

## Response Dynamics

In Figure 4, we plot the means and confidence intervals for the dynamical measures of *latency* and *x-flips*, collected as participants responded to the verification task (encoding) and when accessing the scene from memory (recognition).

We find significantly longer latency times during encoding compared to recognition (main effect of Task; refer to Table 2). Also, incongruent pairs, and implausible stimuli display longer latencies during encoding and recognition (main effects of Congruency and Plausibility). Moreover, interactions between Task and Congruency and Task and Plausibility show that these differences are particularly strong during encoding. In addition, an interaction between Task and Presentation Time indicates that latency time decreases as presentation time increases, again, being most pronounced for encoding.

These results suggest that verifying the congruency of the stimulus pair (sentence-scene) during encoding might accrue a greater cognitive cost than merely recognizing the scene. Verification does require the comparison and integration of linguistic and visual content to make a decision, whereas recognition has relatively fewer demands. This greater cognitive cost is particularly evident for incongruent and implausible scenes: a result corroborating the accuracy analysis, and aligning with previous literature showing greater difficulty in processing inconsistent information. Moreover, presentation time reduces latency time during the more costly encoding task [5].

When examining *x-flips*, we observe greater indecision (more x-flips) during encoding than recognition (main ef-

[5]These results are entirely corroborated by overall response time.

fect of Task). Incongruent pairs also triggered more x-flips (main effect of Congruency). However, increasing presentation time does reduce the number of x-flips for incongruent trials (two-way interaction Congruency:Presentation Time). These results confirm that encoding/verification is associated with more indecisiveness compared to recognition. They also provide converging evidence with the latency results. Crucially, we observe an interesting two-ways interaction between Task and Congruency, whereby visual scenes correctly encoded as incongruent, as compared to scenes encoded as congruent, trigger greater indecisiveness (more x-flips) on the later recognition task. This suggests that congruency of information at encoding mediates responses downstream when the same scene has to be recognized, and corroborates similar "transfer" effects observed in recognition accuracy. We do not find plausibility, however, to play any particular role on the indecisiveness of the response trajectory.

## Conclusion

A momentary glance is sufficient for the visual system to retain complex information about naturalistic scenes with surprising fidelity (Thorpe et al., 1996; Brady et al., 2013). Even so, the memorability of scenes still largely depends on their presentation time, the semantic information they convey, and the way they are encoded (Potter, 1976; Davenport & Potter, 2004; Makovski et al., 2013). These factors have mostly been studied independently, but by uncovering their interdependencies, a more integrated understanding of memory processes for visual scenes might be provided.

Using an action dynamics mouse-tracking approach, we examined how some of these interdependencies mediate recognition accuracy of naturalistic scenes, and differentially modulate the moment-to-moment response dynamics during

their encoding and subsequent recognition.

Our results show that the recognition accuracy of visual scenes improves when the scene was correctly encoded during an initial verification task, where participants had to assess whether its content matched, or not, in content with an associated sentence. The improvement effect was most pronounced in connection with congruent pairs. That is, when participants correctly indicated that sentence-scene pairs shared content, it strengthened the memory for those scenes when assessed in an unexpected follow-up recognition task. Memory might also be strengthened by congruency of the stimuli, because in such a case, the same message (e.g., the boy Moreover, a similar effect was found in connection with plausibility, but where correct recognition of implausible scenes positively correlated with performance on the initial verification (encoding) task.

This effect is interesting in that recognition for implausible scenes has been associated with poorer performance in other studies. But when this information is more actively processed, such as in the verification task employed here, performance is improved.

The selected dynamical response measures also aligned with the accuracy results. We observed, for example, that hesitancy (*latency*) and indecisiveness (*x-flips*) were significantly greater for visual scenes encoded in incongruent pairs. However, hesitancy and indecisiveness decreased as the presentation times of scenes increased during encoding/verification. Presentation times also had an effect on the accuracy analysis, where greater presentation times improved later recognition accuracy, and where it also mediated the effects of congruency, plausibility, and encoding accuracy.

To reiterate, our results highlight the interdependence between presentation time, stimulus plausibility, and active encoding on the long-term memory of visual information. It also raises possibly interesting questions about the role of explicit feedback or memory task expectations during encoding, given none were present yet stable memories still occurred. Moreover, because memories persisted over a somewhat long delay between encoding and a follow-up recognition task, another direction for follow-up research would be to systematically evaluate the resilience of incidental encoding over longer time-scales. Doing so would bolster our current goal of examining the underlying mechanisms contributing to the memory formation of visual scenes.

## Acknowledgments

## References

Barr, D., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., Maechler, M., & Bolker, B. (2011). *lme4: Linear mixed-effects models using s4 classes.*

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

Coco, M. I., & Duran, N. D. (2015). mousetrack: Mouse-tracking measures from trajectory data [Computer software manual]. (R package version 1.0.0)

Coco, M. I., & Duran, N. D. (under review). Action dynamics as revision costs: The interaction of plausibility and congruency during a verification task.

Coco, M. I., Malcolm, G. L., & Keller, F. (2014). The interplay of bottom-up and top-down mechanisms in visual guidance during object naming. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1096–1120.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559–564.

Duran, N., Dale, R., & McNamara, D. (2010). The action dynamics of overcoming the truth. *Psychonomic Bulletin & Review*, *17*(4), 486–491.

Isola, P., Xiao, J., Torralba, A., & Oliva, A. (2011). What makes an image memorable? In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011* (pp. 145–152).

Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010). Scene memory is more detailed than you think the role of categories in visual long-term memory. *Psychological Science*, *21*(11), 1551–1556.

Makovski, T., Jiang, Y. V., & Swallow, K. M. (2013). How do observers responses affect visual long-term memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(4), 1097.

Mudrik, L., Lamy, D., & Deouell, L. (2010). ERP evidence for context congruity effects during simultaneous object - scene processing. *Neuropsychologia*, *48*, 507–517.

Papesh, M. H., & Goldinger, S. D. (2012). Memory in motion: Movement dynamics reveal memory strength. *Psychonomic Bulletin & Review*, *19*(5), 906–913.

Potter, M. C. (1976). Short-term conceptual memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(5), 509.

Spivey, M., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, *15*(5), 207–211.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*(6582), 520–522.