Brain & Language 276 (2026) 105724

Contents lists available at ScienceDirect

Brain and Language

journal homepage: www.elsevier.com/locate/b&l

Research Article

The neural time course of self-oriented pointing and labeling in visual processing

Spencer D. Kelly *, Dylan Vlasak, Binghui Tang

Center for Language and Brain, Department of Psychological and Brain Sciences, Colgate University

ARTICLE INFO

Keywords:
Gesture
Categorization
Color
Perception
Conceptual
ERP
P1
P2
N400

ABSTRACT

Pointing and labeling are traditionally seen as communicative behaviors; the present study examines their neurocognitive consequences when produced for oneself. Twenty-four participants completed a priming task in which they encoded a color prime in three within-subjects conditions: point, label, and point & label. Following each prime, participants judged whether a target object belonged to the same (congruent) or different (incongruent) color category. Target-locked ERPs showed effects of congruence and encoding conditions spanning early and later processing stages. At early latencies, pointing elicited a more positive posterior P1 than labeling, while the combined point & label condition was intermediate and did not differ reliably from either unimodal condition, consistent with a non-additive interaction at early perceptual-attentional stages. At later latencies, pointing alone elicited a larger posterior P2 and smaller central-parietal late negativity (300–500 ms) than labeling alone, and the combined condition diverged from pointing while closely paralleling labeling. Together, these results indicate that pointing and labeling for oneself exert distinct, non-additive influences on downstream object processing, with early processing showing graded interaction and later perceptual and conceptual processing increasingly influenced by labeling.

1. Introduction

In Japan, railway workers practice a relatively simple, yet incredibly effective safety measure known as “point and call” or *shisa kanko* (Richarz, 2017). For instance, when alone in the head car, a train driver might point to a green signal light and audibly say “green” to confirm that conditions are safe before going to the next task. By heightening attention to safety steps, this practice has reduced railway accidents by 85%, leading to its adoption by other high-risk professions, such as aviation, nuclear power, and medicine (Richarz, 2017; Shigemori et al., 2013; Violato et al., 2022). While the present study does not examine *shisa kanko* directly, we draw inspiration from it and explore how pointing and speaking for oneself may influence the earliest neurocognitive stages of perceiving and conceptualizing objects in the environment.

2. Communicative Framework for Speaking and Pointing

We frame our study using Herb Clark’s (1996) typology of communication. Clark distinguishes three communicative modes: description, indication, and demonstration. In this study, we focus on the first two. A *description* uses symbolic language—in the form of conventional signs—that arbitrarily maps onto referents. For example, saying “the green one” when differentiating a green light and a red light located on a control panel provides a verbal label that distinguishes one referent from another through linguistic convention. An *indication*, by contrast, is deictic: it physically connects a signal to its referent. A classic example is pointing. Pointing to a green light directly anchors the gesture to that specific item, while simultaneously excluding surrounding objects (e.g., the proximal red light) from the referential field. Moreover, combining these two modes forms a more powerful composite signal (Clark, 1996; Ferrara & Hodge, 2018). Pointing to a green light while saying “the green one” is a mutually reinforcing message, which can enhance clarity, precision, and ease of processing.

Traditionally, Clark’s framework has been applied to communicative contexts, but inspired by the practice of *shisa kanko*, the present study explores a more cognitive question: does self-oriented pointing and labeling alter how the producer processes the visual world?

This article is part of a special issue entitled: ‘Gesture, Cognition and Language’ published in Brain and Language.

* Corresponding author.
E-mail address: skelly@colgate.edu (S.D. Kelly).

https://doi.org/10.1016/j.bandl.2026.105724
Received 29 May 2025; Received in revised form 16 January 2026; Accepted 4 February 2026

3. Pointing for the Self

Human pointing is arguably the most basic and ancient form of communicative behavior, perhaps serving as a foundation for language in both ontogeny (Bates et al., 1975) and phylogeny (Tomasello, 2010; but see Pika, 2012, for a competing view). Despite it typically being associated with social communication, there is evidence that pointing can serve a myriad of cognitive functions as well (Church et al, 2017; Kelly, 2024).

Cooperrider (2023) describes the phenomenon of "private pointing," which can be seen in everyday examples, such as counting (pointing to numbers or objects), reading (underlining sentences with one's finger), and remembering (gesturing to one's keys before leaving home). Empirical research has shown the benefits of these sorts of behaviors (Carlson et al., 2007; Goldin-Meadow et al., 2001). For example, Goldin-Meadow et al. (2001) found that explaining hard math problems while pointing to the numbers, thereby externalizing thought, helps children free up cognitive resources while speaking. In a pure counting task, Carlson et al. (2007) found that pointing to asterisks on a computer screen reduces the working memory load and increases the accuracy and speed in which participants count them.

Furthermore, positive effects of pointing are also evident in the domain of information encoding and reading comprehension, suggesting that pointing can make information more salient in attention and memory. For instance, Zhang and colleagues (2023) showed that pointing during a split-attention task (in which participants point to a text and an accompanying graph) was positively associated with information retention. Taking this a step further, Park et al. (2023) used eye-tracking to show that self-oriented pointing effectively guides visual attention when scanning a complex illustration, leading participants to analyze the illustration more deeply, focus their attention more readily, and make more connections between the illustration and text. The attention-directing function of self-oriented pointing is supported by neuroimaging research showing that merely preparing to point activates brain regions involved in oculomotor control and attention (Astafiev et al., 2003). In this way, pointing can be framed as an instrumental tool enhancing the pointer's cognitive processing by physically modeling the direction of one's gaze toward a spatial target of attention (Cappuccio et al., 2013).

Pointing for the self can also enhance spatial memory. Research has found that pointing can activate visuospatial maps in the parietal and superior frontal cortex, brain areas associated with spatial selectivity in the visual field (Hagler et al., 2007). Indeed, other research has shown that when completing a picture-location recall task, both young and older participants performed better when pointing during an encoding phase compared to naming or visual observation alone (Ouwehand et al., 2015).[1] Chum et al. (2007) also observed spatial recall benefits in a task where participants pointed to shapes in an array (in the context of their study, they tapped a touch screen) compared to when they observed or labeled the side of the screen where the shapes appeared (i.e., "left" or "right"). The authors proposed that pointing may help people remember locations better by linking a first-person physical perspective to a third person sense of where objects are in space. Taken together, there is empirical support for the intuitive practice of self-oriented pointing: it can reduce load on working memory, better direct one's visual attention, and improve spatial memory.

4. Labeling for the Self

Self-talk is a common activity (Alderson-Day & Fernyhough, 2015), and if done with intention in the right contexts, it can have powerful cognitive benefits. Self-talk can be framed as a form of self-thinking—a transition from social speech to inner speech that underpins verbal thought (Vygotsky, 1987/1934). This view of self-oriented language fits with embodied cognition theories claiming that language does not merely package thought but actually *constitutes* it (Clark, 2006). In this way, language can be seen as a "neuroenhancer," in the sense that it is a conceptual system in its own right, as opposed to just organizing or "scaffolding" meaning through statistical regularities (Dove, 2019).[2]

There is strong empirical evidence that producing language can affect perceptual and conceptual processes (Lupyan & Swingley, 2012; Overkott & Souza, 2022; Souza & Sko´ra, 2017). Lupyan & Swingley (2012) demonstrated that overt labeling can exert top-down influences on perception in a study where participants searched for objects. Labeling improved the participants' reaction times, especially when the target's perceptual features were more associated with their category (e.g., bananas benefited more from labels than Pop-Tarts, which have assorted colors). In this way, category labels can activate canonical examples of a category, thereby aiding perception for prototypical objects.

In addition to perceptual effects, there is evidence that self-generated labels might bias categorical processes in the domain of working and long-term memory (Bae et al. 2015; Regier & Xu, 2017). Focusing on working memory, Bae et al. (2015) had participants perform a color estimation task in which they viewed a color and attempted to replicate it on a color wheel, either simultaneously or after a 900 ms delay. Although the effect was stronger in the delay condition, participants in both conditions biased their estimation away from category boundaries, suggesting that implicitly generated color labels activated color prototypes and biased perception and working memory towards it. This is consistent with a model adopted by Regier & Xu (2017) who argue that color labels exert increasingly stronger effects on memory for color as time passes.

While the above research focused on internally generated labels, other research has shown that explicitly labeling categories can actually *help* with perception and working memory (Overkott & Souza, 2022; Souza & Sk´ora, 2017). For example, in Souza & Sko´ra (2017), participants either explicitly labeled a color on a screen or suppressed the color label (i.e., by saying "ba ba ba") and then completed a test phase where they tried to reproduce the color on a color wheel under low or high cognitive loads. By using categorical-continuous mixture modeling, they found that although explicit labeling increased the number of categorical representations in working memory in the high load condition, it helped subjects more accurately identify the original color. One explanation for this effect is that overtly producing labels served as an additional "reminder" in working memory, which could give a linguistic boost to the actual perceptual memory trace (Overkott & Souza, 2022).

Regardless of the precise mechanism, the evidence suggests that in addition to augmenting early perceptual processes, self-oriented labeling activates categorical knowledge, which has an impact on the earliest stages of conceptual processing.

5. Pointing and Labeling for the Self

In Clark's (1996) semiotic framework, description and indication work together to produce a stronger social signal in communication. Can this kind of composite signaling also enhance perception and attention for the speaker?

The success of *shisa kanko* suggests that the answer is yes (Kishi et al., 2014). Research on the practice indicates that pointing and calling can decrease reaction times without affecting subjective mental workload or accuracy during a task-switching test (Shinohara et al., 2013), and it can also maintain arousal during a vigilance task (Shigemori et al., 2013).

---

[1] However, see Ouwehand et al. (2019) for contradictory results showing that pointing to objects actually hindered memory for those objects.

[2] Noam Chomsky (2011) goes a step further and actually argues that language is primarily a system of thought and only secondarily a system of communication, but this extreme cognitive position is a matter of debate (Clark, 1996; Fedorenko et al., 2024; Tomasello, 2010).

Interestingly, Shigemori et al. (2013) conducted experiments with only pointing and only calling to better understand their separate contributions to “point and call”. They found that pointing—independent of calling—can suppress impulsivity within a task-switching paradigm, and it can also improve accuracy during a symbol counting task. As for calling, they found that verbally stating the names of colored stimuli aloud during an n-back task improved recall. Thus, the cognitive benefits of pointing and labeling described above might be maintained when the two are used together.

Although no neuroimaging research, to our knowledge, has directly investigated the neurocognitive benefits of pointing and calling, research on the effects of hearing labels and seeing pointing gestures can offer some clues. Using fMRI, Peeters et al. (2017) found that being exposed to other people’s pointing gestures and verbal descriptions activated brain regions associated with the semantic unification of word meaning and world knowledge (left IFG), in addition to higher-order integration processes (pMTG). Moreover, this enhanced pMTG disappeared when pointing was replaced with a different visual cue (i.e., highlighting the object with a green box).

Moving from receiver to producer, there is some evidence that pointing and speaking may play different roles in early planning and attentional phases of production (Peeters et al., 2015). Using event-related potentials (ERPs) in a communication task (indicating individual objects in an array), Peeters and colleagues (2015) found that speaking while pointing produced a smaller central positivity than speaking alone starting 200 ms post-stimulus onset, suggesting that the two modalities interact early and may serve different cognitive functions. Although this study focused on speaking and pointing *for others*, it is interesting to ask how the two modalities interact when produced for oneself.

6. The Present Study

To explore this question, we used ERPs because of their excellent millisecond-to-millisecond temporal resolution. Specifically, we employed a priming paradigm roughly modelled on an ERP study by Boutonnet & Lupyan (2015). The Boutonnet & Lupyan study investigated how hearing a label for a category, like “dog” (verbal labels condition), primed a picture of a dog differently from hearing an actual dog barking (environmental sounds condition). Half of the prime-target pairs were congruent, and the other half were incongruent (e.g., the word “dog” preceded a picture of a guitar). The study focused on three ERP components: the P1 (early visual attention), the P2 (perceptual-matching processes), and the N400 (semantic and conceptual integration). The main finding was that hearing verbal labels for categories, such as “dog,” enhanced the parieto-occipital P1 and P2 to target pictures compared to hearing environmental sounds, such as a dog barking. In contrast, there was no difference between the image and sound conditions for the N400 time window. Boutonnet & Lupyan interpreted these findings to suggest that hearing verbal labels activates categorical knowledge and affects perceptual processing of subsequent objects within 100 ms.

Adapting this basic paradigm and focusing on the same three ERP components, the present study used a priming task in which participants encoded a color prime (a blue or green square) in three different modalities: pointing, labeling, and pointing & labeling. Following each prime, participants had to determine whether a target square, which was a different shade of blue and green, belonged to the same (congruent) or different (incongruent) color category as the prime. In measuring the ERPs to the target objects, we made two predictions.

The first prediction concerned the relationship between the prime and target. We predicted that incongruent targets would evoke a more positive P2 compared to congruent targets (Boutonnet & Lupyan, 2015; Freunberger et al., 2007; Luck & Hillyard, 1994), but based on Boutonnet & Lupyan (2015), we did not expect this congruency effect to extend to the P1. For the N400, we expected a more negative-going peak amplitude for incongruent targets compared to congruent targets based on research showing visually based N400 effects (Kelly et al., 2004; McPherson & Holcomb, 1999; Kutas & Federmeier, 2011; Wu & Coulson, 2005).

The second—and primary—prediction was that pointing and labeling would interact in stage-dependent ways across the ERP time course, such that the relationship between pointing, labeling, and their combination will differ for early and late components.[3] Specifically, we predicted that for early components (P1 and P2), the combined point & label condition would more closely mirror pointing than labeling. This prediction was motivated by prior work suggesting that pointing is associated with individuating objects, which has been linked to early attentional and perceptual processing. In contrast, for the later N400 component, we predicted that the combined condition would more closely mirror labeling than pointing, consistent with evidence that labels are associated with conceptual categorization and semantic integration processes. Finally, given these distinct functions of pointing and labeling, we predicted that the two unimodal conditions would differ across all three stages of processing.

7. Methods

7.1. Participants

We tested 24 participants (12 female, 12 male) recruited from the student body at a university in the US northeast.[4] To reach this sample size, we collected data from 32 participants but excluded three participants for not following task instructions and five others for having noisy EEG data. These participants were all native English speakers and right-handed with a mean handedness score of + 94.2 (Oldfield, 1971; Zhang, 2014). Most participants received course credit for their participation, but two received a $20 payment instead.

8. Materials

The visual stimuli were two prime objects, a green square and a blue square, and eight target objects. There were four targets per color category: four squares of different blue shades and four squares of different green shades. Refer to Fig. 1. Variance in shade was made proportional for both within and between color categories using MATLAB (v. R2024b). Specifically, the middle color of each set was the prime color ([0, 175.49, 0] on the RGB scale for green and [0, 0, 175.49] for blue). The surrounding four colors were lighter and darker shades of the prime color (±39.75 and ± 79.48 from the prime color in “G” and “B” respectively).

The priming phase of the experiment presented the two prime objects side-by-side. Under one of the squares was a small circle indicating that it was the prime object. The target phase presented a singular color square in the center of the screen. The prime objects were counterbalanced to appear on both sides of the screen (i.e., 4 possible prime screens) and could be paired with any of the 8 targets, resulting in 32 unique prime-target trials.

Of the 32 unique prime-target trials, half were colors in the same

---

[3] The predictions for encoding condition differ slightly from our OSF pre-registration filed in 2024 (https://osf.io/jwxtr). This adjustment is the result of insightful comments from two reviewers who suggested that without a true baseline condition (passive viewing), it is safest to frame the paper around how pointing and labeling interact across the neural time course of visual processing.

[4] Using a moderate effect size (d = 0.25), power of 0.8, and $p < 0.05$, G*Power suggested a sample size of 19. However, because we had six presentation orders for the three conditions, the sample was set at 24 participants. This is slightly over-powered, but it is necessary to control for order effects, and this plan was stated in our OSF pre-registration.

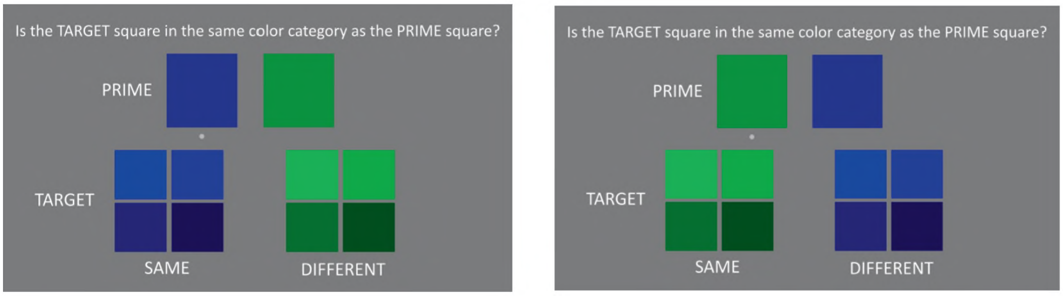

Fig. 1. These images were shown to participants during a pre-experiment introductory presentation (in the actual experiment, the primes and targets never appeared on the same screen). It demonstrated that the prime and target objects would never be the exact same color but would instead belong to congruent or incongruent color categories. Note the small white circle that indicates the prime square (left: blue prime; right: green prime).

category as the prime, and the other half were from different categories. In this way, the Prime-Target relationship was our first independent variable, with levels referred to as Congruent and Incongruent trials. The order of the 32 trial blocks was completely randomized with every presentation. This set was repeated 4 times for a total of 128 prime-target pairs. Congruent and Incongruent trials each accounted for half of the total prime-target pairs (64 each).

8.1. Procedure

Participants entered the lab and were given a consent form. After signing, they were asked to remove any ear jewelry and electronic devices (phones, smart watches, etc.). Next, participants were ushered into an electrically shielded and sound-proof room where the experiment took place. Following head measurements, participants were taken through a brief slideshow by the experimenter explaining the task.

The presentation introduced participants to the stimuli (see Fig. 1), and the experimenter received verbal confirmation that all stimuli appeared visually distinct to the participant. Additionally, participants were shown all four possible configurations of the target screen (bottom part of Fig. 1). Afterwards, they completed four mock trials of the task with the experimenter's aid/supervision. This way, participants were able to practice identifying targets that were congruent and incongruent for both prime colors and ask clarifying questions about the task. It was during this portion of the presentation that the experimental task was introduced: participants were instructed to click a button with their left hand if they thought the prime and target objects were part of the same color category and a different button if they thought the color categories were different. At the end of the presentation, the experimenter emphasized that the participant should try to complete the task with speed and accuracy. The experimenter further cautioned the participant about movement artifacts and specifically explained how the participant should time their eye blinking to reduce noise in the EEG (eye artifacts were the main source of data rejection). The net was then applied and impedances were measured. A team of researchers used small, plastic syringes to better situate electrodes onto the scalp, and when necessary, they used a pipette to apply more electrolyte.

Following a brief explanation of the three "encoding conditions" in the study (more below), the experiment began. The primes, blue and green squares, were presented on the screen for 1000 ms and were followed by a 2000 ms pause with a white fixation cross in the center of the screen. Next, a target that belonged to either the same or a different color category as the prime was presented at the center of the screen (where the fixation cross was located). It remained on the screen until participants made their selection. For an overview of the task, see Fig. 2. Following a prime-target pair trial, there was a 1500–2000 ms variable inter-stimulus interval.

All stimuli were presented through E-Prime 3.0 (Psychological Software Tools), and accuracy scores and reaction times were measured by a Chronos Box (Psychological Software Tools). The stimuli were presented on a Dell C2423H monitor with AMD Radeon™ Graphics on a gray background ([128, 128, 128] on the RGB scale). The display measured 27″ diagonally and had a 1920 x 1080 resolution with a 60 Hz refresh rate.

8.2. Prime Encoding Modality Conditions

Subjects performed the priming task across three Encoding Modality conditions: For the label condition, participants were told to verbally label the indicated color of the prime ("green" or "blue"). For the point condition, participants were told to point with their right hand to the indicated prime color without saying anything. And for the point & label condition, participants were instructed to label and simultaneously point to the indicated prime color with their right hand. The order of conditions (six possible orders) was counterbalanced equally across our 24 subjects. Within each condition, there was a pause every 32 trials when participants could take a self-paced break. Before switching to a new encoding condition after 128 trials, a team of researchers monitoring EEG data checked impedances and determined whether or not the participant required additional electrolyte. In total, each participant completed 384 trials (3 encoding conditions x 128 trials).

8.3. EEG Application

Participants were fitted with a 128-electrode Geodesic Sensor Net. Impedances for the large majority of electrodes were adjusted to register below 50kΩ. In Net Station Tools (v. 5.5 Magstim Electrical Geodesics, Inc), EEG segments ranging from -100 to 800 ms relative to the target object onset were extracted. Trials consisting of movement artifacts and activity exceeding 50kΩ were tagged by an algorithm using Net Station Tools and replaced (off-line) with an average signal taken from surrounding electrodes. Additionally, all electrodes were re-referenced to the average reference during data reduction. Baseline correction was applied relative to the 100 ms before target object onset, and a 0.10–30 Hz bandpass filter was applied with a 60 Hz notch.

As specified in our pre-registration and in accordance with Boutonnet & Lupyan (2015), three ERP components were identified: The P1 and P2 components were measured at left (electrodes 68, 69, 70, 71, 73, and 74) and right (76, 82, 83, 88, 89, and 94) occipital-parietal regions with average amplitudes taken from the 70–125 ms and 190–230 ms time range, respectively. The N400 component was measured at central sites (electrodes 6, 7, 13, 30, 31, 37, 54, 55, 79, 80, 87, 105, 106, and 112) with the average amplitude taken from the 300–500 ms range.

8.4. Design and Analyses

The basic analysis plan involved exporting the pre-processed data from Net Station Tools into MATLAB (v. R2024B), processing it (more below), and then submitting the aggregate file to a linear mixed-effects model conducted in SPSS (v. 29.0.1.0 (171)). Follow-up *a priori*

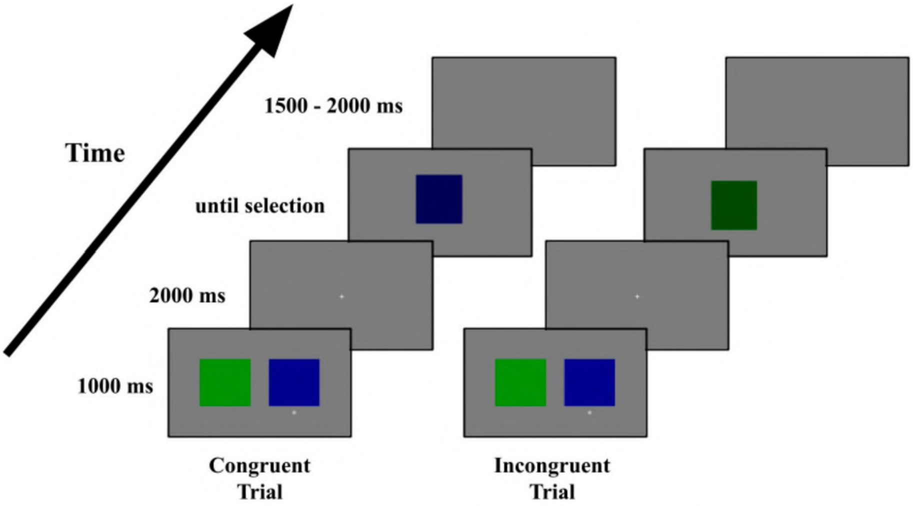

Fig. 2. Visual overview of one trial (Congruent & Incongruent).

Bonferroni-corrected t-tests in SPSS explored significant interactions. This plan complies with our Open Science Framework pre-registration (https://osf.io/jm4qk/), except where indicated otherwise.

8.5. Behavioral data

Reaction Time (RT) averages and accuracy error rates were calculated in MATLAB for each participant. As a first pass, all RTs above 2000 ms were discarded, and for the remainder, we removed trials that were more than two standard deviations above/below the mean within each condition. Only correct responses were included in the RT measure. These files were then exported to SPSS where we ran a linear mixed-effects model with the following parameters: congruence, encoding modality and their interaction were entered as fixed-effects, and subjects were entered as a random effect.[5]

8.6. ERP data

For each participant, all non-rejected ERP trials were exported from Net Station Tools to MATLAB to create average amplitudes for the three components. Next, these files were exported to SPSS and submitted to a linear mixed-effects model. The parameters were as follows: congruence, modality, hemisphere (for the P1 and P2 only), and their interactions were entered as fixed-effects, and subjects were entered as a random-effect. Hemisphere did not interact with congruence and encoding modality in any component, so in the interest of space, we do not report any hemisphere effects. For all ERP components, follow-up *a priori* t-tests explored significant interactions.

9. Results

9.1. Behavioral Results

9.1.1. Error Rates

There was no main effect of congruence, with congruent targets (M = 3%, SD = 17%) and incongruent targets (M = 2%, SD = 16%) producing similar error rates (F(1, 9098) = 2.370, p = 0.124). There was also no main effect of encoding modality, with label (M = 2%, SD = 15%), point (M = 3%, SD = 17%), and point & label (M = 3%, SD = 16%) producing similar error rates (F(2, 9098) = 0.533, p = 0.587). Furthermore, there was no significant interaction between congruence and encoding modality (F(2, 9098) = 1.927, p = 0.146).

9.1.2. Reaction Times

There was a significant main effect of congruence for RTs (F(1, 9108) = 159.058, p < 0.001, η^2p =.607[6]; see Fig. 3), with faster RTs for congruent targets (M = 618 ms, SD = 273 ms) compared to incongruent targets (M = 677 ms, SD = 288 ms). There was no significant main effect of encoding modality (F(2, 9108) = 1.941, p = 0.144), with RTs for label (M = 642 ms, SD = 267 ms), point (M = 648 ms, SD = 295 ms), and point & label (M = 652 ms, SD = 284 ms) having no significant differences. Lastly, there was no significant interaction between congruence and encoding modality (F(2, 9108) = 1.832, p = 0.160).

9.2. Electrophysiological Results

9.2.1. P1 Component

There was a significant main effect of congruence for the P1 (F(1, 15967) = 11.814, p < 0.001, η^2_p = 0.212), with incongruent targets (M = 1.848 μV, SE = .469 μV) eliciting a more positive P1 than congruent targets (M = 1.593 μV, SE = .469 μV). See Fig. 4A. Regarding encoding modality, there was also a significant main effect for the P1 (F(2, 15967) = 6.005, p = 0.002, η^2p = 0.068). Pointing to the prime object elicited a more positive P1 (M = 1.883 μV, SE = .470 μV) than labeling (M = 1.568 μV, SE = .470 μV, p = .002) but not point & label (M = 1.714 μV, SE = .470 μV, p = 0.192). There was also no significant difference between labeling and point & label (p = 0.334). See Fig. 4B and Fig. 6 for the scalp topographies. Lastly, congruence and encoding modality did not significantly interact (F(2, 15967) = 0.220, p = 0.803).

9.2.2. P2 Component

There was a significant main effect of congruence for the P2 (F(1, 15989) = 63.260, p < 0.001, η^2_p = 0.493), with incongruent targets eliciting a more positive P2 (M = 2.214 μV, SE = .434 μV) than congruent targets (M = 1.495 μV, SE = .434 μV). See Fig. 4A. For encoding modality, there was also a significant main effect (F(2, 15989) = 5.41, p = 0.006, ηp^2 = 0.045). Pointing elicited a more positive P2 (M = 2.056 μV, SE = .437 μV) than labeling (M = 1.731 μV, SE = .437 μV, p = 0.010) and point & label (M = 1.776 μV, SE =.437 μV, p = 0.034). There

---

[5] In our OSF pre-registration, we included additional variables, such as prime color and location, as potential random effects, but model fitting indicated that subject was the only variable that should be retained. The same held for the ERP data.

[6] To more accurately represent subject-level effect sizes, partial eta squared values were calculated based on RT averages and ERP components within each subject instead of across items.

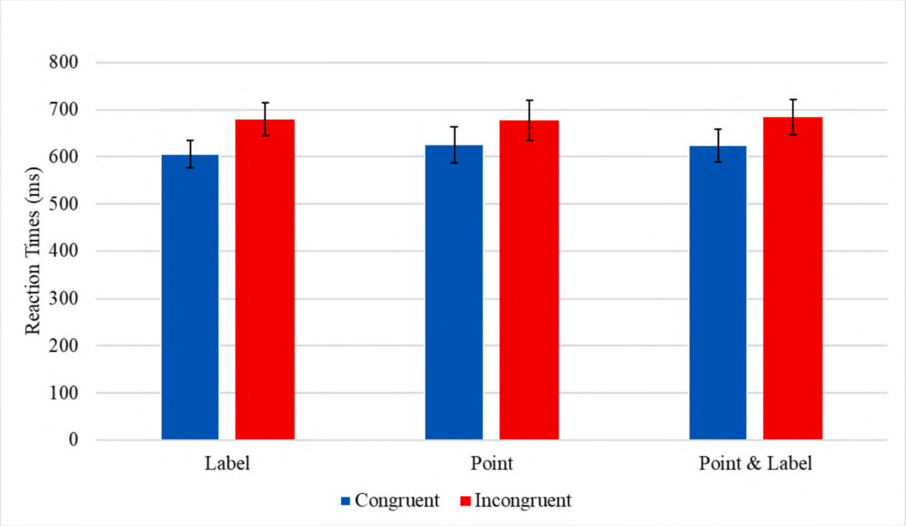

Fig. 3. These RTs were calculated from subject-level averages and not trial-level averages to reduce noise. Congruent targets (blue) were significantly faster than incongruent targets (red) across all encoding modalities ($p < 0.001$). Error bars represent the standard error. There was no significant effect of encoding modality.

was no significant difference between labeling and point & label (p = 1.000). See Fig. 4B and Fig. 6 for the scalp topographies. Lastly, congruence and encoding modality did not significantly interact ($F(2, 15989) = 0.473$, $p = 0.623$).

9.2.3. N400 Component

There was a significant main effect of congruence for the N400 ($F(1, 7972) = 5.715$, $p = 0.017$, $\eta^2_p = 0.149$), with incongruent targets eliciting a more negative N400 (M = 1.539 μV, SE = .406 μV) than congruent targets (M = 1.756 μV, SE = .406 μV). See Fig. 5A.[7] There was also a significant main effect of encoding modality ($F(2, 7972) = 14.717$, $p < 0.001$, η^2p = 0.139). Pointing to the prime object elicited a more positive N400 (M = 1.990 μV, SE = .408 μV) compared to labeling (M = 1.432 μV, SE =.408 μV, $p < 0.001$) and point & label (M = 1.519 μV, SE =.408 μV, $p < 0.001$). See Fig. 5B and Fig. 6 for the scalp topographies. There was no significant difference between labeling and point & label (p = 1.000), and there was no significant interaction between congruence and encoding modality ($F(2, 7972 = 0.110$, $p = 0.896$).

10. Discussion

We found mixed support for our predictions. Regarding Prediction 1, there was a main effect of congruence on the P2, with incongruent targets evoking a more positive average amplitude compared to congruent targets. For the final time window, we confirmed that incongruent targets produced a more negative-going N400 than congruent targets. However, there was also an unpredicted main effect of congruence on the P1 which followed the same pattern as the P2.

For Prediction 2 about encoding effects, we found that pointing and labeling indeed did interact in a stage-dependent fashion in our ERPs—but not exactly in the way we expected. For the N400, we confirmed that point & label behaved more like labeling than pointing. However, we found an unexpected pattern for the P2: point & label was similar to labeling—and both were different than pointing. Finally, the P1 showed the expected directional pattern, but it was not significant: pointing and labeling differed, whereas point & label was intermediate and did not differ from either.

10.1. Congruence Effects

The congruence effect for the P2 aligns with prior research showing that the component is sensitive to color pop-outs (i.e., a green color amongst blue colors; Luck & Hillyard, 1994) and could reflect a perceptual-matching mechanism in parieto-occipital regions (Freunberger et al., 2007). Boutonnet & Lupyan (2015) also found a more positive P2 for incongruent targets within the same ROI, so it follows that this early perceptual-matching process would be sensitive to incongruent targets in our study (e.g., a blue target following a green prime). We also replicated previous research showing an N400 effect between congruent and incongruent visual stimuli (Kelly et al., 2004; Kutas & Federmeier, 2011; McPherson & Holcomb, 1999; Wu & Coulson, 2005).

In addition, we found a more positive P1 for incongruent targets compared to congruent targets. Prior research has found that the P1 reflects very early selective attention (Rutman et al., 2010), and while some studies have not found an effect of congruence on the P1 (Boutonnet & Lupyan, 2015; Freunberger et al., 2007), other research has found a more positive P1 for incongruent targets (de Leeuw, et al., 2021; Kiefer et al., 2011). For example, Kiefer et al. (2011) found that the P1 to pictorial targets was larger when primed by incongruent vs. congruent pictures, suggesting that the P1 might reflect a visual feature extraction mechanism sensitive to low-level visual features. Thus, if attention is primed for a specific feature of an object, any violations might be rapidly processed before higher-level recognition can take place. These perceptual mechanisms may also explain the difference between our results and Boutonnet & Lupyan (2015). Note that Kiefer et al. (2011) found that prime-target congruence modulated the P1 only for pictorial and not word primes. Boutonnet and Lupyan used visual targets, but their primes were *auditory* (i.e., dog barking) whereas our primes and targets were *visual*, so our participants directly activated perceptual mechanisms.

10.2. Encoding Effects

Turning to our most novel contribution, the encoding results show that pointing and labeling have different downstream consequences for visual processing, suggesting that what we do with our words and hands

---

[7] Based on visual inspection of Fig. 5A, there was an additional prominent negativity for incongruent targets at 200–250 ms post-stimulus, which was significantly different from congruent targets ($F(1, 7972.081) = 79.68$, $p < .001$; $\eta^2_p = .640$). The scalp topography and time course of this negativity suggest a central N2 component that is typically associated with early conflict processing in sequential matching tasks (for more, see Folstein & Van Petten, 2008).

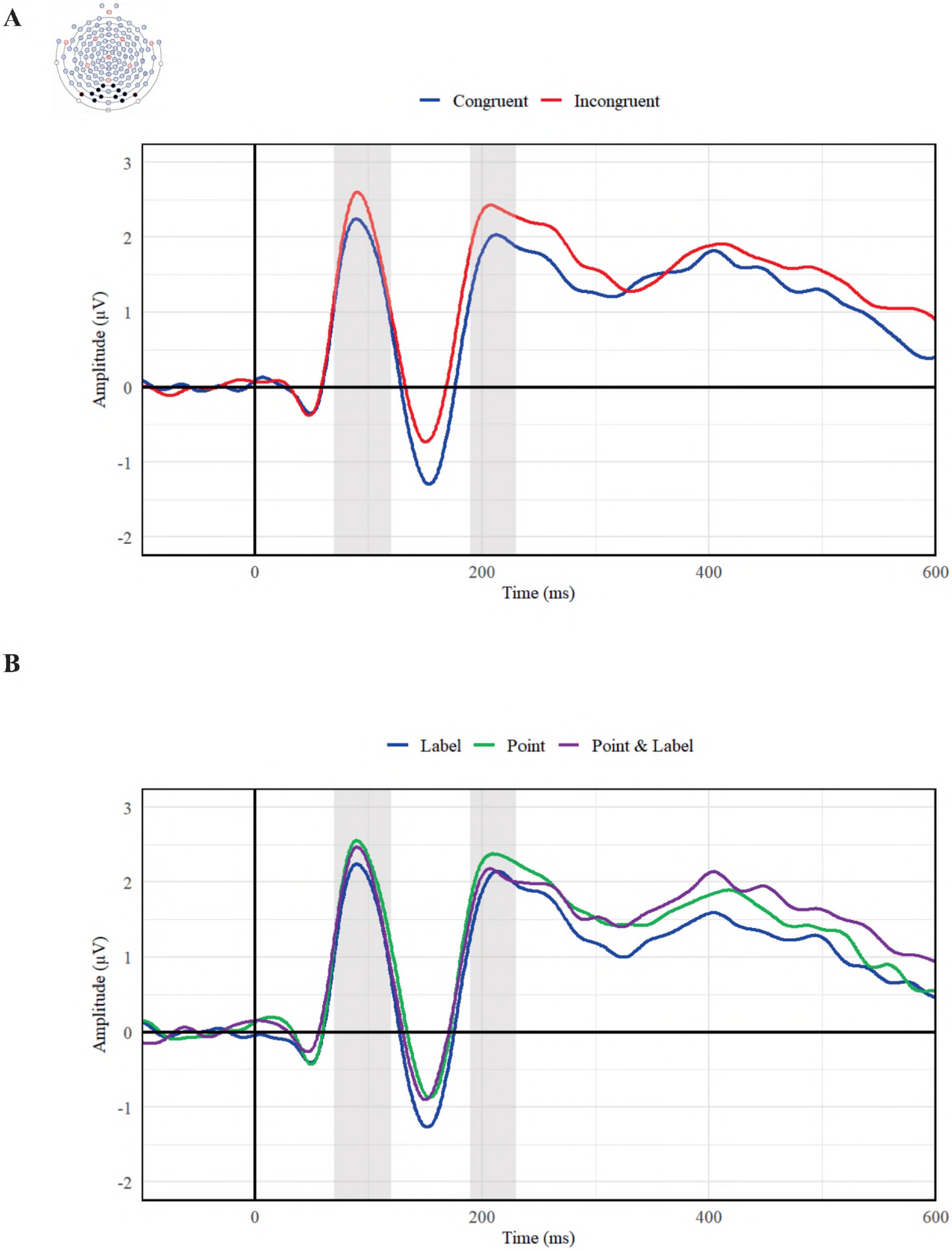

Fig. 4. ERPs for congruence and modality across the parieto-occipital ROI. The shaded regions identify the time windows taken for the P1 average amplitude (70 ms - 125 ms) and for the P2 average amplitude (190 ms - 230 ms). ***A,*** P1 and P2 ERPs across both levels of congruence: In red are ERPs elicited by incongruent targets and in blue are ERPs elicited by congruent targets. ***B,*** P1 and P2 ERPs across three levels of modality: In green are ERPs elicited by pointing, in blue are ERPs elicited by labeling, and in purple are ERPs elicited by pointing & labeling.

affects what we subsequently see. Specifically, we discovered that pointing and labeling interact in dynamic ways across different stages of visual processing. At the early attentional/perceptual stage (P1), the point & label condition was intermediate between the two unimodal conditions and not reliably different from either, consistent with a graded, non-additive interaction. In contrast, at later stages, the point & label condition closely paralleled labeling for perceptual categorization (P2) and conceptual processing (N400), suggesting that labeling eventually dominates downstream processing when both modalities are produced together.

The clearest conclusion from our results is that pointing and labeling do not interact uniformly and additively when processing visual information. This deviates from classic dual-coding accounts that emphasize additive benefits of multimodal representations (Paivio, 1986), but it fits well with neurobiological theories that multimodal integration varies over time and is often non-additive (Stein et al., 2020). If pointing and labeling had interacted in a uniform and additive fashion, we would have expected the combination of modalities to produce larger effects

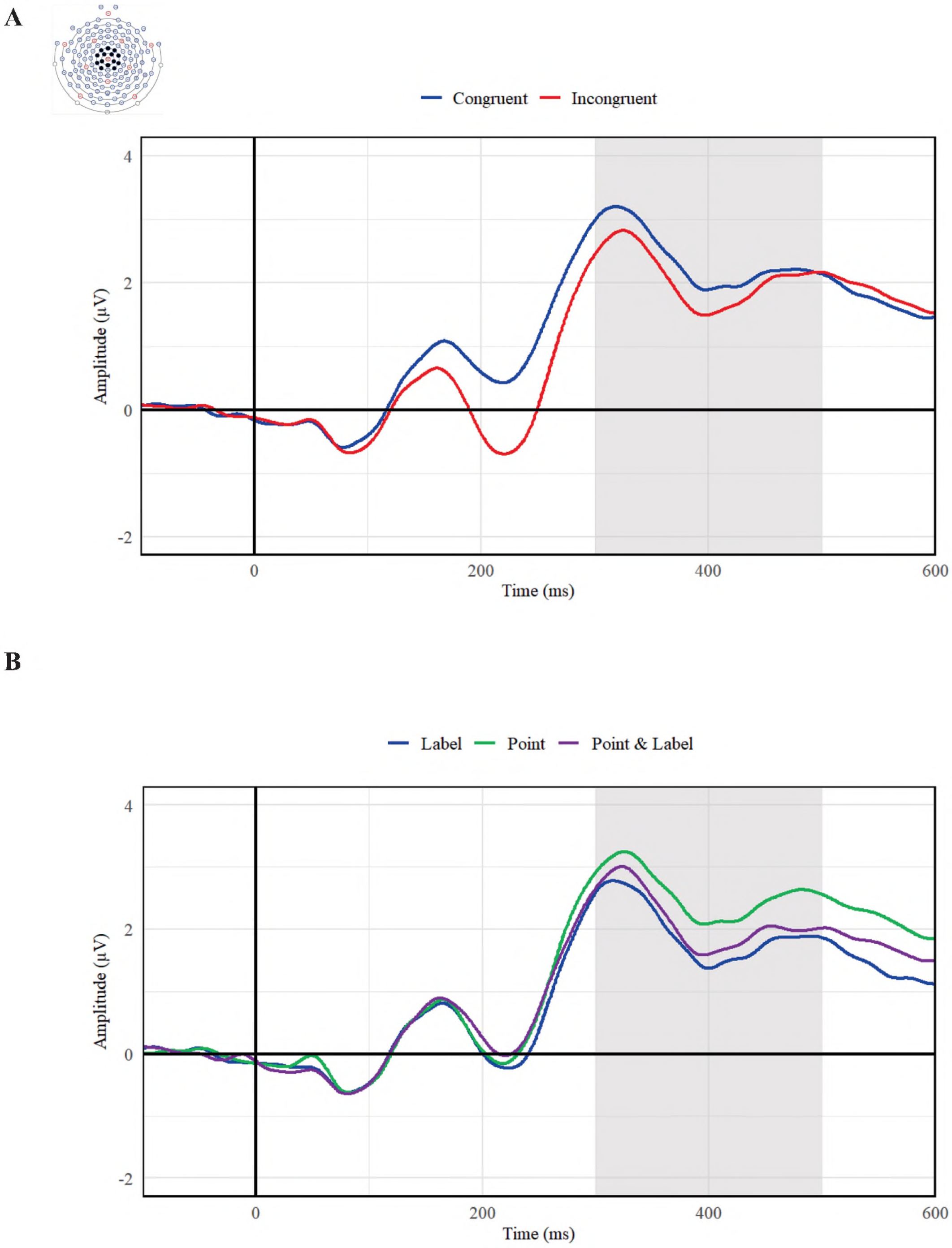

Fig. 5. ERPs for congruence and encoding modality across the central ROI. The shaded region identifies the time window taken for the N400 average amplitude (300 ms - 500 ms). ***A,*** N400 ERPs across both levels of congruence: In red are ERPs elicited by incongruent targets and in blue are ERPs elicited by congruent targets. ***B,*** N400 ERPs across three levels of modality: In green are ERPs elicited by pointing, in blue are ERPs elicited by labeling, and in purple are ERPs elicited by pointing & labeling.

than the unimodal conditions for all three ERP components. Instead, the two modalities interacted in a more complex way: early at the P1, the point & label condition was intermediate between the two unimodal conditions (and was numerically closer to pointing), whereas later at the P2 and the N400, it more closely resembled labeling. This stage-dependent pattern is consistent with a dynamic integration account in which the relative contribution of pointing and labeling may shift over time during visual perception.

The present design did not include a true baseline (passive viewing), so it is not possible to know for sure *how* pointing and labeling interacted in visual processing. However, based on known mechanisms for each of our ERP components, and based on research identifying different functions of pointing and labeling, we can speculate on how each modality may have produced different neural responses while processing the visual targets.

Regarding the attentional-perceptual effects of encoding modality, recall that pointing evoked a more positive P1 and P2 than labeling. These results align with the reported attentional benefits of pointing (Ariga & Watanabe, 2009; Park et al., 2023). Park et al. (2023) found that when participants utilized self-oriented pointing, they were quicker

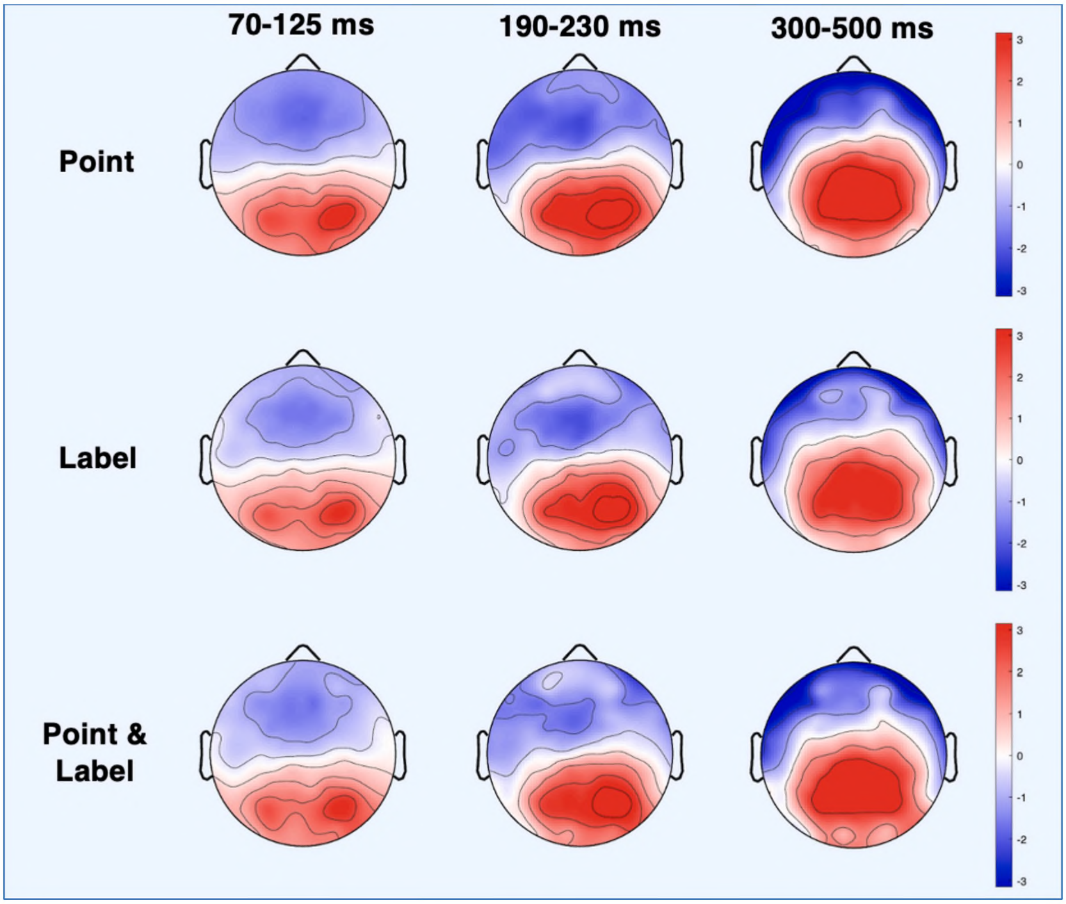

Fig. 6. Target-locked scalp topographies depict mean voltage within each time window. Red indicates positive voltage and blue indicates negative voltage (μV).

to fixate on an illustration and spent a longer time observing it, suggesting that pointing can more readily and deeply focus visual attention. With regard to our P1 finding, pointing might have helped participants more deeply focus their attention on the prime, resulting in increased visual attention to the salient feature (color) of our target. Furthermore, other research suggests that pointing can also facilitate object individuation (Carlson et al., 2007). In the present study, pointing may have helped participants individuate the prime to accentuate the contrast between the non-prime object also on the screen, and/or it may have better individuated the prime object's color, thereby improving early visual feature extraction. The process of object individuation also helps to explain our P2 unimodal differences, suggesting that pointing can better modulate early perceptual-matching processes (Freunberger et al., 2007). If pointing allowed participants to better orient themselves to the prime's most salient physical feature, color, they might have been more perceptually sensitive to the color of the subsequent target stimuli. Both of these accounts, however, cannot explain why we failed to support our prediction that point & label would mirror the point condition for these two components. We return to this unexpected finding later.

Regarding our final component, the fact that labeling elicited a more negative-going N400 than pointing might suggest that labeling introduces a conceptual load that pointing does not. That is, when participants labeled a prime "green" or "blue," they activated a lexical retrieval process that introduced semantic information to working memory (Delogu et al., 2019; Overkott & Souza, 2022; Souza & Sko´ra, 2017), and this semantic trace may have revealed itself during the N400 time window when lexical and conceptual integration processes typically unfold.

In a different vein, but consistent with labels adding a conceptual load, previous research suggests that continuous information about color (i.e., fine-grained perceptual representations) can be biased toward prototypical color categories (Regier &Xu, 2017), which can be heightened by the activation of color labels (Bae et al., 2015). This further supports a more conceptual function for labeling in our study: saying the word "green" might have activated a categorical "magnet" that attracted the memory for that shade of green towards the category center (i.e., a prototype green) (for an alternative account, see Overkott & Souza, 2022). In contrast, pointing to a green prime may have activated less categorical information because an index excels at connecting a signal to a *particular* object in the external world (Clark, 1996; Cooperrider, 2023). The end result is that the N400 may have been sensitive to these different activations of linguistic categories when processing the target.

In the late (300–500 ms) window, the scalp topographies (Fig. 6) appear relatively positive-going at central-parietal sites compared to a canonical negative-going N400, raising the possibility that our 'N400-window' effect reflects a late positivity (e.g., P3b/late positive complex) that partly masks an underlying N400 negativity, rather than a pure N400 component. In other studies, late central-parietal positivities such as the P3b have been linked to post-perceptual stimulus evaluation and context updating/decision-related processing, and related late positive complexes have been associated with controlled evaluative and memory-related operations in some paradigms (Key et al., 2005; Polich, 2007). Thus, the main point here is not the precise component label, but that activity in this time window likely reflects later, post-perceptual processing that is qualitatively distinct from the earlier sensory-perceptual stages indexed by P1/P2. In this context, the fact that labeling alone and pointing & labeling together both diverge from pointing alone in this late window suggests that the presence of a label shifts processing toward more meaning-based evaluative operations during target categorization.

10.3. Pointing and Labeling Across Processing Stages

One way to interpret the overall pattern of results—again, with caveats about having no true baseline—is that pointing may exert its strongest influence at upstream attentional-perceptual stages, whereas labeling may exert its strongest influence at later stages associated with perceptual categorization and conceptual processing. We offer two interrelated considerations in support of this possibility. First, because a key function of pointing is to highlight information outside one's head (Cooperrider, 2023), its effects may be expressed early—by biasing visual attention to exogenous stimuli and facilitating feature selection/extraction. This account is consistent with our P1 pattern: pointing and labeling alone significantly differed, and while the point & label was numerically closer to pointing, it was intermediate overall and did not differ reliably from either unimodal condition. This intermediate positioning of point & label may reflect some sort of early-stage competition between the two modalities.

Second, and inversely, a primary function of words is to activate information *inside* the head. In so doing, the semantic process of labeling might divert resources away from, or add a conceptual load to, earlier perceptual mechanisms. When saying "green" or "blue", this adds high-level lexical retrieval to the low-level neural mechanisms for extracting color features from objects in the environment. This account may help to explain why point & label more closely parallelled the unimodal label condition during later stages of processing: the conceptual and categorical power of labels was so strong that it dominated semantic processing when the modalities were combined. It is noteworthy that this semantic influence of labelling appeared as early as the P2 component, which went against our prediction. This raises the interesting possibility that labeling activates a particularly powerful conceptual process that may override the perceptual functions of pointing within a mere 200 ms of visual processing. Without a baseline condition, this is speculative, but it fits well with visual processing models demonstrating that high-level conceptual information influences perception of complex scenes within 170 ms (Greene & Hansen, 2020). Moreover, it is consistent with previous claims that words exert a top-down influence at very early stages of perception (Boutonnet & Lupyan, 2015; Lupyan et al., 2007; Lupyan, 2008; Lupyan & Swingley, 2012; Lupyan & Ward, 2013).

Metaphorically speaking, the functions of pointing and labeling can be seen as two waves along the shore that may crash at different times but ultimately reach the same destination. Pointing may serve exogenous perceptual functions to better process what is in the outside world, so the wave crashes early; labeling may serve endogenous conceptual functions to process what is inside one's head, so the wave crashes later.

11. Limitations

There are several limitations of the current study that could be addressed in future research. First, as already mentioned, the largest limitation is that our design did not include a true baseline condition of no pointing and no labeling (neither did Boutonnet & Lupyan, 2015, for that matter). Although this limitation does not detract from our main conclusion—that pointing and labeling interact in non-additive and non-uniform ways during different stages of visual processing—it does temper what we can conclude about the *nature* of those interactions. For example, without a baseline condition, we cannot determine whether the unimodal differences reflect enhancement by pointing, suppression by labeling, or the possibility that one modality is inert while the other drives the effect. Better pinning down these precise mechanisms is an interesting direction for future studies.

Second, for our perceptual ERPs, it is difficult to disentangle which function of pointing is contributing to the enhanced P1 and P2: spatial-orienting or feature-extraction? After all, we know that pointing can strongly modulate spatial attention (Cappuccio et al., 2013; Chum et al., 2007; Hagler et al., 2007; Ouwehand et al., 2015). Because our targets always appeared in the center of the screen, and our primes always appeared on either the left or right side, it is possible that the heightened P1 and P2 for pointing might also reflect perceptual sensitivity to the target's changed location. However, if this were exclusively the case, the combined point & label condition should be similarly affected, but it was not. Rather, at least starting at 200 ms, the point & label condition behaved more like the label than the pointing condition. Moreover, spatial location was not relevant to the task; only color was. Given that color can be extracted independently of spatial locations (Zhang & Luck, 2009), it is likely that the spatial function of pointing was not the only driver of the early perceptual effects. To tease apart spatial orientation and feature extraction, future studies could present primes and targets in the same visual field—if the enhanced P1 and P2 effect remain, pointing is doing more than simply directing spatial attention.

Third, in post-experiment interviews, almost all participants reported saying the color word in their heads while they pointed. Given that this subvocalization likely diluted the perceptual function of pointing, it is even more impressive that we still found that it elicits a more positive P1 and P2 than verbal labeling. Even if subvocalization did affect perception (and later conceptual processes), its influence was likely rather small. Prior research suggests that reading a word out loud can improve memory in a semantic recall task compared to reading a word silently (MacLeod et al., 2010). A similar effect might occur in visual matching tasks as well, with one study showing that participants in an overt speech condition had better visual memory compared to those in a quiet condition (Guo and Dobkins, 2023). The authors hypothesize that labeling out loud might increase the meaningfulness of an image, thereby activating long-term memories and improving working memory. To eliminate any effects of subvocalization, future studies could add a verbal interference task, as in Souza & Sko´ra, 2017. With this control, we would predict even larger perceptual and conceptual differences between pointing and labeling.

And fourth, in the prime phase, each prime that participants encoded was explicitly marked by a white circle underneath it. This design was implemented to reduce ambiguity for referential purposes; however, it may have made pointing and labeling redundant (after all, the object was already indicated with the circle). Therefore, the current encoding conditions may not fully represent the most natural case where people socially use pointing or labeling to resolve ambiguity (e.g., "which one is being referred to?"). Future research should explore this relationship in more naturalistic social contexts—where disambiguating information is the primary goal—to determine if the integration of pointing and gesture is enhanced, attenuated, or altered in some other way.

12. Conclusion

It is well-established that gesture and speech form an integrated system in language production (McNeill, 1992). The present study extends this integrated relationship to the subsequent processing of visual information. The results suggest that self-oriented pointing and labeling modulate the downstream neural processing of information in non-additive and non-uniform ways across different stages of vision. This stage-dependent pattern is consistent with the possibility that pointing may be relatively more dominant at very early perceptual stages, whereas labeling may contribute more at later conceptual stages. Future research can help specify the mechanisms through which this multimodal interaction unfolds over time in the visual system.

Declaration of generative AI and AI-assisted technologies in the MS preparation process

During the preparation of this work, the authors used ChatGPT (OpenAI) to:

1) proofread and suggest clearer wording for complex sentences.
2) generate additional keywords and search strategies to complement traditional literature searches.
3) help interpret technical language in published papers (as an aid to

the authors' own reading).

4) assist with drafting and debugging analysis codes and with formatting tables/figures for data presentation (all outputs were checked by the authors)

After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

CRediT authorship contribution statement

Spencer D. Kelly: Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization. **Dylan Vlasak:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Data curation. **Binghui Tang:** Writing – review & editing, Validation, Supervision, Software, Methodology, Investigation, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Alderson-Day, B., & Fernyhough, C. (2015). Inner speech: Development, cognitive functions, phenomenology, and neurobiology. *Psychological Bulletin, 141*(5), 931–965. https://doi.org/10.1037/bul0000021

Ariga, A., & Watanabe, K. (2009). What is special about the index finger?: The index finger advantage in manipulating reflexive attentional shift. *Japanese Psychological Research, 51*(4), 258–265. https://doi.org/10.1111/j.1468-5884.2009.00408.x

Astafiev, S. V., Shulman, G. L., Stanley, C. M., Snyder, A. Z., Van Essen, D. C., & Corbetta, M. (2003). Functional organization of human intraparietal and frontal cortex for attending, looking, and pointing. *The Journal of Neuroscience, 23*(11), 4689–4699. https://doi.org/10.1523/jneurosci.23-11-04689.2003

Bae, G., 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. https://doi.org/10.1037/xge0000076

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly of Behavior and Development, 21*(3), 205–226.

Boutonnet, B., & Lupyan, G. (2015). Words jump-start vision: A label advantage in object recognition. *Journal of Neuroscience, 35*(25), 9329–9335. https://doi.org/10.1523/jneurosci.5111-14.2015

Cappuccio, M., Chu, M., & Kita, S. (2013). Pointing as an instrumental gesture: Gaze representation through indication. *Humana. Mente: Journal of Philosophical Studies, 24*, 125–150.

Carlson, R. A., Avraamides, M. N., Cary, M., & Strasberg, S. (2007). What do the hands externalize in simple arithmetic? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(4), 747–756. https://doi.org/10.1037/0278-7393.33.4.747

Chomsky, N. (2011). Language and other cognitive systems. What is special about language? *Language Learning and Development, 7*(4), 263–278. https://doi.org/10.1080/15475441.2011.584041

Church, R. B., Alibali, M. W., & Kelly, S. D. (2017). *Why Gesture?: How the hands function in speaking, thinking and communicating.* John Benjamins Publishing Company.

Chum, M., Bekkering, H., Dodd, M. D., & Pratt, J. (2007). Motor and visual codes interact to facilitate visuospatial memory performance. *Psychonomic Bulletin & Review, 14*(6), 1189–1193. https://doi.org/10.3758/BF03193111

Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences, 10*(8), 370–374. https://doi.org/10.1016/j.tics.2006.06.012

Clark, H. H. (1996). *Using language.* Cambridge University Press.

Cooperrider, K. (2023). Fifteen ways of looking at a pointing gesture. *Public Journal of Semiotics, 10*(2), 40–84. https://doi.org/10.37693/pjos.2023.10.25120

de Leeuw, J. R., Andrews, J., Barney, L., Bigler, M., Bruna, P. J., Chen, Y., Cherry, R., Dowie, D. R., Forbes, E., Haffey, B., & Hu, X. (2021). Words may jump-start meaning more than vision: A non-replication of early ERP effects in Boutonnet and Lupyan (2015). *Collabra: Psychology, 7*(1), 29763. https://doi.org/10.1525/collabra.29763

Delogu, F., Brouwer, H., & Crocker, M. W. (2019). Event-related potentials index lexical retrieval (N400) and integration (P600) during language comprehension. *Brain and Cognition, 135*, Article 103569. https://doi.org/10.1016/j.bandc.2019.05.007

Dove, G. (2019). More than a scaffold: Language is a neuroenhancement. *Cognitive Neuropsychology, 37*(5–6), 288–311. https://doi.org/10.1080/02643294.2019.1637338

Fedorenko, E., Piantadosi, S. T., & Gibson, E. A. F. (2024). Language is primarily a tool for communication rather than thought. *Nature, 630*(8017), 575–586. https://doi.org/10.1038/s41586-024-07522-w

Ferrara, L., & Hodge, G. (2018). Language as description, indication, and depiction. *Frontiers in Psychology, 9*, 716.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Freunberger, R., Klimesch, W., Doppelmayr, M., & Höller, Y. (2007). Visual P2 component is related to theta phase-locking. *Neuroscience Letters, 426*(3), 181–186. https://doi.org/10.1016/j.neulet.2007.08.062

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*(6), 516–522. https://doi.org/10.1111/1467-9280.00395

Greene, M. R., & Hansen, B. C. (2020). Disentangling the independent contributions of visual and conceptual features to the spatiotemporal dynamics of scene categorization. *Journal of Neuroscience, 40*(27), 5283–5299. https://doi.org/10.1523/JNEUROSCI.2088-19.2020

Guo, X., & Dobkins, K. (2023). Private speech improves cognitive performance in young adults. *Consciousness and Cognition, 116*, Article 103585. https://doi.org/10.1016/j.concog.2023.103585

Hagler, D. J., Riecke, L., & Sereno, M. I. (2007). Parietal and superior frontal visuospatial maps activated by pointing and saccades. *NeuroImage, 35*(4), 1562–1577. https://doi.org/10.1016/j.neuroimage.2007.01.033

Kelly, S. D., Kravitz, C., & Hopkins, M. (2004). Neural correlates of bimodal speech and gesture comprehension. *Brain and Language, 89*(1), 253–260. https://doi.org/10.1016/s0093-934x(03)00335-3

Kelly, S. D. (2024). Gesture. In M. C. Frank, & A. Majid (Eds.), *Open Encyclopedia of Cognitive Science*. MIT Press. https://doi.org/10.21428/e2759450.441037e5.

Key, A. P. F., Dove, G. O., & Maguire, M. J. (2005). Linking brainwaves to the brain: An ERP primer. *Developmental Neuropsychology, 27*(2), 183–215. https://doi.org/10.1207/s15326942dn2702_1

Kiefer, M., Sim, E. J., Helbig, H., & Graf, M. (2011). Tracking the time course of action priming on object recognition: Evidence for fast and slow influences of action on perception. *Journal of Cognitive Neuroscience, 23*(8), 1864–1874. https://doi.org/10.1162/jocn.2010.21543

Kishi, C., Nakano, Y., Bhuiyan, M. S., Kawanaka, H., & Oguri, K. (2014). Verification of the effect on "Finger Pointing and calling" method from observation of brain activity related driver's attention. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*. https://doi.org/10.1109/itsc.2014.6957974

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Luck, S. J., & Hillyard, S. A. (1994). Electrophysiological correlates of feature analysis during visual search. *Psychophysiology, 31*, 291–308. https://doi.org/10.1111/j.1469-8986.1994.tb02218.x

Lupyan, G. (2008). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition, 108*(2), 566–577. https://doi.org/10.1016/j.cognition.2008.03.009

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking. *Psychological Science, 18*(12), 1077–1083. https://doi.org/10.1111/j.1467-9280.2007.02028.x

Lupyan, G., & Swingley, D. (2012). Self-directed speech affects visual search performance. *Quarterly Journal of Experimental Psychology, 65*(6), 1068–1085. https://doi.org/10.1080/17470218.2011.647039

Lupyan, G., & Ward, E. J. (2013). Language can boost otherwise unseen objects into visual awareness. *Proceedings of the National Academy of Sciences, 110*(35), 14196–14201. https://doi.org/10.1073/pnas.1303312110

MacLeod, C. M., Gopie, N., Hourihan, K. L., Neary, K. R., & Ozubko, J. D. (2010). The production effect: Delineation of a phenomenon. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 36*(3), 671–685. https://doi.org/10.1037/a0018785

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought.* Chicago: University of Chicago Press.

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology, 36*(1), 53–65. https://doi.org/10.1017/s0048577299971196

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*(1), 97–113. https://doi.org/10.1016/0028-3932(71)90067-4

Ouwehand, K., van Gog, T., & Paas, F. (2015). Effects of pointing compared with naming and observing during encoding on item and source memory in young and older adults. *Memory, 24*(9), 1243–1255. https://doi.org/10.1080/09658211.2015.1094492

Ouwehand, K., Dijkstra, K., Van Gog, T., & Paas, F. (2019). Effects of pointing gestures on memory for (in) congruent stimuli in children and young adults. *Mind, Brain, and Education, 13*(2), 92–99.

Overkott, C., & Souza, A. S. (2022). Verbal descriptions improve visual working memory but have limited impact on visual long-term memory. *Journal of Experimental Psychology: General, 151*(2), 321–347. https://doi.org/10.1037/xge0001084

Paivio, A. (1986). *Mental representations: A dual coding approach.* Oxford University Press.

Park, B., Korbach, A., Ginns, P., & Brünken, R. (2023). How learners use their hands for learning: An eye-tracking study. *Educational Psychology Review, 35*(4). https://doi.org/10.1007/s10648-023-09833-2

Peeters, D., Chu, M., Holler, J., Hagoort, P., & Ozyürek, A. (2015). Electrophysiological and kinematic correlates of communicative intent in the planning and production of pointing gestures and speech. *Journal of Cognitive Neuroscience, 27*(12), 2352–2368.

Peeters, D., Snijders, T. M., Hagoort, P., & Ozyürek, A. (2017). Linking language to the visual world: Neural correlates of comprehending verbal reference to objects through pointing and visual cues. *Neuropsychologia, 95*, 21–29. https://doi.org/10.1016/j.neuropsychologia.2016.12.004

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology, 118*(10), 2128–2148. https://doi.org/10.1016/j.clinph.2007.04.019

Psychology Software Tools. (n.d.). Psychology Software Tools. https://pstnet.com.

Pika, S. (2012). The case of referential gestural signaling. *Communicative & Integrative Biology, 5*(6), 578–582. https://doi.org/10.4161/cib.22012

Regier, T., & Xu, Y. (2017). The Sapir-Whorf hypothesis and inference under uncertainty. *Wiley Interdisciplinary Reviews: Cognitive Science, 8*(6), e1440.

Richarz, A. (2017, March 29). *Why Japan's rail workers can't stop pointing at things*. Atlas Obscura. https://www.atlasobscura.com/articles/pointing-and-calling-japan-trains.

Rutman, A. M., Clapp, W. C., Chadick, J. Z., & Gazzaley, A. (2010). Early top—down control of visual processing predicts working memory performance. *Journal of Cognitive Neuroscience, 22*(6), 1224–1234. https://doi.org/10.1162/jocn.2009.21257

Shigemori, M., Sato, A., Masuda, T., & Haga, S. (2013). Human error prevention effect of point and call check used by railway workers in Japan. *Rail Human Factors: Supporting Reliability, Safety, and Cost Reduction*, 599–608. https://doi.org/10.1201/b13827-78

Shinohara, K., Naito, H., Matsui, Y., & Hikono, M. (2013). The effects of "finger pointing and calling" on cognitive control processes in the task-switching paradigm. *International Journal of Industrial Ergonomics, 43*(2), 129–136. https://doi.org/10.1016/j.ergon.2012.08.004

Souza, A. S., & Skóra, Z. (2017). The interplay of language and visual perception in working memory. *Cognition, 166*, 277–297. https://doi.org/10.1016/j.cognition.2017.05.038

Stein, B. E., Stanford, T. R., & Rowland, B. A. (2020). Multisensory integration and the society for neuroscience: Then and now. *Journal of Neuroscience, 40*(1), 3–11.

Tomasello, M. (2010). *Origins of human communication.* MIT press.

Vygotsky, L. S. (1987). *The collected works of LS Vygotsky: Volume 1: Problems of general psychology, including the volume Thinking and Speech* (Vol. 1). Springer Science & Business Media.

Violato, E., Chao, I. C. I., McCartan, C., & Concannon, B. (2022). Pointing and calling the way to patient safety: An introduction and initial use case. *Journal of Patient Safety and Risk Management, 27*(2), Article 251604352210780. https://doi.org/10.1177/25160435221078099

Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology, 42*(6), 654–667. https://doi.org/10.1111/j.1469-8986.2005.00356.x

Zhang, S., de Koning, B. B., & Paas, F. (2023). Finger pointing to support learning from split-attention examples. *Educational Psychology, 43*(2–3), 207–227. https://doi.org/10.1080/01443410.2023.2193696

Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience, 12*(1), 24–25. https://doi.org/10.1038/nn.2223

Zhang, Y. (2014). *Online Tool for Handedness Assessment*. Retreived from http://zhanglab.wikidot.com/handedness.