2026 ■ Vol. 22 ■ no. 2

A Replication of Rensink, O'Regan, and Clark (1997): Change Blindness in an Online Flicker Paradigm

M. M. Ouattara [a], E. Alexandre [a], M. Gagnon [a], F. Joanette [a], É. Lamarre [a], S. Mohamed Guirreh [a], O. Aboua Adon [a], A. Bellout [a], B. Billings [a], G. Bisson [a], V. Bouchard [a], M. Boucher [a], R. Boundzanga [a], P. Brousseau [a], L. Clark [a], E. Connors-Lee [a], E. Dabal [a], D. Daho [a], A. Gendron [a], C. Gowan [a], M. Grégoire [a], L. Hadley [a], W. Jean [a], A. Kaba [a], D. Mailhot [a], S. Mesidor [a], E. Minor [a], N. Montpetit [a], K. Prince [a], C. Ready [a], Z. Sayah [a] & Y. Tangba [a]

[a]University of Ottawa

Abstract ■ This study is a partial replication of Experiment 1 of Rensink, O'Regan, and Clark (1997), which introduced the flicker paradigm to investigate change blindness. Change blindness refers to the failure to detect large visual changes when a brief perceptual disruption eliminates the local motion signal that normally captures attention. Although widely cited as a foundational demonstration of attentional limits in visual perception, the original findings have not been subject to systematic independent replication. The present study addressed this gap using a larger sample (N = 69) and an online administration format via jsPsych. Participants were allocated to a flicker condition or a no-blank condition, with recruiters blind to condition assignment. Two hypotheses were tested: that the flicker manipulation would increase change detection reaction times, and that marginal interest changes would be detected more slowly than central interest changes, with this difference amplified under flicker conditions. Both hypotheses were supported. The flicker condition produced substantially longer detection times (Mann-Whitney U = 150.00, Z = -5.318, p < .001, Cohen's d = 1.36). Marginal interest changes were harder to detect than central interest changes across all participants (Z = -6.62, p < .001), with the location effect substantially amplified under flicker conditions. Overall detection accuracy did not differ between conditions, ruling out a speed-accuracy tradeoff. These findings replicate the core results of Rensink et al. (1997), in a larger online sample, supporting the robustness of the flicker effect beyond controlled laboratory settings..

Keywords ■ change blindness, flicker paradigm, visual attention, replication, change detection.

Mouat085@uottawa.ca

10.20982/tqmp.22.2.p056

Acting Editor ■ Denis Cousineau (Université d'Ottawa)

Reviewers
■ One anonymous reviewer

Introduction

Empirical observation is a foundation of science; we trust what we can see. Yet, what if perception itself were fundamentally unreliable? What if a person standing directly in front of you could be replaced by a stranger, mid-conversation, and you would simply fail to notice? To understand why such striking failures of perception occur, it is necessary to examine the mechanisms underlying visual attention and the fragility of short-term visual representations.

Change blindness refers to the failure to detect large, clearly visible changes in a visual scene when those changes coincide with a brief disruption to visual continuity (Simons & Levin, 1997). Although observers typically experience the visual world as rich and continuous, extensive research has shown that even substantial changes often go unnoticed (Rensink, 2002; Simons & Ambinder, 2005). This phenomenon challenges intuitive assumptions about visual perception and has important implications for understanding the limits of attention and conscious awareness.

Early work in this area emerged from two complementary lines of investigation. When brief interruptions are introduced between visual displays, observers show marked impairments in detecting changes, suggesting that visual representations are fragile and easily disrupted (Phillips, 1974). In parallel, research on attention has demonstrated that only a subset of visual information is selectively processed, with unattended elements receiving little to no detailed encoding (Posner, 1980). Together, these findings converged on the view that conscious visual representation is neither rich nor continuous, but rather limited and dependent on moment-to-moment attentional allocation. Successful change detection thus depends critically on attentional allocation, such that changes may go unnoticed when attention is not directed toward the relevant location (Rensink, 2002).

These findings collectively motivated the development of the flicker paradigm, which provided a controlled method for isolating attentional contributions to change detection (Rensink et al., 1997). In this task, an original image and a modified version alternate repeatedly, separated by brief blank screens that eliminate the motion transients normally associated with change. Under these conditions, observers must rely on effortful, attention-driven search to detect differences between images. Change detection is significantly impaired in this paradigm, with longer detection times compared to conditions without visual interruption (Rensink et al., 1997). Importantly, changes to objects of central interest were detected more rapidly than those of marginal interest, highlighting the role of attentional prioritization in guiding visual search (Rensink et al., 1997). Subsequent research has shown that change blindness is a robust phenomenon that extends beyond the original flicker paradigm. Change detection failures have been observed even in the absence of global visual disruptions when transient signals are masked (O'Regan et al., 1999). Similar effects have also been demonstrated in real-world social interactions where participants fail to detect changes in identity (Simons & Levin, 1998). These findings suggest that change blindness reflects a broader limitation in visual attention rather than being tied to a specific experimental setup.

Despite the strong influence of the flicker paradigm on theories of visual attention, direct replications remain rare in psychological science more broadly (Open Science Collaboration, 2015; Simons, 2014), and the original findings of Rensink et al. (1997) are no exception. Although several studies have employed similar approaches and obtained comparable results, providing conceptual support for the robustness of the change blindness effect (Aginsky & Tarr, 2000; Rensink, 2000, 2002), a near-exact replication of the original flicker paradigm has not, to our knowledge, been conducted. Direct replications are therefore essential for assessing the reliability and generalizability of specific experimental findings, particularly in light of broader concerns regarding reproducibility in psychology (Open Science Collaboration, 2015).

The present study aimed to address this gap by partially replicating Experiment 1 of Rensink et al. (1997) under modified testing conditions, focusing on the core manipulation of flicker versus no-blank conditions and the central versus marginal interest distinction. The sample was expanded from 10 to 72 participants, given that small sample sizes are known to reduce the probability of detecting true effects and to systematically inflate effect size estimates (Button et al., 2013). The experiment was administered online, a format that provides access to larger and more demographically diverse samples than traditional laboratory recruitment typically affords (Gosling & Mason, 2015). Successful replication under such conditions would itself speak to the robustness of the original effect. The original stimulus set was obtained directly from the authors, ensuring fidelity to the original materials.

Based on the findings of Rensink et al. (1997) and the broader change blindness literature, two hypotheses were tested. First, participants in the flicker (EXP) condition were expected to require significantly longer reaction times to detect changes compared to participants in the no-blank (CON) condition. Second, changes to objects of central interest (CI) were expected to be detected more rapidly than changes to objects of marginal interest (MI), and this difference was expected to be amplified under flicker conditions.

Method

Participants and recruitment

The total recruited sample comprised 72 participants assigned to one of two conditions: the flicker condition (EXP; $n = 40$) or the no-blank condition (CON; $n = 32$). Participants were assigned to condition using a pre-generated alternating ID list (EXP, CON, EXP, CON, …). Recruiters distributed URLs without knowing which condition each URL corresponded to, and the lead experimenter did not know which participant had received which URL. This double-blind allocation procedure eliminated the possibility of recruiter-driven selection bias commonly associated with non-random assignment. Three participants in the EXP group were excluded due to a technical error in which reaction times were not recorded by the software despite valid textual responses being provided, yielding a final sample of 69 participants (EXP: $n = 37$; CON: $n = 32$). Among these, 65 completed the sociodemographic questionnaire. The four participants who did not complete this questionnaire were nonetheless retained in all performance analyses but excluded from demographic analyses, resulting in sample sizes of 35 (EXP) and 30 (CON) for demographic data. Among the participants who completed the sociodemographic questionnaire, the majority identified as female (56.9%), with the remainder identifying as male (43.1%). The majority fell within the 18–24 age range (67.7%), with participants aged 25–34 (4.6%), 35–44 (6.2%), 45–54 (15.4%), and 55–64 (6.2%) comprising the remainder. The sample was predominantly French-speaking (60.0%), with smaller proportions reporting English (24.6%) or another native language (15.4%). Most participants reported normal or corrected-to-normal vision (76.9%). Two participants in the EXP group reported a history of epilepsy; their data were retained as their performance did not differ markedly from the remainder of the sample. Full demographic information by condition is presented in Table A.1 at the end. Recruitment was conducted in person in public settings and online by undergraduate students enrolled in the PSY 3777 course during the Fall 2025 semester. Prior to participation, all participants provided informed consent. The study was conducted in accordance with prevailing ethical standards. No strict inclusion or exclusion criteria were applied. Sociodemographic variables (including sex, age group, native language, education level, student status, normal or corrected vision, history of epilepsy, attention or memory disorders, head injuries, recent and typical sleep duration, alcohol or drug use within the past 24 hours, and prior participation in perception studies) were collected for descriptive purposes only. Descriptive statistics for participant characteristics are presented in Table A.1.

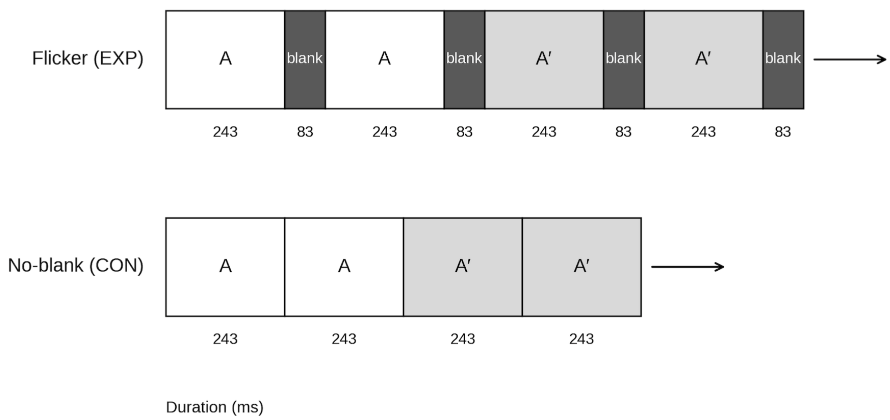
Materials

The study was conducted entirely online and asynchronously using jsPsych (version 7.3.2). Access to a computer equipped with a screen of at least 11 inches and high resolution was required. Stimuli were presented in full-screen mode. Consistent with the flicker paradigm introduced by Rensink et al. (1997), the task was a change detection task in which participants identified a modification between two versions of the same scene. The original set of 24 color image pairs depicting real-world scenes was obtained directly from R. A. Rensink (personal communication), along with the original classification of each image as central interest (CI) or marginal interest (MI). In the original study, interest was determined via an independent experiment in which five naive observers provided brief verbal descriptions of each scene: CIs were defined as objects or areas mentioned by three or more observers, whereas MIs were objects or areas mentioned by none (Rensink et al., 1997). This set comprised 24 original images (A) and 24 modified versions (A'), for a total of 48 images. Changes involved the color, position (translation), or presence/absence of a target item relative to the original version. The 24 pairs were evenly distributed across a 3 (change type: color, translation, presence/absence) × 2 (interest: CI, MI) design, yielding 4 pairs per cell. Because the experiment was administered online on modern LCD screens operating at 60 Hz, stimulus timing was adjusted to conform to the screen refresh rate (1 frame = 16.667 ms). Each image was displayed for approximately 243 ms (14 refresh cycles) and each gray blank screen for approximately 83 ms (5 refresh cycles), closely approximating the 240 ms and 80 ms durations used in the original study. Images were displayed at 700 × 499 pixels in full-screen mode. Participants responded using a standard computer keyboard. Two versions of the task were administered via separate URLs: a flicker version with gray screens interleaved between images (EXP condition) and a no-blank version without interleaved screens (CON condition). The independent variables included the experimental condition (EXP vs. CON; between-subjects), the type of visual change (color, translation, presence/absence; within-subjects), and the location of the change (CI or MI; within-subjects). The dependent variables were reaction time, measured in milliseconds, and detection accuracy.

Procedure

The study employed a between-subjects posttest-only design. The between-subjects independent variable was the presence of the flicker paradigm. In the flicker condition, a gray screen of approximately 83 ms was inserted between each image presentation in the alternating sequence. In the no-blank condition, images alternated without any interleaved gray screen. Each participant was exposed to only one condition, with assignment determined by participant ID as described above. Stimulus pairs were presented in pseudorandom order within each condition. Participants were instructed to complete the task on a computer (not a mobile device) in a quiet environment to minimize distractions. After reading and completing the informed consent form, participants received distinct links to the sociodemographic questionnaire and the experimental task. Participants then read instructions describing the general procedure, which indicated that they should examine the presented scene and press the spacebar as soon as they detected a change, followed by a written description of the change in a text box. This text box served to improve the precision of the detection measure and provided a brief pause between image pairs, reducing the likelihood of eye fatigue interfering with performance (Sheppard & Wolffsohn, 2018; Wolffsohn et al., 2023). Each trial began with a fixation screen (1,000 ms) followed by a gray screen (1,000 ms). In the flicker condition, the alternation sequence followed the pattern: A (243 ms) – gray (83 ms) – A (243 ms) – gray (83 ms) – A' (243 ms) – gray (83 ms) – A' (243 ms)

Figure 1 ■ Schematic of the Trial Sequence in the Flicker and No-blank Conditions. In the flicker condition (top), each image was followed by an 83 ms gray blank screen. In the no-blank condition (bottom), images alternated directly without interleaved blanks. Image frames (A and A') were displayed for approximately 243 ms each. The cycle continued until the participant pressed the spacebar or 59 s elapsed. Frame widths are drawn proportionally to actual stimulus durations.



– gray (83 ms), repeating until the participant pressed the spacebar or a maximum of approximately 59 s had elapsed (see Figure 1 for a schematic of the trial sequence). In the no-blank condition, images A and A' alternated in the same A, A, A', A' sequence at the same display duration (243 ms each) but without the interleaved gray screens, such that each image was presented twice consecutively before the alternation occurred, matching the total exposure time of the flicker condition. The inter-trial interval was 2,000 ms. A practice phase of six trials preceded the experimental phase to familiarize participants with the task. Because no additional stimulus images were available beyond the 24 experimental pairs, the six practice trials were drawn from the experimental pool: for each of the six cells of the 3 (change type) × 2 (interest) design, one image pair was randomly selected for practice and subsequently excluded from the experimental phase. Consequently, each participant completed 6 practice trials followed by 18 experimental trials (3 trials per cell). The transition between the practice and experimental phases was indicated by instructions displayed on a gray screen. Following the spacebar response, the last displayed image (A or A') was frozen on screen, and participants were asked to provide a written description of the change in a text box (minimum 10 characters, maximum 280 characters). Participants then clicked "Continue" to proceed to the next trial. After all trials were completed, participants rated their perceived performance on a 5-point scale, collected for descriptive purposes. A gray screen then displayed a message indicating the end of the study along with contact information for any questions. The entire session lasted approximately 15 to 20 minutes per participant.

Data Preparation

Three participants in the EXP group were excluded from all analyses because a software error resulted in reaction times not being recorded for the majority of their trials (17–18 out of 24 trials), despite valid textual responses having been provided. The final analyzed sample comprised 69 participants (EXP: $n = 37$; CON: $n = 32$). For each trial, detection accuracy was coded by a team of 6 judges (the manuscript co-authors) who evaluated participants' written descriptions to determine whether the change had been correctly identified. Coding categories were: correct detection (hit), incorrect detection (false positive), and no detection (miss). All judges agreed on all coding decisions. Trials in which no response was given within the time limit were coded as misses and excluded from reaction time analyses. The six practice trials per participant were excluded from all analyses. All data preparation and statistical analyses were conducted using IBM SPSS Statistics Version 31, 2025.

Results

Preliminary Analyses

As described in the Data Preparation section, the final analyzed sample comprised 69 participants: 37 in the experimental (EXP) condition and 32 in the control (CON) condition. Four participants (two per group) did not complete the demographic questionnaire; their performance data were retained in all reaction time and accuracy analyses. Descriptive statistics for participant demographics are presented in Table A.1.

To verify that the two groups were comparable on demographic characteristics, chi-square tests of independence were conducted comparing the EXP and CON groups on sex, vision status, attention/memory disorders, alcohol/drug use in the past 24 hours, and prior participation in visual perception studies. Mann-Whitney U tests were used for ordinal variables, including age group and sleep duration. None of these comparisons reached statistical significance, the two groups did not differ significantly on any measured demographic variable (all ps $> .05$).

Prior to conducting inferential analyses, the normality of reaction time distributions was assessed using the Shapiro-Wilk test for each dependent variable, separately by group (see Table A.2). Results indicated significant departures from normality for the CON group across all reaction time variables ($p < .001$), while the EXP group showed mixed results, with some variables meeting the assumption of normality (e.g., overall mean RT: $W = .958$, $p = .172$; marginal: $W = .982$, $p = .785$) and others violating it (e.g., present/absence – central: $W = .769$, $p < .001$). Given these consistent departures from normality in the CON group, Mann-Whitney U tests were selected as the primary inferential approach for group comparisons. However, because parametric tests such as the independent-samples t test and analysis of variance (ANOVA) are generally considered robust to violations of normality with sample sizes exceeding 30 per group, parametric analyses were also conducted to verify the consistency of results. Both approaches yielded the same pattern of findings; parametric results are reported alongside non-parametric tests for completeness.

Overall Reaction Time: EXP Versus CON

The primary analysis compared overall mean reaction time between the EXP and CON groups. Descriptive statistics for all reaction time variables by group are presented in Table 1, and the overall group comparison is illustrated in Figure 2. The EXP group ($M = 8.47s$, $SD = 2.69$) took on average $4.15s$ longer per trial to detect changes than the CON group ($M = 4.33s$, $SD = 3.43$); mean difference = $4.15s$, 95% CI [2.67, 5.62], a large effect (Cohen's $d = 1.36$). A Mann-Whitney U test yielded $U = 150.00$, $Z = -5.318$, $p < .001$ (see Table 2). An independent-samples t test corroborated this result, $t(67) = 5.623$, $p < .001$ (see Table A.3). On average, participants in the EXP condition required approximately twice as long as those in the CON condition to identify changes. A sensitivity analysis excluding the four participants who did not complete the demographic questionnaire produced a consistent result ($U = 107.00$, $Z = -5.500$, $p < .001$).

Mixed ANOVA: Change Type, Location, and Condition

To examine the joint effects of change type, change location, and experimental condition on reaction time, a 3 (change type: color, translation, present/absence) × 2 (location: marginal, central) × 2 (condition: EXP, CON) mixed-design ANOVA was conducted, with change type and location as within-subjects factors and condition as the between-subjects factor. Mauchly's test indicated that the assumption of sphericity was met for change type, $W = .946$, $\chi^2(2) = 3.66$, $p = .160$, and for the change type × location interaction, $W = .939$, $\chi^2(2) = 4.13$, $p = .127$; sphericity-assumed values are therefore reported. Levene's test of equality of error variances was significant for two of the six dependent variables (color-marginal, $F(1;67) = 12.30$, $p < .001$; translation-marginal, $F(1;67) = 7.76$, $p = .007$), indicating unequal variances between groups for those conditions. However, because the primary inferential approach used non-parametric Mann-Whitney U tests, which do not assume homogeneity of variance, this violation does not affect the main conclusions. Full results are presented in Table 3.

The between-subjects main effect of condition was significant, $F(1;67) = 32.38$, $p < .001$, $\eta_p^2 = .33$, confirming that the EXP group was significantly slower overall than the CON group. Within subjects, there was a significant main effect of change type, $F(2;134) = 12.50$, $p < .001$, $\eta_p^2 = .16$, indicating that reaction times varied across color, translation, and present/absence changes. There was also a significant main effect of location, $F(1;67) = 80.57$, $p < .001$, $\eta_p^2 = .55$, which was the largest effect observed in the analysis: changes in marginal areas of the scene took considerably longer to detect than changes in central areas. All two-way interactions were significant. The Change Type × Condition interaction, $F(2;134) = 9.62$, $p < .001$, $\eta_p^2 = .13$, indicated that the magnitude of the group difference varied by change type, with translation changes showing the largest EXP–CON discrepancy. The Location × Condition interaction, $F(1;67) = 49.81$, $p < .001$, $\eta_p^2 = .43$, revealed that the difference between marginal and central changes was substantially larger in the EXP group (Mean difference = 5.43 s) than in the CON group (Mean difference = 0.74 s). The Change Type × Location interaction was also significant, $F(2;134) = 16.50$, $p < .001$, $\eta_p^2 = .20$, indicating that the advantage for central over marginal changes was not uniform across change types. The three-way interaction of change type × location × condition was also significant, $F(2;134) = 5.36$, $p = .006$, $\eta_p^2 = .07$, indicating that the pattern of change type × location differences was itself modulated by experimental condition. This interaction is consistent with the follow-up analyses reported below, which reveal that the EXP–CON group difference was present across nearly all conditions except one: present/ab-

Table 1 ■ Descriptive Statistics for Mean Reaction Times (Seconds) by Experimental Condition

	EXP ($n = 37$)		CON ($n = 32$)	
Variable	$M(s)$	SD	$M(s)$	SD
Overall				
Mean RT	8.47	2.69	4.33	3.43
By change type				
Color	8.06	3.70	4.27	3.84
Translation	10.41	3.72	4.50	3.53
Present/absence	7.03	2.98	4.21	3.58
By location				
Marginal	11.22	3.78	4.72	3.93
Central	5.79	2.61	3.98	3.17
By change type × location				
Color – marginal	8.64	5.22	3.99	3.10
Color – central	7.46	4.30	4.48	5.01
Translation – marginal	14.80	6.32	5.32	4.54
Translation – central	6.40	3.33	3.71	2.76
Present/absence – marginal	10.50	4.53	4.75	5.24
Present/absence – central	3.61	2.40	3.90	3.56

Note. Values reported in seconds. EXP = experimental (flicker) condition; CON = control (no-blank) condition. M = mean; SD = standard deviation.

Table 2 ■ Mann-Whitney U Test Comparing Overall Mean Reaction Time Between Groups

Variable	EXP Mean Rank	CON Mean Rank	U	Z	p
Mean RT (overall)	46.95	21.19	150.00	-5.318	$< .001$

Note. EXP = experimental (flicker) condition; CON = control (no-blank) condition. A higher mean rank indicates slower reaction times.

sence changes in central locations.

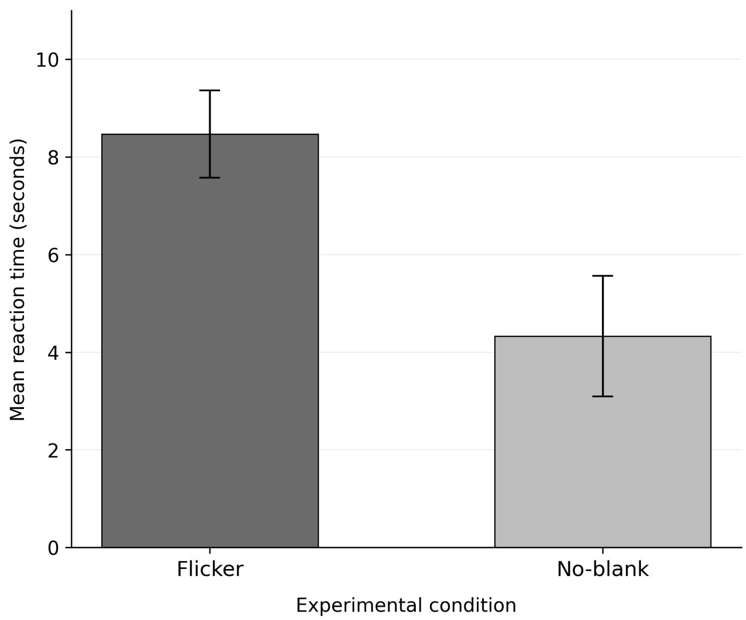
Follow-Up Comparisons by Condition

To decompose the significant interactions, Mann-Whitney U tests were conducted comparing EXP and CON groups on each combination of change type and location (see Table 4 and Figure 3). All comparisons were significant at $p < .001$, with one notable exception: present/absence changes in central locations did not differ between groups ($U = 515.00$, $Z = -0.93$, $p = .354$). For this condition, the EXP group ($M = 3.61s$) and the CON group ($M = 3.90s$) performed nearly identically, indicating that this particular type of change was equally detectable regardless of experimental condition. The largest group difference was observed for translation changes in marginal locations, where the EXP group ($M = 14.80s$) required nearly three times as long as the CON group ($M = 5.32s$) to detect the change, $U = 120.00$, $Z = -5.68$, $p < .001$. Independent-samples t tests confirmed the same pattern of significance across all conditions (see Table A.3). When collapsed across change types, both marginal ($U = 113.00$, $Z = -5.76$, $p < .001$) and central ($U = 290.00$, $Z = -3.63$, $p < .001$) changes showed significant group differences, though the difference was considerably larger for marginal changes (mean difference = 6.50 s) than for central changes (mean difference = 1.81 s). Similarly, when collapsed across locations, all three change types showed significant group differences: color ($U = 209.00$, $p < .001$), translation ($U = 138.00$, $p < .001$), and present/absence ($U = 228.00$, $p < .001$).

Marginal Versus Central Changes

To further examine the effect of change location, Wilcoxon signed-rank tests were conducted comparing marginal and central changes across all participants (see Table 5). Reaction times were significantly faster for central changes than for marginal changes, Z = -6.62, p < .001, with 62 of 69 participants showing this pattern. Detection accuracy was also significantly higher for central changes, Z = -4.56, p < .001, with 30 participants detecting more central changes and only 3 detecting more marginal changes (36 ties).

Figure 2 ■ Mean Reaction Time in seconds by Experimental Condition for change detection in the flicker (EXP) and no-blank (CON) conditions. Error bars represent 95% confidence intervals. EXP $n = 37$; CON $n = 32$.



Accuracy Analyses

Mann-Whitney U tests were conducted to compare detection accuracy between the EXP and CON groups across all conditions (see Table 6). Overall hit rates did not differ significantly between the EXP ($Mhits = 16.54$ out of 18) and CON groups, $U = 476.00$, $Z = -1.49$, $p = .136$. Similarly, no significant group differences were found for false positives ($U = 456.50$, $p = .075$) or misses ($U = 521.50$, $p = .182$). When examined by condition, only one comparison reached significance: the EXP group detected significantly more color changes than the CON group ($U = 466.00$, $Z = -2.06$, $p = .039$). All other accuracy comparisons by change type, location, and their combination were non-significant. The absence of significant group differences in accuracy, combined with the large and consistent group differences in reaction time, indicates that the experimental manipulation primarily affected the speed of change detection rather than the ability to detect changes. Importantly, these results rule out a speed–accuracy trade-off: the CON group was not faster because they were less careful or less accurate. Rather, both groups achieved comparable detection rates, but participants in the EXP condition required substantially more time to do so. The one significant accuracy difference, whereby the EXP group detected more color changes than the CON group, is an unexpected finding that warrants further consideration in the discussion.

Perceived Performance

At the end of the task, participants rated their perceived performance on a 5-point scale. Ratings were similar across the EXP ($M = 4.51$, $SD = 0.73$, $Mdn = 5.0$) and CON ($M = 4.47$, $SD = 0.92$, $Mdn = 5.0$) groups, $U = 573.50$, $Z = -0.22$, $p = .797$. The majority of participants in both groups rated their performance at the highest level (EXP: 62.2%; CON: 68.8%), suggesting that participants in both conditions had comparable and generally positive subjective impressions of their performance despite the substantial objective differences in reaction time

Summary of Findings

Taken together, these results reveal a clear and robust pattern. First, the experimental condition significantly increased reaction times for change detection, with a large effect size (Cohen's d = 1.36). Second, changes in marginal locations were significantly more difficult to detect than changes in central locations, and this location effect was substantially amplified in the EXP group relative to the CON group. Third, change type modulated reaction times, with translation changes eliciting the longest detection times, particularly in marginal locations. Fourth, accuracy did not differ between groups, indicating that the experimental

Table 3 ■ Mixed ANOVA Results: Change Type (3) × Location (2) × Condition (2)

Source	F	df	p	η_p^2
Between-subjects				
Condition (EXP vs. CON)	32.378	1; 67	$< .001$	.326
Within-subjects: Main effects				
Change type	12.502	2; 134	$< .001$	.157
Location	80.566	1; 67	$< .001$	.546
Within-subjects: Two-way interactions				
Change type × Condition	9.619	2; 134	$< .001$	.126
Location × Condition	49.809	1; 67	$< .001$	.426
Change type × Location	16.496	2; 134	$< .001$	.198
Within-subjects: Three-way interaction				
Change type × Location × Condition	5.358	2; 134	.006	.074

Note. Sphericity assumed. η_p^2 = partial eta squared.

Table 4 ■ Mann-Whitney U Tests Comparing Mean Reaction Times Between Groups by Condition

Variable	U	Z	p
By change type × location			
Color – marginal	237.00	-4.272	$< .001$
Color – central	282.00	-3.730	$< .001$
Translation – marginal	120.00	-5.679	$< .001$
Translation – central	265.00	-3.935	$< .001$
Present/absence – marginal	142.00	-5.415	$< .001$
Present/absence – central	515.00	-0.927	.354
By location			
Marginal	113.00	-5.764	$< .001$
Central	290.00	-3.634	$< .001$
By change type			
Color	209.00	-4.609	$< .001$
Translation	138.00	-5.463	$< .001$
Present/absence	228.00	-4.380	$< .001$

Note. Grouping variable: Condition (EXP vs. CON). Two-tailed significance reported. All comparisons were significant at $p < .001$ except present/absence – central ($p = .354$).

manipulation affected processing speed rather than detection ability. Finally, present/absence changes in central locations were the only condition that showed no significant group difference in reaction time.

Discussion

The present study partially replicated Experiment 1 of Rensink et al. (1997), to assess the robustness of the flicker paradigm as a measure of change blindness. The core findings of the original study were largely replicated: the flicker condition significantly increased reaction times for change detection, changes to marginal interest objects were harder to detect than changes to central interest objects, and accuracy did not differ between conditions.

Comparison With the Original Study

A direct comparison of the present results with those of Rensink et al. (1997), reveals both convergences and divergences. In the flicker condition, mean detection times for marginal interest changes (11.22 s) closely approximated the original values (10.9 s), and central interest changes (5.79 s) were also comparable to the original (4.7 s; see Figure 4). The CI advantage over MI was replicated across all change types, and the overall pattern of results, including the interaction between change type, location, and condition, was consistent with the original findings.

However, a notable divergence emerged in the no-blank condition: participants in the present study required 4.33 s on average to detect changes, compared to only 0.9 s in Rensink et al. (1997). Furthermore, whereas Rensink

Figure 3 ■ Mean Reaction Time (in seconds) in the flicker (left panel) and no-blank (right panel) conditions, broken down by change type (Color, Translation, Presence/Absence) and location (Central vs. Marginal). Error bars represent 95% confidence intervals. Flicker $n = 37$; No-blank $n = 32$.

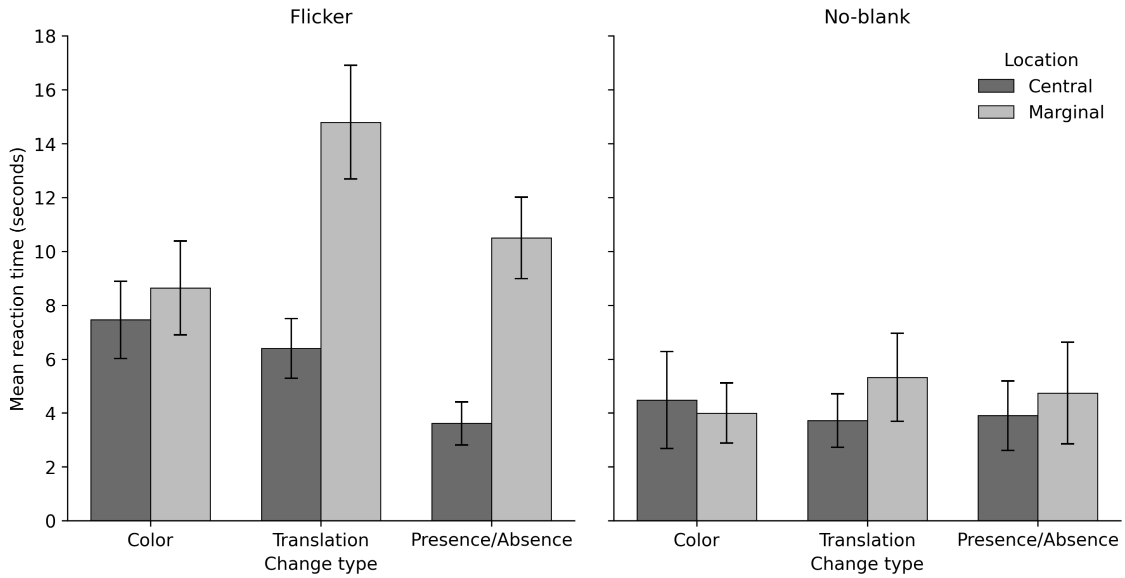


Table 5 ■ Wilcoxon Signed-Rank Tests Comparing Marginal Versus Central Changes

Variable	Direction	Z	p
Reaction time	Central < Marginal (62 of 69)	-6.616	$< .001$
Accuracy (hits)	Central > Marginal (30 of 69)	-4.560	$< .001$

Note. $N = 69$. For reaction time, 62 participants were faster for central than marginal changes. For accuracy, 30 participants had more hits for central changes; 36 participants had tied scores.

et al. (1997) found no significant difference between CI and MI in the no-blank condition, the present study found that marginal changes (4.72 s) were still somewhat slower than central changes (3.98 s), although this difference was considerably smaller than in the flicker condition. This discrepancy is particularly noteworthy given that motion is generally considered a pop-out feature in visual search, capable of capturing attention rapidly and automatically (Wolfe & Horowitz, 2004). In the original laboratory-based study, changes in the no-blank condition were detected in less than one second on average, suggesting that they may have been captured through rapid, bottom-up attentional processes. The substantially longer detection times observed here suggest that the online testing environment may have disrupted the conditions necessary for efficient pop-out. Online administration introduces variability in screen size, viewing distance, ambient lighting, and hardware latency that can elevate baseline reaction times relative to laboratory conditions (Crump et al., 2013; Gosling & Mason, 2015). Additionally, sources of attentional distraction inherent to unsupervised online testing, such as browser notifications or divided attention, may have interfered with the automatic capture of attention by motion transients. A direct comparison of in-person and online versions of the no-blank condition would clarify whether the pop-out effect is preserved under controlled viewing conditions, or whether it is attenuated by factors associated with remote testing.

An additional divergence concerns the single non-significant comparison: present/absence changes in central locations did not differ between the EXP and CON groups. This suggests that highly salient changes may bypass the flicker manipulation through bottom-up attentional capture. They may thus represent a boundary condition for the flicker effect, where salience is sufficient to trigger automatic orienting regardless of visual disruption. Future research could examine this possibility by systematically varying the salience of target changes to identify

Table 6 ■ Mann-Whitney U Tests Comparing Detection Accuracy Between Groups

Variable		U	Z	p
Overall				
	Hit total	476.00	-1.492	.136
	False positives	456.50	-1.783	.075
	Misses	521.50	-1.334	.182
	Accuracy rate	476.00	-1.492	.136
By change type				
	Color	466.00	-2.060	.039
	Translation	503.50	-1.185	.236
	Present/absence	554.50	-0.709	.478
By location				
	Marginal	463.50	-1.676	.094
	Central	506.00	-1.477	.140
By change type × location				
	Color – marginal	500.00	-1.813	.070
	Color – central	501.00	-1.796	.073
	Translation – marginal	529.00	-0.853	.394
	Translation – central	562.00	-0.891	.373
	Present/absence – marginal	520.50	-1.473	.141
	Present/absence – central	568.50	-0.630	.529

Note. Grouping variable: Condition (EXP vs. CON). Two-tailed significance reported. Hit total = number of correctly detected changes. False positives = incorrect identifications. Misses = undetected changes. Accuracy rate = hit total / 18.

the threshold at which flicker-induced change blindness breaks down.

Theoretical Interpretation

Although the present results are consistent with the original findings, several of Rensink et al. (1997) theoretical claims have since been refined or challenged, including by the authors themselves (Rensink, 2018; Simons & Rensink, 2005). Rensink et al. (1997) proposed that observers never form a coherent visual representation of a scene and that unattended items are simply overwritten by subsequent stimuli. This proposal is consistent with subsequent findings that individuals can retain approximately four items in visual working memory (Cowan, 2001; Luck & Vogel, 1997), which corresponds to the number of items that can be simultaneously attended (Rensink, 2000). As Rensink (2000) noted, the distinction between short-term visual memory and attention remains difficult to draw, and information may be extracted from the scene and entered into memory via a series of attentional shifts or eye movements, potentially forming the basis for longer-lasting scene representations (Hollingworth & Henderson, 2002; Hollingworth et al., 2001; Vierck & Kiesel, 2008).

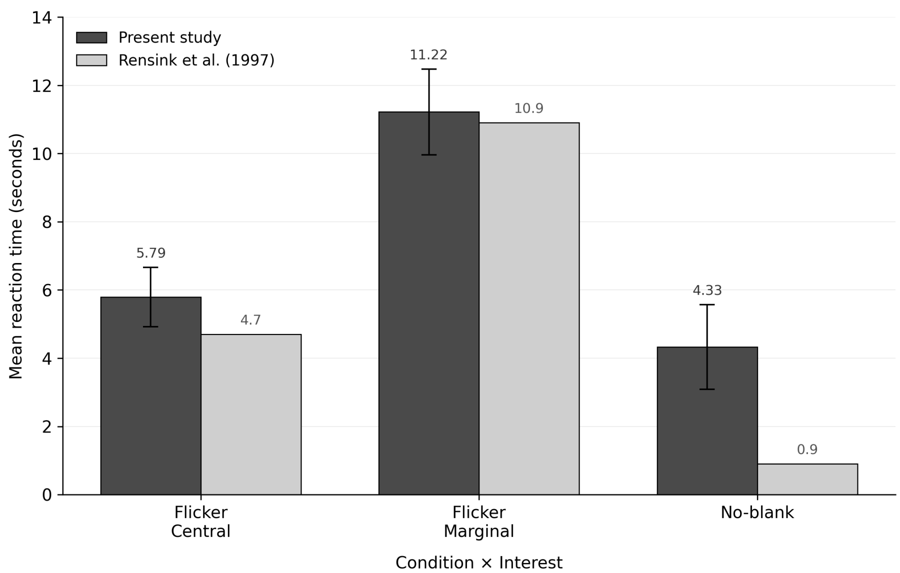
Rensink et al. (1997) further argued that attention was the key factor in the visual perception of change. While the present results are consistent with this claim, subsequent work has shown that the relationship between attention and change detection is more complex than originally proposed. It has been demonstrated that individuals can report sensing a change without being able to identify it, suggesting that seeing and sensing change may involve separate mechanisms (Rensink, 2004). Current models propose that change detection requires at least three processes: attention, representation, and comparison (Simons, 2000).

Strengths and Limitations

The present study offers several strengths as a replication effort. The sample size ($N = 69$) substantially exceeds that of the original study ($N = 10$), increasing statistical power and the precision of effect estimates. The use of the original stimulus set obtained directly from Rensink, with the original CI/MI classification, ensures high fidelity to the original materials. Additionally, the written description requirement and independent coding by a team of 6 judges provided a more rigorous measure of detection accuracy than verbal report alone.

Several limitations must also be acknowledged. First, the online administration introduced uncontrolled variability in viewing conditions (screen size, resolution, viewing distance, ambient lighting), which may have influenced detection performance. Second, technological heterogeneity across participants, including variability in browser la-

Figure 4 ■ Comparison of Present Study with Rensink et al. (1997): Mean reaction time (in seconds) for change detection in the present study (dark bars) and the original Rensink et al. (1997) study (light bars), shown for the Flicker condition split by interest level (Central, Marginal) and for the No-blank condition. Error bars represent 95% confidence intervals; bars are unavailable for the Rensink et al. (1997) values.



tency, monitor refresh rates, and overall hardware performance, may have introduced additional noise into stimulus presentation timing and reaction time measurements. Although this variability is unlikely to have biased the between-group comparison given random allocation and the large effect sizes observed, it may have inflated within-group variance and limited the precision of individual reaction time estimates. Third, the present study used only 24 of the original 48 image pairs, with 18 experimental trials per participant (3 per cell), limiting the reliability of cell-level comparisons. This reduction was nonetheless a deliberate tradeoff to minimize fatigue effects in an unsupervised online setting. Fourth, no strict inclusion or exclusion criteria were applied.

Future Directions

Future research should address several questions raised by the present findings. First, the substantial divergence in no-blank condition performance between the present study and the original warrants investigation into how online administration affects baseline change detection. A direct comparison of in-person and online versions of the flicker paradigm using the same stimulus set would clarify the extent to which viewing conditions modulate the phenomenon. Second, the non-significant group difference for present/absence changes in central locations could be further examined by manipulating the salience of changes systematically to identify the threshold at which flicker-induced change blindness breaks down. Third, replication of Experiments 2 and 3 of Rensink et al. (1997), which examined extended display duration and verbal cueing respectively, would provide a more complete assessment of the original findings.

Conclusion

In sum, this partial replication supports the robustness of the flicker paradigm as a tool for investigating change blindness. The core findings of Rensink et al. (1997), were replicated with a larger sample and in an online format, demonstrating that flicker-induced visual interruption is sufficient to impair change detection even under less controlled conditions. The unexpected finding that the EXP group detected significantly more color changes than the CON group in accuracy analyses warrants further investigation, as it may reflect a response strategy difference between conditions rather than a true perceptual effect. The present study does not directly test the role of attention in change detection, which remains a complex and actively debated topic in the field (Rensink, 2002; Simons, 2000). The divergence in no-blank performance highlights the importance of considering methodological context when comparing results across studies and underscores the need for continued replication efforts in cognitive psychology.

Authors' note

All authors contributed to data collection as part of an undergraduate course project. Alexandre, E., Gagnon, M., Joanette, F., Lamarre, É., and Mohamed Guirreh, S. contributed equally to the preparation of this manuscript. Ouattara, M.M. conceptualized and supervised the project, led the design and implementation of the experimental paradigm, and had primary responsibility for the writing and preparation of this manuscript. All authors read and approved the final version of this manuscript. No funding was received for this work. The data and accompanying documentation for this experiment have been made available online: https://osf.io/ef769/.

References

Aginsky, V., & Tarr, M. J. (2000). How are different properties of a scene encoded in visual memory?: Change blindness and visual memory. *Visual Cognition*, *7*(1-3), 147–162.

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. doi: 10.1038/nrn3475.

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. doi: 10.1017/S0140525X01003922.

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating amazon's mechanical turk as a tool for experimental behavioral research. *PLOS ONE*, *8*(3), e57410. doi: 10.1371/journal.pone.0057410.

Gosling, S. D., & Mason, W. (2015). Internet research in psychology. *Annual Review of Psychology*, *66*, 877–902. doi: 10.1146/annurev-psych-010814-015321.

Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(1), 113–136. doi: 10.1037/0096-1523.28.1.113.

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review*, *8*(4), 761–768. doi: 10.3758/BF03196215.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. doi: 10.1038/36846.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science [aac4716]. *Science*, *349*(6251). doi: 10.1126/science.aac4716.

O'Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change blindness as a result of "mudsplashes." *Nature*, *398*(6722), 34. doi: 10.1038/17953.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*(2), 283–290. doi: 10.3758/BF03203943.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. doi: 10.1080/00335558008248231.

Rensink, R. A. (2000). The dynamic representation of scenes. *Visual Cognition*, *7*(1-3), 17–42. doi: 10.1080/135062800394667.

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*, 245–277. doi: 10.1146/annurev.psych.53.100901.135125.

Rensink, R. A. (2004). Visual sensing without seeing. *Psychological Science*, *15*(1), 27–32. doi: 10.1111/j.0963-7214.2004.01501005.x.

Rensink, R. A. (2018). To have seen or not to have seen: A look at rensink, o'regan, and clark (1997). *Perspectives on Psychological Science*, *13*(2), 230–235. doi: 10.1177/1745691617707269.

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368–373. doi: 10.1111/j.1467-9280.1997.tb00427.x.

Sheppard, A. L., & Wolffsohn, J. S. (2018). Digital eye strain: Prevalence, measurement and amelioration. *BMJ Open Ophthalmology*, *3*(1), e000146. doi: 10.1136/bmjophth-2018-000146.

Simons, D. J. (2000). Current approaches to change blindness. *Visual Cognition*, *7*(1-3), 1–15. doi: 10.1080/135062800394658.

Simons, D. J. (2014). The value of direct replication. *Perspectives on Psychological Science*, *9*(1), 76–80. doi: 10.1177/1745691613514755.

Simons, D. J., & Ambinder, M. S. (2005). Change blindness: Theory and consequences. *Current Directions in Psychological Science*, *14*(1), 44–48. doi: 10.1111/j.0963-7214.2005.00332.x.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267. doi: 10.1016/S1364-6613(97)01080-2.

Simons, D. J., & Levin, D. T. (1998). Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review*, *5*(4), 644–649. doi: 10.3758/BF03208840.

Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, *9*(1), 16–20. doi: 10.1016/j.tics.2004.11.006.

Vierck, E., & Kiesel, A. (2008). Change detection: Evidence for information accumulation in flicker paradigms. *Acta Psychologica*, *127*(2), 309–323. doi: 10.1016/j.actpsy.2007.06.004.

Wolfe, J. M., & Horowitz, T. S. (2004). Opinion: What attributes guide the deployment of visual attention and how do they do it? nature reviews. *Neuroscience*, *5*(6), 495. doi: 10.1038/nrn1411.

Wolffsohn, J. S., Lingham, G., Downie, L. E., Huntjens, B., Inomata, T., Jivraj, S., Kobia-Acquah, E., Muntz, A., Mohamed-Noriega, K., Plainis, S., Read, M., Sayegh, R. R., Singh, S., Utheim, T. P., & Craig, J. P. (2023). TFOS lifestyle: Impact of the digital environment on the ocular surface. *The Ocular Surface*, *28*, 213–252. doi: 10.1016/j.jtos.2023.04.004.

Appendix A

Tables A1, A2 and A3 follows in the next pages.

Open practices

The *Open Data* badge was earned because the data of the experiment(s) are available on osf.io/ef769/

The *Open Material* badge was earned because supplementary material(s) are available on osf.io/ef769/

Citation

Ouattara, M. M., Alexandre, E., Gagnon, M., Joanette, F., Lamarre, É., Mohamed Guirreh, S., Aboua Adon, O., Bellout, A., Billings, B., Bisson, G., Bouchard, V., Boucher, M., Boundzanga, R., Brousseau, P., Clark, L., Connors-Lee, E., Dabal, E., Daho, D., Gendron, A., … Tangba, Y. (2026). A replication of Rensink, O'Regan, and Clark (1997): Change blindness in an online flicker paradigm. *The Quantitative Methods for Psychology*, *22*(2), 56–70. doi: 10.20982/tqmp.22.2.p056.

Received: 31/05/2026 ∼ Accepted: 15/06/2026

Table A.1 ■ Participant Demographics by Experimental Condition ($N = 65$)

Variable		EXP ($n = 35$)	CON ($n = 30$)	Total ($N = 65$)
Sex				
	Male	17 (48.6%)	11 (36.7%)	28 (43.1%)
	Female	18 (51.4%)	19 (63.3%)	37 (56.9%)
Age group				
	18–24	23 (65.7%)	21 (70.0%)	44 (67.7%)
	25–34	2 (5.7%)	1 (3.3%)	3 (4.6%)
	35–44	2 (5.7%)	2 (6.7%)	4 (6.2%)
	45–54	6 (17.1%)	4 (13.3%)	10 (15.4%)
	55–64	2 (5.7%)	2 (6.7%)	4 (6.2%)
Primary language				
	French	20 (57.1%)	19 (63.3%)	39 (60.0%)
	English	9 (25.7%)	7 (23.3%)	16 (24.6%)
	Other	6 (17.1%)	4 (13.3%)	10 (15.4%)
Highest education				
	Secondary	13 (37.1%)	13 (43.3%)	26 (40.0%)
	College	10 (28.6%)	8 (26.7%)	18 (27.7%)
	Bachelor's	9 (25.7%)	8 (26.7%)	17 (26.2%)
	Master's	2 (5.7%)	0 (0.0%)	2 (3.1%)
	PhD	1 (2.9%)	0 (0.0%)	1 (1.5%)
Currently studying				
	Yes	20 (57.1%)	19 (63.3%)	39 (60.0%)
Normal/corrected vision				
	Yes	30 (85.7%)	20 (66.7%)	50 (76.9%)
History of epilepsy				
	Yes	2 (5.7%)	0 (0.0%)	2 (3.1%)
Attention/memory disorder				
	Yes	2 (5.7%)	4 (13.3%)	6 (9.2%)
Head injury				
	Yes	3 (8.6%)	2 (6.7%)	5 (7.7%)
Sleep last night				
	3–4 hours	1 (2.9%)	3 (10.0%)	4 (6.2%)
	5–6 hours	10 (28.6%)	7 (23.3%)	17 (26.2%)
	7–8 hours	18 (51.4%)	18 (60.0%)	36 (55.4%)
	9+ hours	6 (17.1%)	2 (6.7%)	8 (12.3%)
Typical sleep amount				
	Yes	24 (68.6%)	20 (66.7%)	44 (67.7%)
Alcohol/drugs (past 24h)				
	Yes	5 (14.3%)	5 (16.7%)	10 (15.4%)
Prior perception studies				
	Yes	8 (22.9%)	3 (10.0%)	11 (16.9%)

Note. Sixty-nine participants completed the experimental task; four did not complete the demographic questionnaire (EXP: $n = 2$; CON: $n = 2$). Percentages are based on valid responses. "Other" languages include Pulaar, Arabic, Somali, Greek, Darija, Spanish, and Urdu. "Prefer not to respond" responses were treated as missing: education ($n = 1$), epilepsy ($n = 2$), attention disorder ($n = 2$), currently studying ($n = 1$).

Table A.2 ■ Shapiro-Wilk Tests of Normality for Reaction Time Variables by Group

	EXP ($n = 37$)			CON ($n = 32$)		
Variable	W	df	p	W	df	p
Mean RT (overall)	.958	37	.172	.746	32	$< .001$
Color – marginal	.936	37	.034	.749	32	$< .001$
Color – central	.937	37	.037	.589	32	$< .001$
Translation – marginal	.969	37	.392	.758	32	$< .001$
Translation – central	.907	37	.005	.795	32	$< .001$
Present/absence – marginal	.964	37	.262	.571	32	$< .001$
Present/absence – central	.769	37	$< .001$	.706	32	$< .001$
Marginal (overall)	.982	37	.785	.734	32	$< .001$
Central (overall)	.885	37	.001	.744	32	$< .001$
Color (overall)	.964	37	.266	.686	32	$< .001$
Translation (overall)	.948	37	.082	.787	32	$< .001$
Present/absence (overall)	.914	37	.007	.731	32	$< .001$

Note. Significant values ($p < .05$) indicate departure from normality. For data to be considered normal, both groups must have $p > .05$.

Table A.3 ■ Independent Samples t-Tests Comparing Mean Reaction Times (Seconds) Between Groups by Condition

Variable	t	df	p	Mean Diff (s)	95% CI	Cohen's d
Overall						
Mean RT	5.623	67	$< .001$	4.15	[2.67, 5.62]	1.357
By change type × location						
Color Marginal	4.405	67	$< .001$	4.65	[2.54, 6.75]	1.063
Color Central	2.661	67	.010	2.98	[0.75, 5.22]	0.642
Trans. Marginal	7.048	67	$< .001$	9.48	[6.79, 12.16]	1.702
Trans. Central	3.614	67	$< .001$	2.69	[1.20, 4.17]	0.873
Pres./abs. Marginal	4.895	67	$< .001$	5.75	[3.41, 8.10]	1.182
Pres./abs. Central	-0.392	67	.696	-0.28	[-1.73, 1.16]	-0.095
By location						
Marginal	6.989	67	$< .001$	6.50	[4.64, 8.36]	1.687
Central	2.600	67	.011	1.81	[0.42, 3.20]	0.628
By change type						
Color	4.166	67	$< .001$	3.79	[1.97, 5.60]	1.006
Translation	6.742	67	$< .001$	5.91	[4.16, 7.66]	1.627
Present/absence	3.573	67	$< .001$	2.82	[1.24, 4.40]	0.862

Note. Equal variances assumed. Values reported in seconds. 95% CI = 95% confidence interval of the mean difference. Cohen's d reported for overall comparison only; $d > 0.80$ indicates a large effect. Both parametric and non-parametric approaches yielded the same pattern of results.