# AWARENESS AND INTEGRATION: UNDERSTANDING THE CHALLENGES OF INFERRING MULTISENSORY INTEGRATION OUTSIDE OF AWARENESS

BY

DANIEL CHRISTOPHER JENKINS

A thesis submitted to the Victoria University of Wellington
in fulfilment of the requirements for the degree of Master of Science
in Cognitive and Behavioural Neuroscience

Victoria University of Wellington
2019

## Abstract

*Multisensory integration* describes the cognitive processes by which information from various perceptual domains is combined to create coherent percepts. For consciously aware perception, multisensory integration can be inferred when information in one perceptual domain influences subjective experience in another. Yet the relationship between integration and awareness is not well understood. One current question is whether multisensory integration can occur in the absence of perceptual awareness. Because there is subjective experience for unconscious perception, researchers have had to develop novel tasks to infer integration indirectly. For instance, Palmer and Ramsey (2012) presented auditory recordings of spoken syllables alongside videos of faces speaking either the same or different syllables, while masking the videos to prevent visual awareness. The conjunction of matching voices and faces predicted the location of a subsequent Gabor grating (*target*) on each trial. Participants indicated the location/orientation of the target more accurately when it appeared in the cued location (80% chance), thus the authors inferred that auditory and visual speech events were integrated in the absence of visual awareness. In this thesis, I investigated whether these findings generalise to the integration of auditory and visual expressions of emotion. In *Experiment 1*, I presented spatially informative cues in which congruent facial and vocal emotional expressions predicted the target location, with and without visual masking. I found no evidence of spatial cueing in either awareness condition. To investigate the lack of spatial cueing, in *Experiment 2*, I repeated the task with aware participants only, and had half of those participants explicitly report the emotional prosody. A significant spatial-cueing effect was found only when participants reported emotional prosody, suggesting that audiovisual congruence can cue spatial attention during aware perception. It remains unclear whether audiovisual congruence can cue spatial attention without awareness, and whether such effects genuinely imply multisensory integration.

## Acknowledgements

A Master's thesis is a work of much time, thought and effort—but that work is made manageable by the guidance and encouragement of many advisors, mentors, colleagues and supporters. I would like to extend a warm and heartfelt thank-you to the many people who have played these roles for me over the past year.

To my supervisor, Reader Gina Grimshaw, I would like to thank you for your guidance, support and encouragement throughout this project. I have learned a great deal from you about how to think critically about my work, how to conduct research productively and with integrity, and how remain open to whatever paths your findings present. In particular, I thank you for your encouragement throughout some of the challenges I have faced, and the reminders that you can continue to do good science even when you don't obtain the results that you expect.

To my partner, Grace, thank you for your constant support and encouragement throughout the year. Your care for me has helped me to stay healthy and hopeful, despite feeling variously challenged, disheartened and overwhelmed. You brighten my days, and I feel incredibly grateful for everything that you do.

Thank you, also, to the various people that provided technical support for various aspects of the project: to Al Abenoja, who worked tirelessly on programming the continuous flash suppression (CFS) experiments in MATLAB, and provided technical guidance throughout; to Michael Tooley, who provided advice and technical assistance around the use of CFS; to David Carmel, for advice on experimental design; and to Angus Chapman and Laina Isler for providing sound statistical advice.

Additionally, thank you to the members of the Cognitive and Affective Neuroscience lab at Victoria University of Wellington, and to Sumaya Lamb, for advice and feedback on various aspects of my project, and for volunteering your time as pilot participants while I was refining the programming for each of my experiments. In particular, thank you to Amy Walsh, Hazel Godfrey, and Sophie Hedley for your suggestions and guidance.

To my family, thank you for your emotional and financial support. Thank you for your encouragement, for your patience for my lack of availability, and your willingness to let me talk about my work even when my enthusiasm outlasted your interest or understanding.

Thank you to my remaining contributors and critics. Particilarly, to my voice actor, Charlotte Cook; and proofreaders, Gordon Jenkins, Siobhan Molloy, and Alan Chapman. Thank you to the organisers and attendees of the 2017 KiwiCAM conference, for the opportunity to present my research and engage with the thoughts and questions of other aspiring scientists. And thank you to everyone else who offered encouragement and support; who helped me to disengage and relax when I needed to recharge; and who let me sound out ideas or helped debrief disasters. I am grateful for all of your assistance.

Finally, I am grateful to have received the Victoria Master’s (by thesis) Scholarship to help meet the financial costs of engaging in this project.

# Table of Contents

## List of Figures

# List of Tables

## Introduction

Our experience of the world is shaped by our senses. We acquire information about ourselves and our environment through many channels (vision, hearing, touch, and so on). Each channel contributes different qualities to our experience. Yet we do not experience these qualities independently, or in isolation. Instead, we combine them into a single, unified perceptual experience (Treisman, 1996), in which various features are bound or attributed to various external objects or causes. In this way, we build a perceptually rich and informative model of the world from the various sensations that occur to us moment-to-moment.

The cognitive processes responsible for this perceptual binding are typically described as *multisensory integration*. These processes rely on the communication of perceptual information from primary sensory regions in the brain (i.e., those responsible for processing information in one specific perceptual domain) to other primary sensory regions (Macaluso, Frith, & Driver, 2000) or to higher-level, multimodal brain regions that respond to information from multiple perceptual domains (Meredith, 2002; Meredith & Stein, 1986). These processes contribute to the content of our experience, by *integrating* the various features of the environment that we perceive.

Yet it remains less clear how *conscious* experience of the world should arise at all. Of all the information sampled by our senses, only a subset enters into conscious awareness and is therefore consciously experienced. It's clear that a relationship exists between consciousness and integration, but the nature of that relationship is not well understood (Mudrik, Faivre, & Koch, 2014). One question, currently being asked, is whether multisensory integration can occur in the absence of perceptual awareness (e.g., Faivre, Mudrik, Schwartz, & Koch, 2014; Palmer & Ramsey, 2012).

In this thesis, I set out to conduct a conceptual replication of a recent study that claimed to show evidence of unconscious multisensory integration (Palmer & Ramsey, 2012), and to investigate whether those findings generalised to the integration of emotional information conveyed by voices and facial expressions. After failing to find evidence that auditory and visual congruence cues spatial attention either *with or without* visual awareness (*Experiment 1*), I shifted my focus to the investigation of the conditions under which cross-modal congruence may cue spatial attention during full perceptual awareness (*Experiment 2*). In doing so, I have attempted to understand how we might validate tasks that assess multisensory integration indirectly, and apply them to the original question of whether integration occurs outside of conscious awareness.

**Multisensory Integration and Perceptual Illusions**

Multisensory integration occurs when information from different sensory domains is combined to produce coherent and unified perceptual representations. Integration binds perceptual qualities that occur close together in time and space (Koelewijn, Bronkhorst, & Theeuwes, 2010) by attributing those sensations to a common source. This attribution occurs effortlessly, such that our aware experience of the world comprises percepts that appear already integrated. That is, the perceptual features of the world occur to us as features of distinct objects or events, without requiring deliberate attributions.

Perceptual properties are processed via functionally specialised pathways and neurons. For instance, in early visual processing, single cells track specific properties of objects in the visual scene, such as orientation (Hubel & Wiesel, 1959, 1977) or movement (Schiller, Finlay, & Volman, 1976). Later stages of perceptual processing combine this information into increasingly complex representations in heteromodal brain regions (Calvert & Thesen, 2004; Macaluso & Driver, 2005; Stein & Stanford, 2008), including the superior colliculus (Leo, Bertini, Di Pellegrino, & Làdavas, 2008; Meredith & Stein, 1996; Wallace & Stein, 1997, 2001), superior temporal sulcus (Barraclough, Xiao, Baker, Oram, & Perrett, 2005; Calvert, Campbell, & Brammer, 2000) and insula (Bushara et al., 2003). Yet these feedforward convergence processes—from early unimodal to later heteromodal processing—are not the only neural correlates of multisensory integration. Multisensory integration also involves crosstalk or feedback mechanisms that permit signalling between unimodal sensory areas (Calvert et al., 2000; Driver & Spence, 2000; Senkowski, Talsma, Grigutsch, Herrmann, & Woldorff, 2007). Collectively, these processes reduce perceptual noise in the environment (Stein, Stanford, Wallace, Vaughan, & Jiang, 2004), allowing us to pick out salient information more easily—e.g., audiovisual integration facilitates identification of purely visual targets (Noesselt, Bergmann, Hake, Heinze, & Fendrich, 2008; Van der Burg, Olivers, Bronkhorst, & Theeuwes, 2008).

Typically, we remain unaware of these processes. Because multisensory integration occurs automatically and pre-attentively (Massaro, 1987; Soto-Faraco, Navarra, & Alsius, 2004; Spence & Driver, 2000; Vroomen, Bertelson, & De Gelder, 2001), our perceptual experience is of a world of coherent objects, for which the individual features appear already bound together.

Yet, sometimes we misattribute sensory stimuli to an incorrect source, giving rise to perceptual illusions. Cross-modal illusions can occur when events in two or more sensory domains appear in close proximity in time and space. For instance, ventriloquism effects occur when individuals misattribute the source of a sound to a co-occurring visual source (Thurlow & Jack, 1973). Following this misattribution, an individual’s subjective experience of the spatial location of the sound is pulled towards the spatial location of the attributed source and away from its actual spatial origin. For instance, subjective judgements about the spatial origin of a simple auditory tone can be involuntarily shifted towards a brief flash of light, provided that the two events occur concurrently (Slutsky & Recanzone, 2001). You may experience something similar when the voice of a film actor appears to emanate from the character on the screen, rather than from speakers that are spatially offset—perhaps even behind you. In such cases, the voice (i.e., auditory stimulus) and the actor’s lip movements (i.e., visual stimulus) are temporally aligned, but spatially offset. When you attribute these events to the same source—the actor—your subjective experience of the spatial position of the sound is pulled towards the spatial location of the actor onscreen. (Conversely, when the auditory and visual streams of a film fall out of synchrony, viewers may find it difficult to attribute the voices they hear to the actors that they see.)

Some percepts, especially those that convey semantic information, may occur in congruent or incongruent pairings. For instance, happy voice prosodies canonically occur in conjunction with happy facial expressions—and it would be unusual to hear a particularly happy voice projected from an unhappy face, or vice versa. When this is the case, multisensory integration can help us interpret otherwise ambiguous percepts. For instance, in noisy environments, it can be difficult to discriminate the content of speech from sounds alone—but attending to a speaker’s lip movements can improve auditory perception (Sumby & Pollack, 1954). Additionally, the co-occurrence of congruent auditory and visual events more effectively captures attention than auditory or visual events alone (Spence & Santangelo, 2009).

These observations suggest that multisensory integration can shape percepts in one sensory domain in response to co-occurring stimuli in another domain. Further evidence is provided by a different perceptual illusion—known the McGurk effect—that arises when a voice is misattributed to an incompatible source. The McGurk effect occurs when incongruent auditory and visual recordings of spoken consonant–vowel syllables are presented together. For some pairings of visual and auditory speech events, participants report hearing an illusory, fused syllable (McGurk & MacDonald, 1976). For instance, when participants listened to an auditory recording of a voice saying “ba-ba” alone, they had no trouble correctly reporting the sounds that they heard. However, when the same voice recording was presented with a concurrent video of a face saying “ga-ga”, participants reported hearing the fused—and illusory—syllable “da-da”. Because the sound “ba” and lip movements “ga” are incompatible, participants’ subjective auditory experience is shifted towards their visual experience in an apparent attempt to make sense of the competing information. Thus, in this case, the subjective experience of hearing spoken words is influenced by the automatic integration of auditory and visual percepts.

Audiovisual integration of faces and voices also occurs during the perception of emotion (Massaro & Egan, 1996). For instance, Collignon et al. (2008) reported that people discriminate emotion in congruent face–voice pairs (i.e., same emotion) faster and more accurately than in either facial expressions or emotional voices alone. When face–voice pairs were emotionally incongruent, participants predominantly reported the emotion expressed in the face rather than the voice, suggesting a visual advantage for emotion perception. However, when participants were instructed to report only the face emotion, and the face images were degraded, subjective perception of visual emotion shifted towards the emotion expressed by the voice. Similarly, De Gelder and Vroomen (2000) generated emotional face images along a morph continuum between *happy* and *sad*, thus creating emotional facial expressions of varying ambiguity. They presented face morphs with concurrent happy or sad emotional voices, or with no voice. Participants were instructed to report only the emotion expressed by the face, ignoring the voice. Faces were more frequently judged to be sad when presented with a sad voice, and to be happy when presented with a happy voice—especially for otherwise ambiguous expressions. These studies suggest that, in contrast to the McGurk effect, auditory information can also influence visual perception.

These perceptual illusions provide strong behavioural evidence for the occurrence of multisensory integration during aware perception (i.e., integration of percepts that are consciously experienced). Such illusions also show that multisensory integration contributes to the contents of our conscious experience, or even shapes it. However, the nature of the relationship between awareness and integration remains unclear (Mudrik et al., 2014).

**Conscious and Unconscious Perception**

What is clear, is that perception is not entirely a conscious process. The primary sensory cortices receive and process much more information than we experience moment-to-moment. Indeed, there is much evidence to suggest that many perceptual processes persist in the absence of awareness. For instance, in blindsight—a condition of complete cortical blindness resulting from lesions in the primary visual cortex—the phenomenal experience of vision is extinguished. Despite lacking visual awareness, both human patients and lesioned monkeys have shown preserved physiological and forced-choice responses to visual targets (Stoerig & Cowey, 1992, 1995, 1997), suggesting that visual information can be processed and made available to other cognitive processes even in the absence of conscious awareness.

Individuals with no perceptual impairments also process perceptual information without awareness. For instance, participants asked to attend to one of two competing speech streams, presented dichotically (i.e., one stream to each ear) may not consciously perceive words spoken in the unattended ear (Cherry, 1953) and yet find that salient words and phrases still capture attention unintentionally (e.g., Gray & Wedderburn, 1960; Treisman, 1960). Such observations suggest that attention may mediate which features of our environment arise into conscious awareness and which go unnoticed.

More recent empirical work has provided a wealth evidence for unconscious perceptual processing in healthy individuals, especially in the visual domain. For instance, some forms of semantic priming persist when visual primes are masked from conscious awareness (Almeida, Mahon, & Caramazza, 2010; Almeida, Mahon, Nakayama, & Caramazza, 2008; Almeida, Pajtas, Mahon, Nakayama, & Caramazza, 2013). Almeida and colleagues (Almeida et al., 2010; Almeida et al., 2008) presented images of various objects (*primes*) masked by continuous flash suppression (CFS; Carmel, Arcaro, Kastner, & Hasson, 2010; Tsuchiya & Koch, 2005), and measured participants' response times for the categorisation of subsequent, *supraliminal* (i.e., consciously perceived) target images. For some image categories, such as tools, participants were faster to categorise the target following a congruent prime, than an incongruent prime, suggesting that the category of the masked image was processed outside of awareness. Interestingly, while priming occurred for images of tools, it did not occur for images of animals and vehicles. Almeida and colleagues attributed this dissociation to unconscious dorsal stream activation (including the activation of the motor cortices which may be involved in the processing of tool-related information), but not ventral stream activation. More broadly, these findings suggest that semantic priming may occur via dissociable cognitive processes, and that some operations for perceptual stimuli may require conscious awareness while others do not. Separately, Almeida et al. (2013) obtained emotional priming effects of masked images of emotional facial expressions on preference judgements of novel objects. Participants judged objects that followed happy faces more favourably and objects that followed angry faces less favourably, compared to preferences for objects with no emotional prime. Taken together, these findings suggest that masked semantic information for visual stimuli, including emotional information, can be processed without awareness.

While unconscious priming studies infer unconscious visual processing from the influence of masked primes on later aware behaviour, physiological studies can provide an alternative measure. In our own lab, experimenters presented highly arousing images of erotica and mutilation scenes (i.e., positively and negatively valenced emotional images, respectively) under visual masking by CFS (Tooley, Carmel, Chapman, & Grimshaw, 2017). They found that some physiological markers of emotional arousal (increased skin conductance, appetitive priming) persisted in the absence of subjective awareness, while other responses (heart-rate deceleration, defensive priming) did not. Thus, some physiological processes may be activated during unconscious perception, while others may critically depend on awareness.

Finally, converging evidence is provided by neuroimaging studies. Functional magnetic resonance imaging (fMRI) has shown sustained and object-specific activation of dorsal pathways in the absence of visual awareness (Fang & He, 2005). In contrast, activation of the fusiform face area (FFA), which is known to preferentially respond to visual representations of faces, seems to be contingent on visual awareness (Tong, Nakayama, Vaughan, & Kanwisher, 1998). Additionally, research using single-cell recordings of neural activation in monkeys has suggested that activation of neurons in the striate cortex is typically insensitive to awareness (Leopold & Logothetis, 1996; Logothetis & Schall, 1989), while activation of areas involved in the later stages of visual processing, such as the superior temporal sulcus, tended to arise only during conscious perception (Sheinberg & Logothetis, 1997). These findings provide further support for the hypothesis that early perceptual processing can occur in the absence of awareness, but only some information may be made available to downstream cognitive processes.

Of particular relevance to this thesis, a collection of behavioural, physiological and neuroimaging studies have suggested that emotional information can be abstracted from images of human faces outside of conscious awareness (reviewed by Diano, Celeghin, Bagnis, & Tamietto, 2017; Izatt, Dubois, Faivre, & Koch, 2014; Pessoa, 2005). For instance, viewing masked emotional faces can induce unconscious facial mimicry (Dimberg, Thunberg, & Elmehed, 2000; Tamietto et al., 2009) and activate subcortical pathways that preferentially respond to emotional information (Lerner et al., 2012; Liddell et al., 2005; Morris, Öhman, & Dolan, 1999; Whalen et al., 1998). Additionally, blindsight patients have been reported to discriminate emotional facial expressions with greater-than-chance accuracy despite reporting no subjective awareness of the face image (De Gelder, Vroomen, Pourtois, & Weiskrantz, 1999; Tamietto et al., 2009). Finally, there is already some evidence that unconsciously recognised emotional faces may be integrated with emotional voices, as blindsight patients' recognition of emotional vocal prosody can be modulated by the presentation of images depicting emotional faces (De Gelder, Pourtois, & Weiskrantz, 2002).

Such investigations of conscious and unconscious cognitive processes in humans has been greatly aided by the development of several new perceptual masking techniques (Breitmeyer, 2015; Kouider & Dehaene, 2007). The development of CFS (Tsuchiya & Koch, 2005) is particularly useful for the investigation of unconscious visual perception because the technique allows for relatively sustained visual masking, with suppression of masked images potentially lasting several minutes (Carmel et al., 2010). In CFS, masking occurs as a result of interocular competition ("binocular rivalry") that arises when two competing images are simultaneously presented, one to each eye. Typically, when each eye views a different image, the images appear to oscillate—one image appears briefly stable, before fading out and being replaced by the other image, and so on (Blake & Logothetis, 2002; Lin & He, 2009). Binocular rivalry occurs for images of comparable intensity (i.e., with similar low-level visual features, such as contrast, luminance and motion). CFS exploits this phenomenon by presenting one low-contrast target image in competition with a high contrast, dynamic mask (usually comprised of colourful patches that appear and disappear at random; Carmel et al., 2010). Under these conditions, the dynamic mask dominates perceptual awareness, allowing the masked target to go unnoticed for a sustained period of time (Faivre, Berthet, & Kouider, 2012; Yang, Brascamp, Kang, & Blake, 2014).

The advent of these tools for the sustained presentation of visual images outside of awareness has allowed researchers to ask new questions about which cognitive processes persist in the absence of awareness. Unsurprisingly, one of these questions is whether multisensory integration occurs when information from one—or both—perceptual domains is unconsciously perceived.

## Consciousness and Integration

The possibility of unconscious multisensory integration is of particular interest because it falls at the intersection of various lines of scientific inquiry. For perception researchers, unconscious integration would extend our scientific knowledge of the ways in which sensory processing occurs, and may provide new insights into how percepts are generated (or not generated). For consciousness researchers, the occurrence of unconscious multisensory integration—or conversely, the absence of integration for unconscious information—has important implications for various theoretical models of consciousness.

Although several competing theories of consciousness have been proposed, integration features heavily in all of them (Mudrik et al., 2014). Present in all three leading models of consciousness, is the suggestion that conscious awareness cannot be explained by functional specialisation alone—that is, we should not expect to find a seat of consciousness buried deep inside some substructure of the brain. Rather, consciousness arises when long-range and complex integration processes occur. Multisensory integration is one such process (or class of processes), though other instances of information sharing within the brain are also *integrative* by this definition.

Broadly, the current theories make three competing claims about the relationship between consciousness and integration. First, consciousness may be necessary for integration to occur at all (e.g., Baars, 1999; Baars, 2002; Dehaene & Naccache, 2001). Second, integration may a necessary but not sufficient condition for consciousness (e.g., Engel, Fries, König, Brecht, & Singer, 1999; Kanwisher, 2001; Marcel, 1983). Third, consciousness and integration may be identical (Balduzzi & Tononi, 2008; Tononi, 2008; Tononi & Edelman, 1998).

Advocates of the first claim—that *consciousness is necessary for integration*—suggest that conscious awareness arises as the result of a “global broadcasting” of information from otherwise functionally distinct brain modules to a global workspace

(Baars, 1999, 2002; Changeux & Dehaene, 2008; Dehaene & Changeux, 2003, 2005, 2011; Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006; Dehaene, Charles, King, & Marti, 2014; Dehaene, Kerszberg, & Changeux, 1998; Dehaene & Naccache, 2001; Sergent & Dehaene, 2004). Without this broadcasting—*without awareness*—information is available only within these task-specific, unimodal modules. Thus consciousness is adaptive *because* it enables the integration of information from diverse cognitive processes, such as unimodal perceptual processing. Such theories may be collectively referred to as *global workspace theories*.

One prominent example of this claim is made by Global Neuronal Workspace theory (GNWT; Dehaene et al., 1998; Dehaene & Naccache, 2001; Sergent & Dehaene, 2004), which proposes that a network of long-range axons connects disparate and functionally specialised regions of the brain, facilitating global communication and giving rise to conscious awareness. These specialised brain regions may engage in unimodal, task-specific processes in the absence of consciousness—but cannot exchange information without it (except in a limited, feedforward fashion). When activity in these areas exceeds some critical threshold, it catalyses a global "ignition" of activity across the global workspace, which raises the outputs of that module to conscious awareness. Only then do these outputs become available for integration with information in other cognitive modules. Thus, consciousness facilitates integration, allowing information that would otherwise remain isolated to be combined and used in conjunction. This view predicts that multisensory integration should occur only for perceptual information that is consciously experienced.

The second claim—that *integration is necessary for consciousness*—is typically suggested as the consequence of two prior observations about perception. First, the contents of aware experience occur to us already integrated. That is, we do not experience the different perceptual features of objects in our environment independently: they appear to us already bound together, or bound to particular objects or events. This suggests that we cannot experience complex stimuli without integration.[1] This account suggests, prima facie, that whatever processes facilitate integration must occur prior to whatever processes facilitate awareness. Second, the various instances of unconscious perception

[1] A counterpoint is that some percepts may require minimal (if any) integration, such as the detection of a single point of light or the experience of a homogenous field of colour (Mudrik et al., 2014).

(e.g., blindsight) suggest that some perceptual binding, at least within unimodal sensory domains, must occur in the absence of awareness. For instance, if some form of visual integration did not occur unconsciously, we should not observe physiological responses to emotional images (as in Tooley et al., 2017) or priming for tools (as in Almeida et al., 2010; Almeida et al., 2008) in the absence of visual awareness.

The claim that integration may be necessary for consciousness has also been stated on logical grounds. Kanwisher (2001) argues that while neuroscientists should pursue the *neural correlates* of consciousness in order to understand awareness, we should be cautious in proposing causal explanations from neuroimaging data alone. If awareness covaries with activation of specifically multimodal—and thus integrative—brain regions, this may suggest that integration is *necessary* for awareness without saying anything about its *sufficiency*.[2]

Together, these observations—and the hypothesis that integration is a necessary, but not sufficient, prerequisite for awareness—predict that integration *should* occur in the absence of awareness, while awareness should *never* occur in the absence of integration. Given our current understanding that perception can occur unconsciously, this view suggests that we should expect multisensory integration to also occur unconsciously.

The third claim—that consciousness and integration are identical—suggests that any complex system is conscious to the extent that information integration occurs. Tononi (2008) proposes a calculus for the amount of integrated information in a system, which he terms Φ. Systems with maximal Φ are maximally conscious. However, this *Integrated Information Theory* (ITT) has been criticised on the grounds that it lacks explanatory power, and implies panpsychism and mind-body dualism (Cerullo, 2015; Mindt, 2017; Pockett, 2014; Searle, 2013). As far as I am aware, no empirical test of ITT has been attempted—though Tononi and Koch (2015) argue that empirical testing is possible in principle. Although ITT may be taken to imply that neither awareness nor integration should occur independently, it can be difficult to test this in practice because the theory holds that smaller integrative networks remain independently conscious even if our ongoing conscious experience reflects only the largest integrated network within the brain

[2] While Kanwisher (2001) prefers the explanation that integration gives rise to awareness, it is worth noting that the logical argument could also be applied the other way around. That is, the covariance of awareness and integration may suggest that awareness is necessary, but not sufficient, for integration.

at any given point in time (Tononi, 2008). To avoid the theoretical and empirical challenges of validating ITT, I have restricted the scope of this thesis to examining only the first two claims.

## Empirical Investigations of Unconscious Multisensory Integration

One straight-forward test of whether multisensory integration occurs in the absence of perceptual integration, is to ask whether multisensory perceptual illusions, such as the McGurk effect, persist when stimuli in the *influencing modality* are presented outside of awareness. Palmer and Ramsey (2012) investigated this possibility by presenting videos of the lower half of a speaking face (visual speech events) under CFS masking, alongside supraliminal auditory recordings of speaking voices (auditory speech events). Both visual and auditory speech events included syllable pairs (e.g., “ba-ba”). If multisensory integration acted completely independently of awareness, we should expect participants to report hearing illusory fused syllables for incongruent auditory–visual speech pairs, just as they do when they are aware of the lip movements (McGurk & MacDonald, 1976). Instead, Palmer and Ramsey (2012) found no evidence that suppressed visual speech stimuli influenced auditory perception at all.

On their own, such findings might suggest that multisensory integration occurs only when all perceptual stimuli are consciously perceived. However, Palmer and Ramsey (2012) reasonably point out that, according to global workspace theories, the expected auditory illusion would require the global broadcasting of *visual information* that was not consciously perceived, which those theories explicitly rule out. Conversely, the theory predicts that only the consciously perceived stimulus—in this case, the voice—should be globally broadcast for integration. Thus, it may still be possible for the integration of auditory and visual information to occur within the constraints of visual-domain processes, but not elsewhere. That is, consciously perceived auditory information could influence unconscious visual processing, but unconsciously perceived visual information should not influence conscious auditory perception.

In light of this possibility, a more appropriate measure of *integration without awareness* would be to investigate whether audiovisual stimulus pairings—where the visual information is unconsciously perceived and auditory information is consciously perceived—can influence behaviour for some other task that operates specifically within the visual domain. If it did, we could infer multisensory integration for the two stimuli indirectly. Specifically, Palmer and Ramsey (2012) suggested that the presentation of matching or mismatching face–voice pairs might influence the allocation of spatial attention within the visual domain, provided that their conjunction provided an informative spatial cue. If participants allocated spatial attention in response to audiovisual congruence, despite remaining unaware of the visual cue, this might reasonably imply that the auditory and visual speech events were integrated in the absence of visual awareness.

To investigate this possibility, Palmer and Ramsey (2012) developed a novel spatial attention task, using CFS to prevent visual awareness. In the classic spatial attention paradigm, a central visual cue (an arrow pointing left or right) precedes a lateral visual target (Posner, 1980). The arrow correctly predicts the target location for 80% of trials (*valid* cues) and indicates the wrong location for the remaining 20% of trials (*invalid* cues). Participants typically respond more quickly, or with greater accuracy, for validly cued targets, as they can voluntarily attend to the cued location before the target even appears. In the adapted paradigm, Palmer and Ramsey (2012) presented two simultaneous video streams depicting the lower half of faces speaking short strings of syllables (e.g., “ba-ba”) on either side of a central fixation cross. Both videos were shown to a single eye, while two simultaneous CFS masks were shown to the other eye, in the corresponding positions.[3] On each trial (Figure 1), the two masked videos displayed faces speaking two *different* syllable strings. At the same time, an unmasked—and therefore consciously perceived—voice spoke a single syllable string that matched exactly one of the two masked video streams (i.e., the voice spoke the same syllables as one of the two masked faces). Following a short interstimulus interval (ISI) of 110 ms, a low-contrast Gabor grating—the *target*—was presented for 70 ms, without masking. Gratings were oriented either rightward or leftward (rotated 45° clockwise or anticlockwise from vertical, respectively) and appeared in the same location as one of the preceding video streams (to the left or right of fixation).

[3] This dual-CFS method was pioneered by Jiang, Costello, Fang, Huang, and He (2006) while investigating unconscious cueing of *exogenous* (i.e., involuntary) spatial attention. Palmer and Ramsey (2012) adapted the stimuli so as to cue *endogenous* (i.e., voluntary) attention.

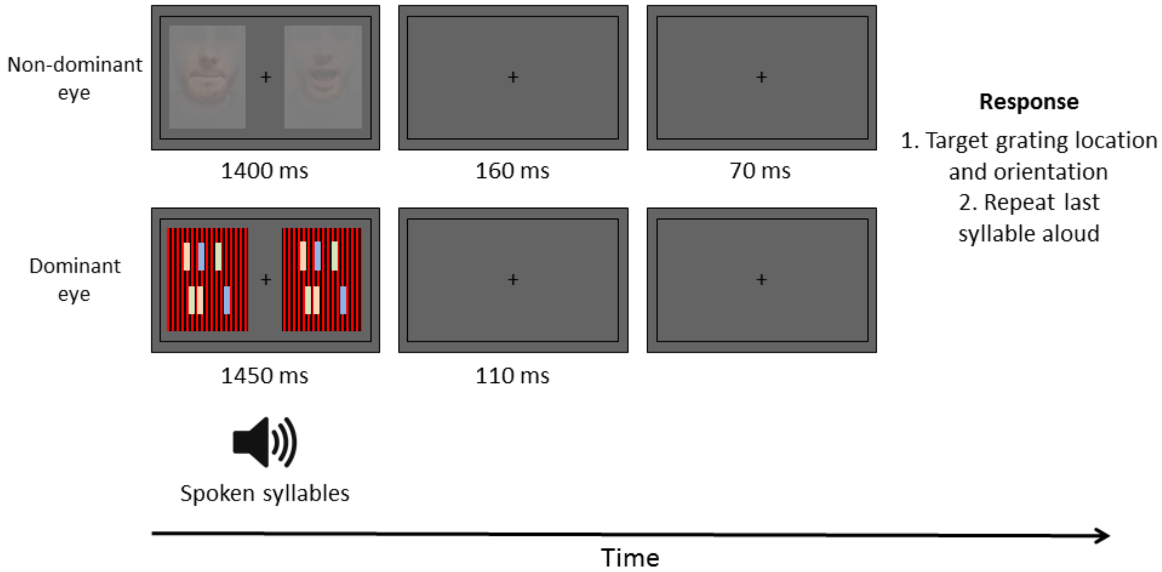


*Figure 1*. Trial diagram for the original congruence-cued spatial attention task (Palmer & Ramsey, 2012).

Participants were asked to accurately report the location and orientation of the target gratings. Unbeknownst to the participants, the congruence between the heard voice and masked lip videos predicted the most likely location of the target. On 80% of trials, the target grating appeared in same location as the congruent lip movements (*valid* trials); on 20% it appeared in the location of the incongruent lip movements (*invalid* trials). Thus, the congruence of the auditory and visual speech events served as an informative cue to the most likely location of the target. Shifting spatial attention in the direction of the congruent face would therefore lead to more accurate target discrimination on valid trials, because the participants would already be attending to the target location before the target even appeared. Conversely, accuracy would be reduced for invalid trials, because participants would be attending to the wrong location, and would need to shift their attention before being able to process the orientation of the target. Greater accuracy for valid trials than for invalid trials would therefore indicate that participants had allocated their attention to the location of the congruent face video. Critically, because the informative feature of the audiovisual cues was the congruence between the face and voice, participants would only be able to allocate attention in this way if they could discriminate congruent and incongruent face–voice pairs in the absence of visual awareness. Thus, a spatial-cueing effect—shown by an accuracy advantage for valid trials—would indicate that multisensory integration had occurred in the absence of visual awareness. Conversely, the authors reasoned, if multisensory integration did not occur, then participants should have no way of discriminating congruent and incongruent locations and thus be unable to allocate attention in accordance with the cues, and would therefore show equal accuracy for valid and invalid trials.

In three experiments, the authors found greater accuracy for valid trials than for invalid trials (Palmer & Ramsey, 2012), suggesting that the faces and voices had been integrated outside of awareness. In each experiment, they varied the paradigm slightly. In the first experiment, congruent face–voice pairs predicted the location of the target, requiring participants to allocate attention to the congruent location. Thus, the significant advantage for validly cued trials may have been due to either the top-down allocation of spatial attention in response to an informative cue (*endogenous cueing*) or to attentional capture by the masked, congruent video stream (*exogenous cueing*). In the second experiment, the cue contingency was reversed, such that incongruent face–voice pairs predicted the location of the target (i.e., 80% of targets appeared in the location of the incongruent lip movements). Participants' accuracy remained greater for validly cued targets, which now appeared in the incongruent location—suggesting that the effect reflected endogenous orienting in response to the cue that predicted the location of the target (in this case the incongruent face-voice pair), and not exogenous orienting to the face that matched the voice. In the third experiment, congruent face–voice pairs again cued target location, and the ISI between the offset of the face–voice cue and onset of the target grating was increased to 500 ms (from 100 ms). Because involuntary attention capture occurs only over short time frames (Müller & Rabbitt, 1989), the observed spatial-cueing effects should only persist over longer ISIs if cueing was endogenous. Again, participants' accuracy was greatest when targets appeared in the cued location, suggesting that attention orienting occurred voluntarily in response to the spatially informative audiovisual cues.

Taken together, all three experiments suggest that participants were sensitive to audiovisual congruence in the absence of visual awareness, consistent with the hypothesis that conscious auditory and unconscious visual information may be integrated within the visual domain. Thus, the authors concluded that unconscious multisensory integration can occur only when the "influencing modality" is consciously perceived. While these findings do constitute evidence of unconscious multisensory integration, the requirement that the influencing modality is consciously experienced means that the reported effects remain consistent with the predictions of global workspace theories. Indeed, Palmer and

Ramsey (2012) explicitly argued that their findings provide evidence in support of these models.

However, other studies have suggested that multisensory integration—evidenced by the discrimination of crossmodal congruence—may occur in the total absence of perceptual awareness. Faivre et al. (2014) presented masked audiovisual primes—variously masking the visual, auditory or both components across three experiments—and supraliminal targets in a variation of the congruency-priming paradigm (Van Opstal, Gevers, Osman, & Verguts, 2010). Participants in this study indicated whether audiovisual targets, composed a written letter (e.g., *m*) and spoken letter name (e.g., *em*), contained matching or mismatching spoken and written letters (i.e., a congruence judgement). Each target was preceded by an audiovisual prime, composed of a written numeral (e.g., *2*) and spoken number (e.g., *two*) that were either matching or mismatching. When participants consciously perceived both audiovisual primes and targets, they judged the congruence of the targets more quickly following primes with identical congruence relationships. That is, when primes presented matching written and spoken numbers, participants identified matching letter pairs more quickly than mismatching pairs; when primes presented mismatching numbers, participants identified mismatching letters more quickly. Thus, participants' congruence judgements for the targets were *primed* by the congruence of the preceding audiovisual number pairs. The authors then systematically masked (a) written number primes only, (b) spoken number primes only, and (c) both written and spoken number primes, using a combination of visual and auditory masking techniques. Congruence priming—that is, faster responses for targets that followed primes with the same congruence relationship—persisted in all three masking conditions, suggesting that multisensory integration of the audiovisual primes occurred even in the complete absence of both auditory and visual awareness.

While the findings reported by Faivre et al. (2014) and Palmer and Ramsey (2012) both suggest that unconscious perceptual information may be integrated, they have different implications for global workspace theories. In the congruence-cued spatial attention task (Palmer & Ramsey, 2012), evidence for multisensory integration was only found when the influencing modality was consciously perceived, consistent with global workspace theories. These findings support the hypothesis that only consciously perceived stimuli are globally broadcast, and therefore available for integration within other functionally specialised cognitive modules. In contrast, in the unconscious congruence priming task (Faivre et al., 2014), crossmodal congruence priming persisted in the complete absence of both auditory and visual awareness, suggesting that consciousness is not a prerequisite for multisensory integration, contrary to the predictions of global workspace theories.

Both studies join a growing literature reporting auditory–visual integration effects outside of conscious awareness (e.g., Alsius & Munhall, 2013; Chen & Spence, 2010, 2011). However, we should be cautious in inferring multisensory integration until we can demonstrate reproducible effects and rule out alternative explanations for these findings. Indeed, recent findings of unconscious perceptual integration within the visual domain (Mudrik, Breska, Lamy, & Deouell, 2011; Mudrik & Koch, 2013), inferred from scene–object congruence judgements, could not be reproduced in later replication attempts (Biderman & Mudrik, 2017; Moors, Boelens, van Overwalle, & Wagemans, 2016). For novel paradigms, such as the congruence-cued spatial attention task, replication is especially important, to ensure that the reported findings are real. Additionally, such studies must provide clear evidence that perceptual masking has successfully suppressed subjective awareness, and provide some evidence that the chosen task genuinely assesses the cognitive mechanisms for which it was designed (e.g., multisensory integration).

In this thesis, I applied these requirements to the validation of the congruence-cued spatial attention task and to replicating the spatial-cueing effects reported by Palmer and Ramsey (2012). Specifically, I conducted a conceptual replication of the congruence-cued spatial attention task with a novel set of cue stimuli and the introduction of a comprehensive suppression check to ensure successful CFS masking. If spatial cueing by cross-modal congruence genuinely requires multisensory integration, we should expect to find spatial-cueing effects for any audiovisual cues for which multisensory integration effects are known to occur (e.g., emotional expressions in faces and voices). A conceptual replication of this kind would provide support for the hypothesis that the task assesses multisensory integration. (It would not, however, rule out alternative explanations for the observed spatial cueing; nor would it test the negative predication that congruence-cueing of visual spatial attention should perish in the absence of auditory awareness.

**Experiments**

The primary aim of this thesis was to investigate whether the prior findings of unconscious audiovisual integration (Faivre et al., 2014; Palmer & Ramsey, 2012) would generalise to the audiovisual integration of emotional expressions. The choice of facial and vocal expressions of emotion was made because these stimuli seem well suited for testing unconscious multisensory integration. Importantly, visual and auditory expressions of emotion can be combined into a multimodal stimulus in which the emotional dimensions in each modality are either congruent or incongruent. Emotional expressions in both domains can be readily recognised and categorised by participants (Adolphs, Tranel, & Damasio, 2003; Öhman, Flykt, & Esteves, 2001; Schirmer & Adolphs, 2017), and are known to produce a variety of multisensory integration effects during aware perception, as evidenced by converging behavioural (Collignon et al., 2008; De Gelder & Vroomen, 2000; Gerdes, Wieser, & Alpers, 2014; Massaro & Egan, 1996) and neurological evidence (De Gelder, Böcker, Tuomainen, Hensen, & Vroomen, 1999; Föcker, Gondan, & Röder, 2011). Finally, converging behavioural, physiological and neuroimaging evidence points to the abstraction of emotional information from unconsciously perceived face images (De Gelder et al., 2002; Dimberg et al., 2000; Izatt et al., 2014; Lerner et al., 2012; Liddell et al., 2005; Morris et al., 1999; Pessoa, 2005; Tamietto et al., 2009; Whalen et al., 1998)

At the outset, my expectation was that my experiments would extend our current knowledge of unconscious multisensory integration by the conceptual replication of Palmer and Ramsey (2012), or the failure to replicate their unconscious spatial-cueing effects. However, it became clear that our understanding of the congruence-cued spatial attention task—and the role of multisensory integration in directing attention—is limited. Prior research into cross-modal spatial cueing has been primarily concerned with the cueing of attention in one sensory domain via a unimodal cue in a different modality (e.g., Butter, Buchtel, & Santucci, 1989; Spence, Nicholls, Gillespie, & Driver, 1998), or with involuntary capture of attention by multimodal events (i.e., exogenous cueing; Spence & Santangelo, 2009; Van der Burg et al., 2008). In contrast, Palmer and Ramsey (2012) are the only researchers, so far as I could find, to have investigated *endogenous* spatial cueing using cross-modal congruence as an informative cue to the target location.

When investigating unconscious perceptual processes, it is important to understand how those processes occur during aware perception. When an aware process is well understood, and is then seen to *persist* in the absence of awareness, we can conclude that the underlying cognitive processes occur unconsciously. However, in the case of the congruence-cued spatial attention task, we have no empirical evidence that the observed cueing effects genuinely assess multisensory integration—which would be necessary to validly infer unconscious multisensory integration from the observed spatial-cueing effects. Based on our current knowledge of multisensory integration for audiovisual speech events (e.g., McGurk & MacDonald, 1976) and audiovisual emotional expressions (e.g., De Gelder & Vroomen, 2000), we should expect participants to integrate such stimuli during aware perception. Thus, if the congruence-cued spatial attention task genuinely assesses multisensory integration, perceptually aware participants should show the same spatial-cueing effects that Palmer and Ramsey (2012) have reported for visually unaware participants.

Thus, the aims for this thesis were expanded to include the investigation of spatial cueing by cross-modal congruence in both aware and unaware participants. If spatial cueing were observed for aware participants, this would support the use of the congruence-cued spatial attention task as an indirect indicator of multisensory integration—making it suitable for the investigation of unconscious multisensory integration.

In *Experiment 1*, I adapted the original paradigm to present audiovisual emotional cues to both aware and unaware participants. In my version of the task, the emotional congruence between a spoken voice and one-of-two lateralised facial expressions predicted the location of an upcoming Gabor grating (i.e., target) with 80% reliability. Consistent with Palmer and Ramsey (2012), greater target-discrimination accuracy for valid trials (i.e., for targets that appear in the same location as the congruent emotional face) than for invalid trials (i.e., for targets that appear in the opposite location to the congruent emotional face) would imply successful endogenous cueing to the target location. With this design, I was able to test two separate hypotheses. First, spatial cueing by cross-modal congruence may imply multisensory integration. If this were the case, a significant spatial-cueing effect should be observed for aware participants, regardless of whether that effect persists in the absence of visual awareness. Second, consciously perceived auditory and unconsciously perceived visual information may be integrated, at least within the constraints of visual-domain cognitive processes such as the allocation of spatial attention. If this were the case, a significant spatial-cueing effect should be observed for *both* aware and unaware participants. Unexpectedly, the results of *Experiment 1* revealed no significant spatial-cueing effects in either awareness condition, thus providing no support for either hypothesis.

Thus, in *Experiment 2*, I shifted my focus to the question of whether cross-modal congruence is able to cue endogenous spatial attention during perceptual awareness. A significant spatial-cueing effect in this experiment would show that cross-modal congruence can cue spatial attention, at least during aware perception. As argued for the previous experiment, if the congruence-cued spatial attention task were a suitable test for *unconscious multisensory integration*, then the expected spatial-cueing effect should be obtainable for perceptually aware participants. If aware spatial cueing were not possible, that would imply that the spatial attention paradigm developed by Palmer and Ramsey (2012) does not genuinely assess multisensory integration, and thus cannot be used to answer the question of whether integration can occur outside of awareness.

## Experiment 1

Palmer and Ramsey (2012) reported a spatial-cueing effect that they interpreted as evidence for unconscious multisensory integration of visual and auditory speech events. In this experiment, I adapted their *congruence-cued spatial attention task* to investigate whether their findings would extend to the integration of emotional facial expressions and vocal prosody. If multisensory integration does contribute to spatial cueing by cross-modal congruence, it should be possible to obtain similar cueing effects for other classes of auditory and visual stimuli that are typically integrated during aware perception—such as emotional faces and voices (Collignon et al., 2008; De Gelder, Böcker, et al., 1999; De Gelder et al., 2002; De Gelder & Vroomen, 2000; Föcker et al., 2011; Massaro & Egan, 1996).

To investigate this possibility, I replaced the original visual and auditory speech-event cues with static images of emotional facial expressions and short speech tokens spoken in emotional prosodies. Consistent with Palmer and Ramsey (2012), participants completed a spatial cueing task in which the single emotional voice cue matched exactly one of two concurrently presented emotional face images. The location of the matching (i.e., emotionally congruent) face accurately predicted the location of a subsequent target grating for 80% of trials (*valid* trials) and miscued the incorrect location on the remaining 20% of trials (*invalid* cues). Also consistent with the original paper, trials included a relatively long stimulus onset asynchrony (SOA) between the emotional cues and target gratings to allow time for voluntary orienting towards the cued location (i.e., endogenous cueing). On each trial, participants indicated the location and orientation of the target grating. Both target-discrimination accuracy and response times were recorded for analysis.

While adapting the task, I considered a number of limitations in the original study and made five changes to the procedure in order to address these. Below, I describe each of these changes in turn, and explain the rationale for making them.

*Aware control group.* The original paper Palmer and Ramsey (2012) reported three spatial attention experiments that tested the hypothesis that multisensory integration can occur outside of conscious awareness—none of which included an aware control group.[4] As such, there is no current evidence that cross-modal congruence cues spatial attention during perceptual awareness, and no empirical support for the assumption that a spatial-cueing effect implies multisensory integration. If the spatial-cueing effects reported by Palmer and Ramsey (2012) do genuinely imply multisensory integration, they should replicate for perceptually aware participants. Such a finding would support the claim that the congruence-cued spatial attention task validly assesses multisensory integration, thus legitimising its use to investigate integration in the absence of awareness. Then, if spatial cueing *persisted* under visual masking, it would imply that multisensory integration of unconscious perceptual information can occur—at least within the modality of the unconscious stimulus. This logic—that an unconscious process can be inferred from the persistence of behavioural markers for that process in the absence of perceptual awareness—has typically motivated the inclusion of aware controls in other investigations of conscious and unconscious cognitive processes (e.g., Faivre et al., 2014; Hedger, Adams, & Garner, 2015; Tooley et al., 2017). The inclusion of an aware control condition would therefore both strengthen the claim that the paradigm genuinely assesses multisensory integration, and potentially support the conclusion of unconscious integration if the unconscious cueing effect was replicated.

*Suppression check.* Whenever an experiment involves a manipulation of conscious awareness, it is necessary to ensure that participants in unaware conditions remain subjectively unaware of the masked stimuli (Yang et al., 2014). For continuous flash suppression (CFS; Tsuchiya & Koch, 2005) in particular, unintentional awareness of the masked stimuli—termed *breakthrough*—can occur intermittently throughout an experiment (Carmel et al., 2010). To control for breakthrough, Palmer and Ramsey (2012) had participants complete a post-test questionnaire, intended to probe subjective awareness. Participants who reported seeing “lip movements, faces, or anything resembling lip movements or faces” were excluded from the experiment. Over three experiments, the authors excluded five participants (out of a total of 25).

[4] The authors did report two control experiments that investigated whether their continuous flash suppression (CFS; Tsuchiya & Koch, 2005) protocol adequately masked the speech videos Palmer and Ramsey (2012). However, while these control experiments did contrast aware and unaware presentation conditions, they did not include a spatial cueing task. As such, these experiments cannot provide any indication of whether cross-modal congruence can cue spatial attention during aware perception.

While this post-test questionnaire may have provided some insight into participants' subjective awareness, it requires participants to make a one-off judgement across many trials (which may have included only occasional, fleeting instances of breakthrough). It also relied on participants' accurate recall of earlier events. Our lab's prior experience with CFS (e.g., Tooley et al., 2017) has made us aware that breakthrough is seldom all-or-nothing. Rather, participants tend to experience subjective breakthrough on some trials, but not others—as indicated via trial-by-trial self-report ratings of subjective awareness. To adequately control for breakthrough, Yang et al. (2014) recommend the inclusion of both subjective (i.e., self-report) and objective (e.g., forced-choice discrimination of masked images) measures of subjective awareness in CFS paradigms.

A subjective measure of awareness, such as a trial-by-trial self-report can be easily included. In the case of this experiment, participants were asked whether each trial included only "colours" (i.e., the CFS mask) or also included "other images" (e.g., emotional faces), and rated their experience on a four-point scale (described in the *Procedure* below). However, because self-report measures of awareness are dependent on accurate self-report, they can underestimate the number of trials in which participants become partially aware of the masked images. Thus, it is useful to contrast subjective awareness with a more objective measure.

In objective measures of awareness, participants are typically considered unaware of a masked stimulus when they are unable to discriminate features of the stimulus above chance accuracy (Cheesman & Merikle, 1986; Holender, 1986; Yang et al., 2014). For example, Tooley et al. (2017) presented a single CFS-masked image on each trial—either an emotional scene or a meaningless scramble—and asked participants to indicate the content of the image at the end of each trial. While this kind of trial-by-trial objective suppression check is ideal, it is not always possible to incorporate during an experimental task. For this experiment, however, the 4:1 ratio of valid-to-invalid trials is a critical feature of the spatial cueing paradigm, so it would be impossible to include scrambled images (which would have no cue information at all) during experimental trials without disrupting the cueing manipulation. Therefore, I included a post-experiment task to measure objective awareness, which was then used to validate participants' subjective ratings of breakthrough.

*Target calibration.* For accuracy to be a useful dependent variable, it is necessary to control the task difficulty to maintain accuracy between floor (i.e., chance) and ceiling (i.e., perfect accuracy). Participants should reliably make some errors, but not too many. Because perceptual sensitivity differs between individuals, this may require calibrating the task difficulty separately for each participant, or excluding a number of participants who fall outside the desired accuracy range.

Palmer and Ramsey (2012) calibrated task difficulty by increasing or decreasing the contrast of the target gratings periodically throughout the experiment. Whenever participants' mean accuracy rose too high, contrast was reduced; when accuracy fell too low, contrast was raised. Using this dynamic calibration method, they obtained mean accuracy scores within the desired range (30–70%) for most participants. Contrast adjustments were made periodically, such that the ratio of valid to invalid trials (4:1) remained constant for each target level—thus avoiding a confound of target contrast.

However, a consequence of this procedure is that participants' mean accuracy scores were averaged over unequal numbers of trials for each level of target contrast. Unless we can be sure that participants engaged the same strategies and cognitive processes for each level of target contrast, then dynamic calibration may have conflated different behaviours. To avoid this issue, it is typical to calibrate task difficulty before a participant begins the experimental task (e.g., Rohenkohl, Gould, Pessoa, & Nobre, 2014), and to then use the calibrated set of stimuli throughout all experimental trials. For this reason, I calibrated task difficulty at the outset of the experiment, in a pre-experimental block of trials, and not during experimental trials.

*Sample sizes.* Palmer and Ramsey (2012) reported consistent and significant congruence-cued spatial attention effects across three experiments, however the power of each experiment is limited by its small sample size. After excluding five participants who reported awareness of the masked face videos, and a further three who were unable to discriminate target gratings above a predefined threshold for chance accuracy (30%), only 17 participants remained across all three experiments ($N_1 = 5$, $N_2 = 7$, $N_3 = 5$). Thus, each experiment had low power to detect a real spatial-cueing effect. While all three studies found significant effects of spatial cueing (suggesting that the effect is robust), underpowered studies run a heightened risk of finding spuriously significant results (Button et al., 2013), making each reported effect individually less compelling. Moreover, significant effects obtained with small sample sizes—even when they reflect a true difference—are likely to overestimate the effect size (Lakens & Evers, 2014), making it difficult to determine the power of future replications to detect the same effect. To address both concerns, I used conservative effect-size estimates extrapolated from the original paper (Palmer & Ramsey, 2012) to determine an appropriate sample size for a conceptual replication—ultimately increasing the sample size to 24 participants per condition.

*Verbal task.* In the original study, participants repeated the last syllable of the spoken voice cue aloud at the end of each trial (Palmer & Ramsey, 2012). Responses were not recorded, nor were they analysed, but the task was included to encourage participants to attend to the spoken voice cue. In adapting the task for this experiment, I elected to remove this verbal task for two reasons. First, I expected participants would process the emotional prosody of the voice tokens automatically, as has been suggested by electrophysiology experiments (e.g., Wambacq, Shea-Miller, & Abubakr, 2004). It therefore seemed unlikely that a spatial-cueing effect would critically depend on participants having explicitly identified the cue-relevant features (i.e., prosody) of each voice token. Second, requiring participants to explicitly identify the emotional prosody on each trial would add an additional task to the experiment—potentially increasing cognitive load to the point where the spatial cueing might be impacted. While participants in the original experiments showed significant spatial-cueing effects despite engaging in two different tasks on each trial (i.e., target discrimination and verbal report), including the verbal report in the current experiment would require unaware participants to engage in three different tasks on each trial (i.e, target discrimination, subjective awareness probe, and discrimination of emotional prosody. Thus, having participants explicitly identify the emotional prosody in the current experiment, may contribute nothing to participants' ability to integrate the audiovisual cues, while simultaneously and unnecessarily adding to participants' cognitive load.

Thus, participants performed an adapted version of the congruence-cued spatial attention task with novel audiovisual cues (i.e., emotional faces and voices), and a number of methodological changes. The cueing contingency of the original task was preserved, such that the emotional congruence of face images and voice tokens predicted the location of the target grating with 80% validity. For each trial, participants indicated both the location (left or right) and orientation (leftward or rightward) of the target grating.

Participants were assigned to one of two awareness groups. Aware-group participants completed trials in which the emotional face images, emotional voice tokens and target gratings were all presented supraliminally. Unaware-group participants completed trials in which emotional voice tokens and target gratings were presented supraliminally, but emotional face images were masked using CFS (Tsuchiya & Koch, 2005), rendering them subjectively invisible. Thus, *Experiment 1* employed a 2 × 2 mixed factorial design, with awareness as a between-subjects factor (two levels: aware, unaware) and cue validity as a within-subjects factor (two levels: valid, invalid). Response accuracy was considered the primary dependent measure, consistent with Palmer and Ramsey (2012) and the fact that the task design had been optimised for accuracy. Response times were also recorded and analysed, but these analyses should be considered secondary to the accuracy results.

If congruence cues spatial attention as a result of multisensory integration, *aware* participants should allocate spatial attention to the cued location, and therefore show greater accuracy (and possibly faster response times) for valid trials over invalid trials. This prediction was based on prior evidence of multisensory integration for emotional voices and facial expressions in other experimental paradigms (Collignon et al., 2008; De Gelder & Vroomen, 2000; Massaro & Egan, 1996).

Additionally, if audiovisual integration occurs in the absence of visual awareness, *unaware* participants should also allocate spatial attention to the cued location, and therefore show greater accuracy (and possibly faster response times) for valid trials over invalid trials. This prediction constitutes a replication of the unconscious spatial-cueing effects reported by Palmer and Ramsey (2012).

## Method

### Participants

Sixty students enrolled in a 100-level introductory psychology course at Victoria University of Wellington participated for course credit. All participants reported normal or corrected-to-normal vision. No other demographic information was recorded.

Forty-eight participants were randomly assigned to aware and unaware task conditions (24 per condition). Aware participants completed trials in which all stimuli were presented supraliminally; unaware participants completed trials in which the emotional face images were masked by CFS (Tsuchiya & Koch, 2005). An additional 12 participants were recruited to the unaware group, replacing participants who were excluded. Exclusions included nine participants who experienced excessive breakthrough (as assessed by subjective and objective suppression checks, described below). A further three participants were replaced, having not completed the task, respectively due to nausea, inability to detect target gratings, and an experimenter error that omitted auditory stimuli from the experiment.

I determined the sample size based on a conservative estimate of the effect sizes for the spatial-cueing effect in the first two experiments reported by Palmer and Ramsey (2012), omitting the third experiment that used longer stimulus onset asynchronies. Due to the small samples used in the original study ($N_1$ = 5, $N_2$ = 7, after exclusions), the reported effect sizes most likely overestimated the true effect size (congruent face–voice pairs as cue: $\eta^2_p$ = .921, $\omega^2_p$ = .883, $d_z$ = 3.05; incongruent face–voice pairs as cue: $\eta^2_p$ = .510, $\omega^2_p$ = .396, $d_z$ = 0.94). I estimated the true effect size to be 80% of the weighted average for the two experiments ($d_z$ = 1.596). Using G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007), I calculated the minimum sample size necessary to have 90% power to detect a cueing effect of this magnitude (*N* = 16 participants per condition). Finally, because I also wanted to be able to detect a possible interaction between awareness and cue validity, I further increased the sample size to 24 participants per condition.

This research was approved by the Victoria University of Wellington Human Ethics Committee (application reference #RM24597).

## Stimuli

**Visual mask.** A dynamic colourful mask, generated by rapidly alternating (10 Hz) colour patches to create Mondrian-like images, was used to suppress cue images from conscious awareness.

**Cues.** Novel cue sets were created by pairing emotional face images from an existing, normed image set (NimStim Set of Facial Expressions; Tottenham et al., 2009) with a new set of emotional voice tokens recorded for this experiment. Emotional faces and voices were chosen so that emotional congruence between auditory and visual stimuli could be used as an informative cue. An additional set of images, depicting emotionally neutral scenes and objects, was collated for use during a pre-test calibration task, so that informative cues were not provided before the experimental trials began (described below).

***Voice tokens.*** Sixteen voice tokens of emotional speech (eight *happy*, eight *sad*) were spoken by a female voice actor. Each voice token was a two-syllable spoken McGurk-like nonsense word. Each of the following eight words featured once in each emotional prosody: *gaga*, *gaka, kaga*, *kaka*, *baba*, *bava*, *vaba*, *vava*. Nonsense words were chosen to match those used by Palmer and Ramsey (2012) or by McGurk and MacDonald (1976).

Although only happy and sad voice tokens were used in the experiment, these were piloted amongst a broader set of voice stimuli representing the six basic emotion categories (happiness, sadness, anger, fear, surprise, and disgust; Ekman, 1992) to ensure that they were unambiguous exemplars of happiness and sadness, respectively. A shortlist of suitable voice tokens was compiled for the pilot study (see *Voice Pilot* below), with tokens selected to ensure that the recording was clean, that the total speaking duration was between 700 ms and 1200 ms, and that tokens for each prosody included a range of intensities (e.g., “anger” could range from irritation through to rage).

Voice tokens were recorded in a single session that included eight takes for each nonsense word in both *happy* and *sad* prosodies, and four takes for each nonsense word in each of the remaining four emotional prosodies. Voice acting was provided by a paid, female actor, recruited through the *School of English, Film, Theatre, and Media Studies* at Victoria University of Wellington. Tokens for each emotional prosody were recorded in blocks: *happy* and *sad* recordings were recorded over two blocks each, at the start and end of the recording session, while all other emotional prosodies were recorded in a single block. The voice actor was encouraged to adopt a strategy of method acting for each block, in order to aid authenticity.

*Voice Pilot.* Twenty-four participants (18 male, 5 female, 1 preferred not to say; age: *M* = 34.59, *SD* = 11.13) were recruited via Amazon Mechanical Turk and paid US$1.50 for participating. Enrolment was restricted to individuals registered in the United States of America, with a history of 50 prior "hits" (i.e., completed Mechanical Turk tasks) and an approval rating of at least 95% across all prior hits. Data from one participant was excluded, as their individual accuracy did not exceed chance on a $\chi^2$ test for goodness of fit with the expected emotional categories.

Participants categorised 192 voice tokens, containing four unique recordings for each possible prosody–word pair (including the eight nonsense words selected for *Experiment 1* and four additional fillers). Voice tokens were selected from the full set of recordings, so that each recording was clean, included between 700 ms and 1200 ms of speech, and expressed a range of intensities for each emotional prosody. The pilot task was presented in an online survey, hosted by TESTABLE (https://www.testable.org/). On each trial, participants listened to a single emotional nonsense word and categorised it as belonging to one of the six basic emotions or as unidentifiable ("no good match"). Participants heard each token exactly once, presented in random order, and were instructed to respond by keypress using their right hand and the 0–6 keys on their numeric keypad (1 = anger, 2 =disgust, 3 = fear, 4 = happy, 5 = sad, 6 = surprise, 0 = no good match). A response legend remained onscreen for each trial, to remind participants of the categories associated with each key.

Participants selected the expected emotional category for a majority of voice tokens (*M* = .703, *SD* = .129), and rarely indicated that the voice did not fit any of the emotional categories (i.e., No good match response; *M* = .034, *SD* = .030). A repeated-measures analysis of variance (ANOVA) revealed a main effect of emotion, $F(5, 22) = 18.51$, $p < .001$, $\eta^2_p = .46$ (Figure 2). Pairwise comparisons indicated that the main of effect of emotion reflected greater response accuracy for happy, sad and surprised voice tokens, relative to all other emotional categories (all *p*s < .003). Critically, accuracy for happy and sad voice tokens did not differ from one another, $t(14) = .920$, $p > .999$, $d = .192$, suggesting that happy and sad voice prosodies were comparably recognisable. Moreover, participants made few errors in which happy voice tokens were categorised as sad (14 errors total, across 24 participants) and vice versa (also 14 errors; see Table 1).

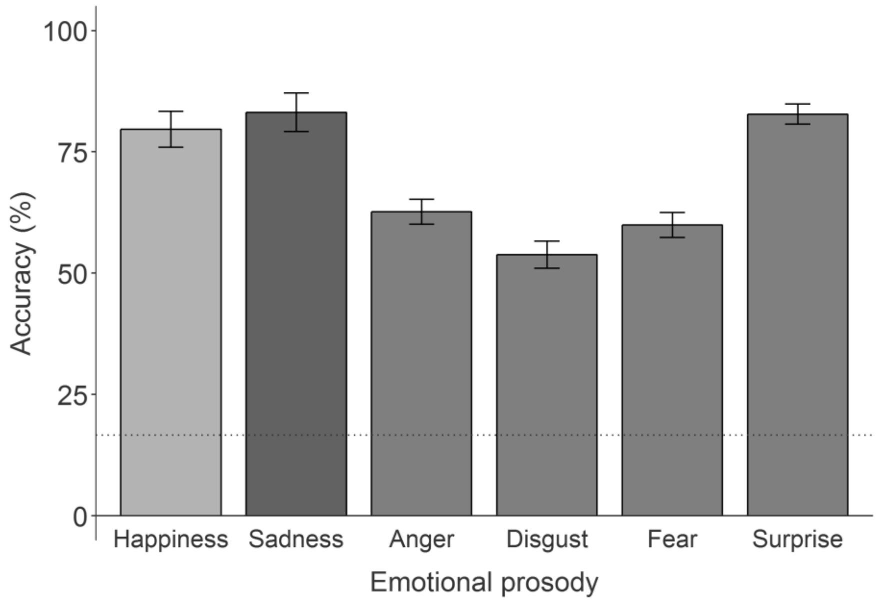


*Figure 2*. Categorisation accuracy for emotional prosody (*Voice Pilot*). No significant difference was observed in the critical comparison of *happy* and *sad* voice tokens. Dotted line represents chance accuracy.

Table 1

*Confusion Matrix for Emotional Categorisation of Emotional Voice Tokens (Pilot Study)*

| | Correct emotion | | | | | |
|---|---|---|---|---|---|---|
| Response | Happiness | Sadness | Anger | Disgust | Fear | Surprise |
| Happiness | *586* | **14** | 52 | 51 | 29 | 52 |
| Sadness | **14** | *612* | 22 | 42 | 174 | 9 |
| Anger | 26 | 10 | *461* | 134 | 15 | 24 |
| Disgust | 19 | 9 | 150 | *396* | 13 | 18 |
| Fear | 15 | 72 | 10 | 20 | *441* | 13 |
| Surprise | 42 | 11 | 23 | 26 | 54 | *609* |
| No good match | 34 | 8 | 18 | 67 | 10 | 11 |
| Total | 736 | 736 | 736 | 736 | 736 | 736 |

*Note*. Correct responses are marked with italics and highlighted in light grey. Confusions of *happy* voice tokens for expressions of sadness and *sad* voice tokens for expressions of happiness are highlighted in boldface.

From the piloted voice tokens, I selected one *happy* and one *sad* voice token for each nonsense word for use in the spatial-attention experiment (*Experiment 1*). Selection produced a set of 16 voice tokens (8 *happy*, 8 *sad*). Tokens were selected such that *happy* and *sad* recordings were equally recognisable, according to the categorisation accuracy data obtained during piloting (happy: $M$ = .847, $SD$ = .020; sad: $M$ = .837, $SD$ = .030).

***Images.*** Cue images included 16 images of emotional facial expressions (used for trials in the experimental and awareness control blocks), and 10 images of emotionally neutral scenes (used for trials in the calibration block).

*Emotional face images.* Sixteen photographs of *happy* and *sad* facial expressions displayed by eight individuals were selected from the *NimStim Set of Facial Expressions* (Tottenham et al., 2009). The *NimStim* database includes emotion-identification accuracy scores for each face, allowing for the selection of highly recognisable emotional facial expressions. Female faces were chosen, to match the gender of the speaker in the voice tokens. For each female face identity in the *NimStim* database, I selected the most accurately identified closed-mouth *happy* and *sad* facial expressions, producing a shortlist of 18 image pairs. For each pair, I calculated a mean accuracy score from the normative data. Pairs were ranked by mean accuracy, and the eight image pairs with the highest scores were selected for use as emotional face cues. A comparison of the chosen *happy* ($M$ = .963, $SD$ = .020) and *sad* ($M$ = .960, $SD$ = .014) emotional face images, based on the *NimStim* data, revealed no difference in accuracy for emotion identification, $t(7) = .251$, $p > .999$, $d = .192$.

Scrambled versions of all emotional face images were created for use in the post-test awareness control task. Scrambles were generated in PhotoScape v.3.7 by splitting each image into 1656 tiles (14 × 14 pixels) and reassembling them in a random order.

*Neutral scene images.* To avoid introducing informative emotional face–voice cues during calibration, I prepared an alternative set of 10 emotionally neutral scene images. Images were chosen from the *International Affective Picture System* (IAPS) database (Lang, Bradley, & Cuthbert, 2008) to include low-arousal, neutral-valence images of recognisable objects. The selected images were image numbers 5800, 7000, 7006, 7010, 7020, 7031, 7110, 7175, 7217, and 7950.

*Image processing.* All images were equated for low-level visual features using the SHINE toolbox (Willenbockel et al., 2010) for MATLAB. Images were converted to greyscale, equated for contrast and luminance, and restored to full colour at 10% contrast. Low-contrast images are typically used for CFS paradigms in order to minimise breakthrough, as full-contrast images are more likely to compete for awareness (Carmel et al., 2010). Low-contrast face images, and examples of low-contrast scrambles, are presented in Figure 3; examples of IAPS images are not presented, as is standard practice for images sourced from that database.

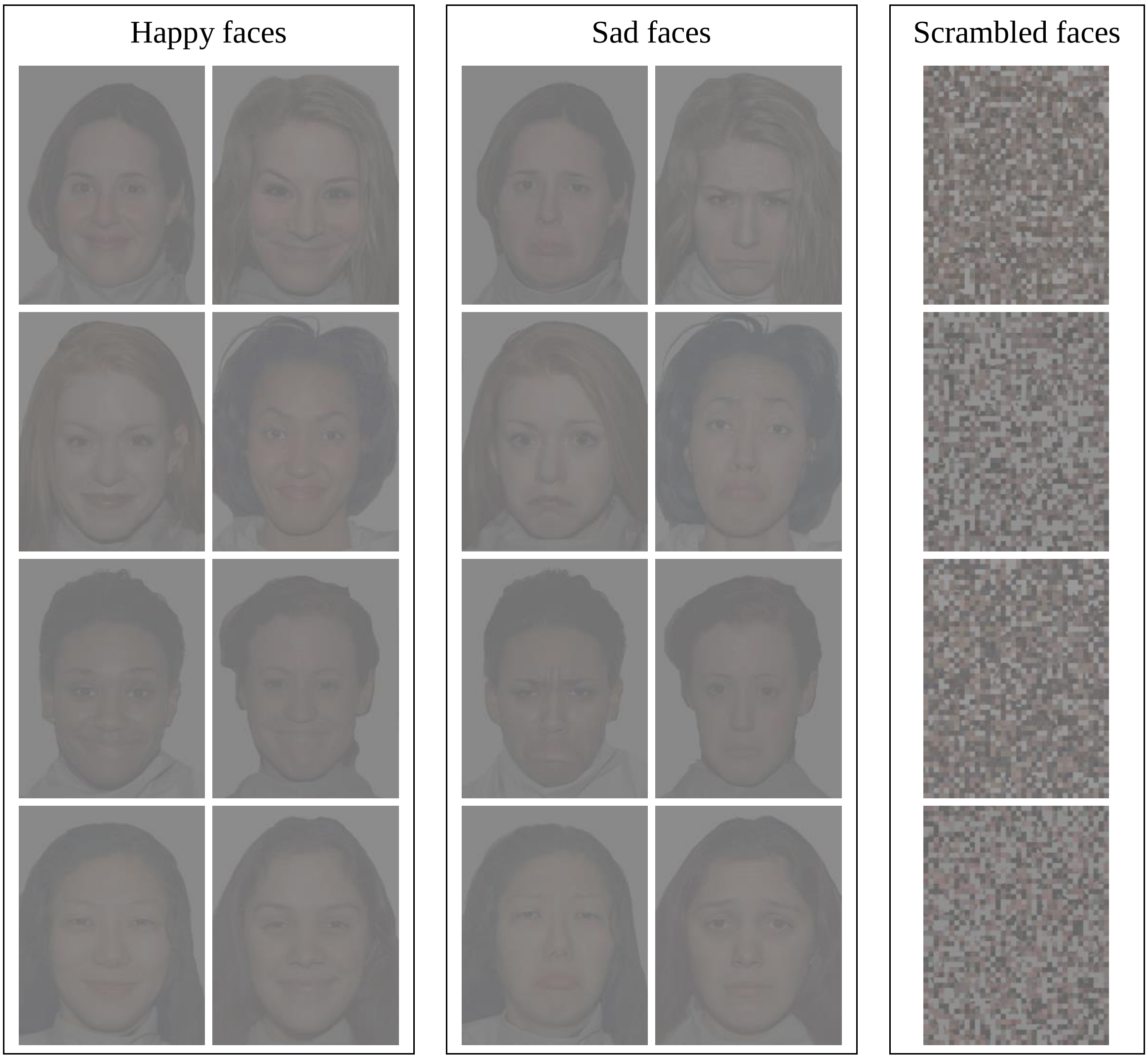


*Figure 3*. Processed cue images of happy and sad facial expressions, and selected examples of scrambled face images.

**Targets .** Ten target Gabor gratings (image dimensions: 156 × 200 pixels, grating dimensions: ~ 112 × 112 pixels; contrast: 30%; frequency: 0.05 cycles per pixel) were created in the Psychophysics Toolbox (Brainard, 1997) for MATLAB 2016a. All gratings appeared on a background of 50% grey (RGB: 128, 128, 128). A single 50% grey rectangle, sans grating, was also created as a foil that would appear opposite the target grating on each trial, to prevent exogenous attention capture by the onset of a singleton target stimulus.

To calibrate task difficulty, targets were generated for five orientations (i.e., grating angles 1, 2, 3, 4, and 5 degrees from vertical) and two directions of rotation (leftward or rightward). As the grating orientation approaches vertical, target discrimination becomes more difficult. Target gratings are presented in Figure 4.

***Pilot studies.*** Target gratings needed to be suitably difficult for participants to discriminate, so that I could detect a spatial-cueing effect (should it exist) and maintain participants' response accuracy midway between floor (25%) and ceiling (100%). To find an appropriate target set for the congruence-cued spatial attention task, I trialled a range of target contrasts and orientations over two pilot experiments. In these studies, I used a simple visual spatial cueing task (Posner, 1980) and systematically varied either target contrast (*Target Pilot 1*; cf. Palmer & Ramsey, 2012) or target orientation (*Target Pilot 2*). Thus, I was able to gradually increase the task difficulty to the point where target-discrimination accuracy fell midway between floor and ceiling (i.e., ~65%) and a significant spatial-cueing effect was observed for both accuracy and response times.

Pilot sessions were conducted in groups of 3–4 participants, in a darkened testing room. Each participant was assigned a private booth, with 24.1-inch computer monitor, standard keyboard, and adjustable chinrest that fixed viewing distance at 58 cm. Instructions and stimuli were presented on a black background that occupied the full screen. Both pilot experiments were programmed and displayed using PsychoPy v1.85.1 (Peirce, 2007, 2009).

Trials for both pilot experiments followed the same structure. On each trial, participants fixated on a central cross that displayed for 800 ms and subtended .79 × .79 degrees of visual angle. An arrow cue, composed of three greater-than or less-than symbols (i.e., < < < or > > >), briefly replaced the fixation cross for 500 ms. Arrows indicated the most likely position of the upcoming target. Targets were validly cued (i.e., appeared in the indicated location) for 80% of trials, and were invalidly cued for the remaining 20% of

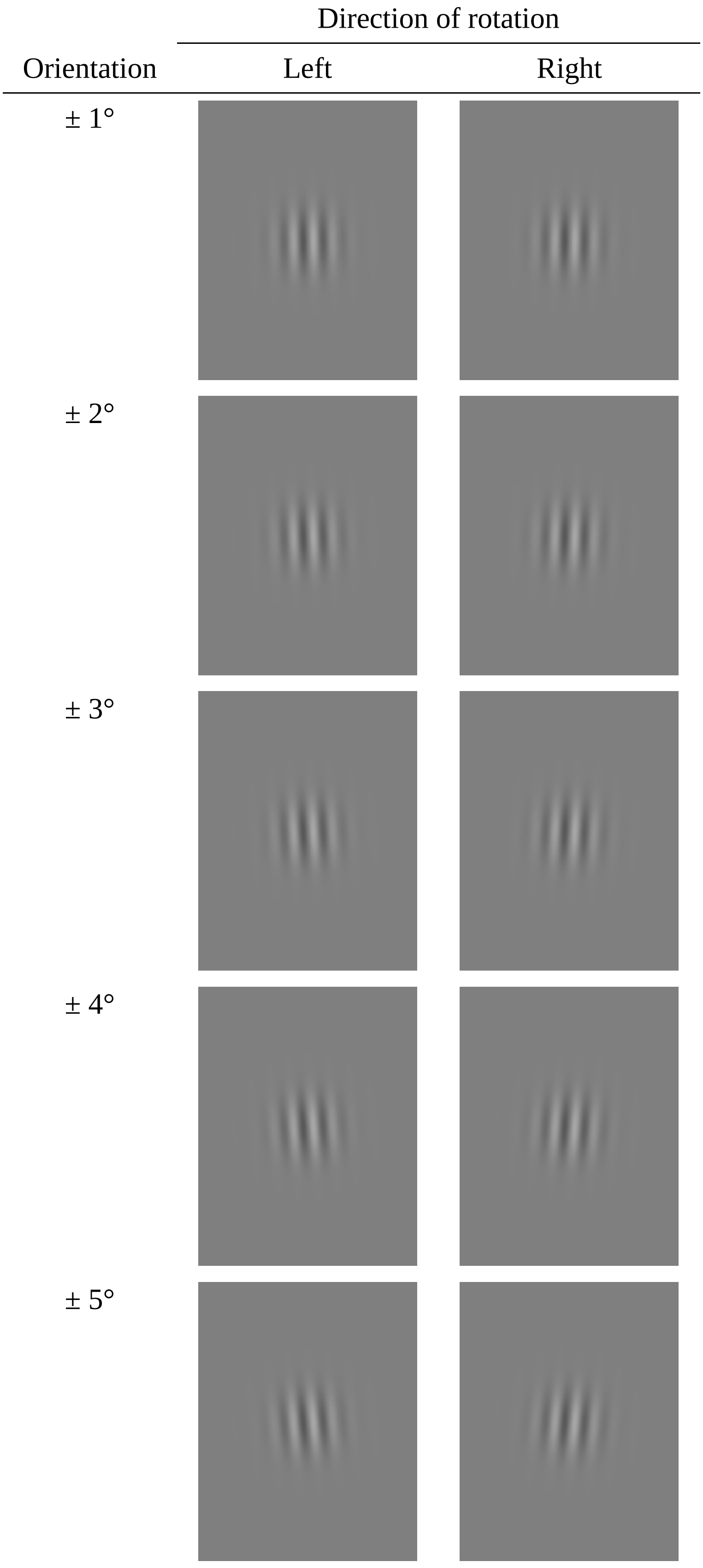


*Figure 4.* Target Gabor gratings for *Experiment 1*.

trials. The unusually long cueing time was intended to allow time for participants to voluntarily orient in the direction of the cue (Posner, 1980), and to approximate the longer SOAs in the main experiment. Target gratings were displayed with an SOA of 610 ms—following an interstimulus interval (ISI) of 110 ms—and remained onscreen for 70 ms. Gratings appeared 6.22 degrees of visual angle to the left or right of fixation (measured from the centre of the fixation cross to the centre of the target). Target images (i.e, Gabor grating and grey background) subtended 3.06 × 3.06 degrees of visual angle (properties of the target gratings themselves are described separately for each pilot experiment below). A matching grey square, without a grating, displayed opposite the target on all trials to prevent exogenous orienting to the target location.

Participants indicated the location and orientation of the target grating by single keypress. Response keys were mapped so that the key positions corresponded to the location and orientation of the target (Figure 5). For targets on the left, participants responded with the left hand—the *Z* key indicated a leftward orientation, the *X* key a rightward orientation. For targets on the right, participants responded with the right hand, pressing *N* for leftward-oriented targets and *M* for rightward-oriented targets. Response windows were open-ended, and trials ended when a response was received. Participants were encouraged to respond quickly and accurately to each target.

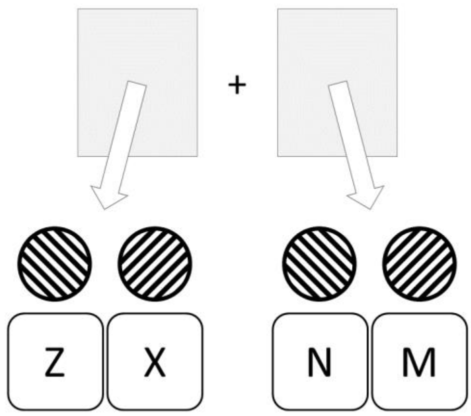


*Figure 5.* Response key mapping for *Target Pilots 1* and *2*. Participants used the left hand to indicate a target on the left of fixation, and the right hand to indicate a target on the right. The first two fingers of each hand rested on the *Z* and *X*, and *N* and *M* keys respectively, and were used to report the direction in which the target was oriented.

Participants received written instructions explaining the task, and completed a short practice block to familiarise themselves with the response keys. The practice block included three stages, in which the task was introduced gradually. First, four trials were presented with no arrow cues and target gratings that remained onscreen until a response was received, allowing participants to practice finding the correct response key. Second, ten trials were presented with arrow cues and 80% cueing validity, allowing participants to both practice responding and to become familiar with the cueing procedure. Finally, a further ten trials were presented, in which target gratings displayed for only 70 ms (thus matching the experimental trial procedure), allowing participants to become familiar with the timing of experimental trials.

I recorded participants' target-discrimination accuracy and response time (from target onset) for each trial, and calculated mean scores of each measure for each cueing condition (i.e., valid or invalid trials). Mean response times were calculated for correct responses only, and excluded outliers (responses more than two standard deviations slower than a participant's mean response time) and anticipations (responses faster than 200 ms).

Participants' mean accuracy and response times were submitted to separate two-way repeated measures ANOVAs. *Target Pilot 1* included two within-subjects variables: cue validity (valid, invalid) and target contrast (five levels of low- through high-contrast target gratings). *Target Pilot 2* also included two within-subjects variables, this time cue validity (valid, invalid) and target orientation (five levels of targets oriented ± 1–5° from vertical). Greenhouse-Geisser corrections were applied, where applicable, and post-hoc analyses were corrected for multiple comparisons using Bonferroni corrections where appropriate. Inferential statistics were calculated using JASP and R.

*Target Pilot 1*. The first pilot presented target gratings with a constant orientation ± 45° to the right or left of vertical, thus matching the orientation of the target gratings used by Palmer and Ramsey (2012). Gratings varied in amplitude, with contrast values from 30.2% to 100% (Figure 6).

Fifteen students (13 female, 2 male; $M_{age}$ = 19.13, $SD_{age}$ = 1.51), enrolled in a 100-level Psychology paper, participated for course credit. Participants completed five blocks of 100 trials each. Each block contained 80 valid trials and 20 invalid trials, and comprised a complete counterbalance for all conditions.

Mean target-discrimination accuracy (Figure 7) was subjected to a 2 × 2 repeated measures ANOVA with cue validity and target contrast as within-subjects variables. A main effect of cue validity, $F(1, 14) = 12.46$, $p = .003$, $\eta^2_p = .471$, indicated that participants' target-discrimination accuracy was greater for valid trials ($M = .887$, $SD = .094$) than for invalid trials ($M = .859$, $SD = .111$). This valid-trials advantage suggests that participants allocated spatial attention to the cued location, and represents a

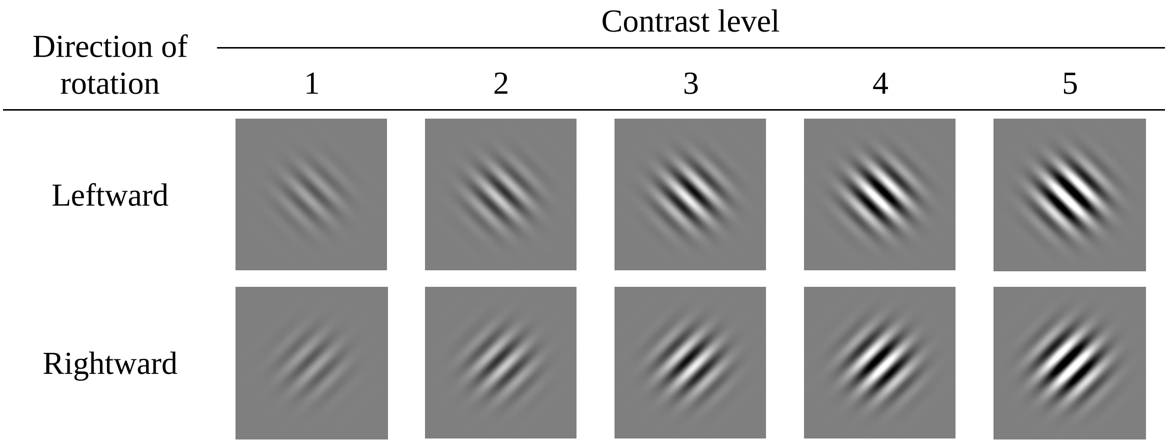


*Figure 6*. Target gratings for *Target Pilot Study 1*.

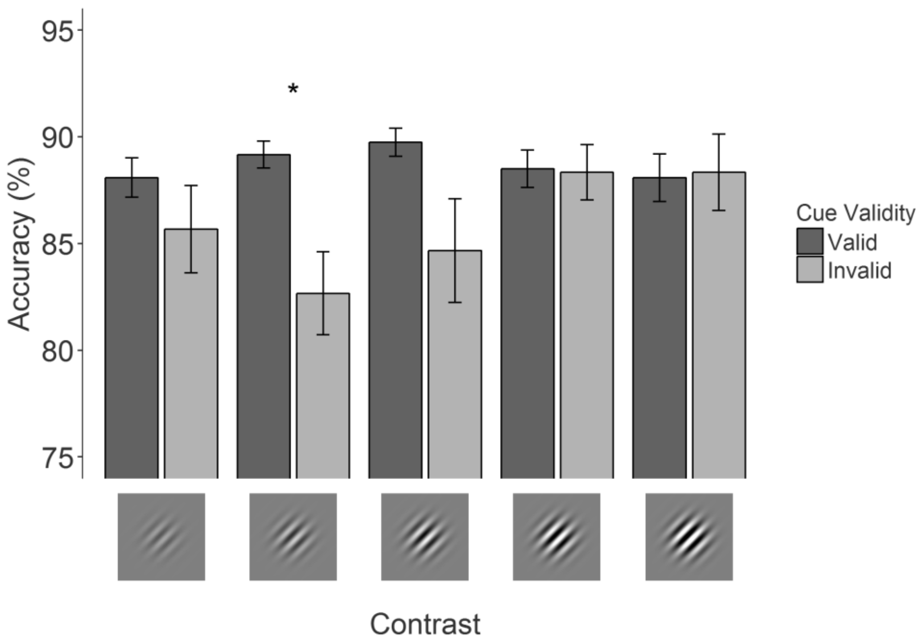


*Figure 7*. Mean target-discrimination accuracy for valid and invalid trials in *Target Pilot 1*. Error bars represent within-subjects standard errors (Morey, 2008).

significant spatial-cueing effect. No other effects reached significance, suggesting that response accuracy was unaffected by target contrast—perhaps because the task remained too easy, as indicated by the high accuracy scores obtained for contrast conditions.
A separate 2 × 2 repeated measures ANOVA, with cue validity and target contrast as within-subjects variables, was conducted for response times (Figure 8). Main effects of cue validity, $F(1, 14) = 14.478$, $p = .002$, $\eta^2_p = .508$, and target contrast, $F(4, 56) = 4.785$, $p = .002$, $\eta^2_p = .255$, were qualified by a significant interaction, $F(4, 56) = 2.581$, $p = .047$, $\eta^2_p = .156$. Participants responded faster to target gratings following a valid cue ($M$ = 632 ms, $SD$ = 108 ms) than an invalid cue ($M$ = 701 ms, $SD$ = 102 ms). Post-hoc comparisons indicated that this speed advantage for valid trials was significant at all levels of target contrast (all *p*s < .05), though the significant interaction term suggests that the response-time cueing effect was largest for higher contrast targets. Additionally, participants responded faster for the three highest contrast levels ($M_3$ = 603 ms, $SD_3$ = 118 ms, $M_4$ = 601 ms, $SD_4$ = 116 ms, $M_5$ = 599 ms, $SD_5$ = 112 ms) than for the lowest ($M_1$ = 628 ms, $SD_1$ = 118 ms), suggesting that lower contrast targets were somewhat more difficult to discriminate even though accuracy was not significantly reduced. All told, response time data also suggested that participants allocated attention towards the cued location.

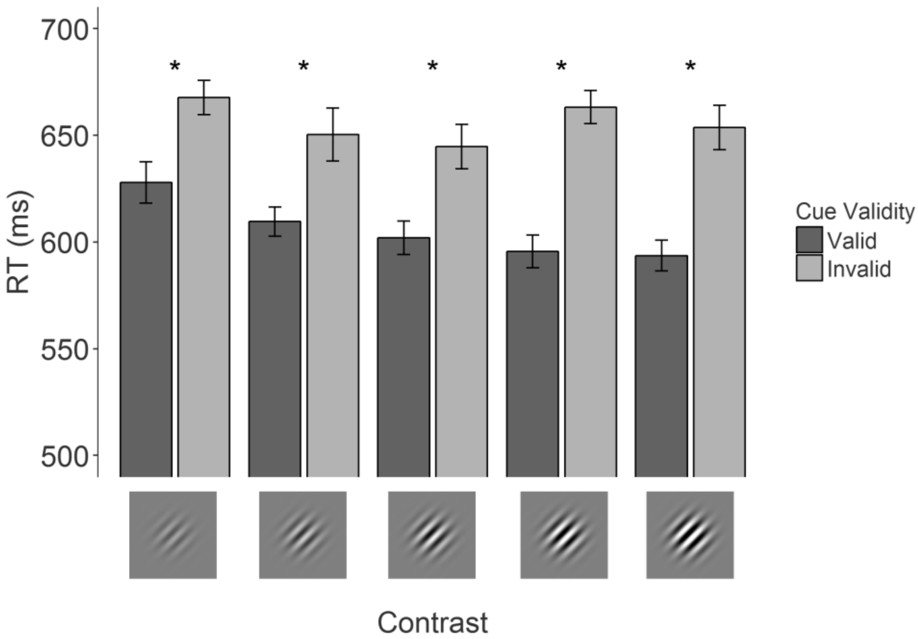


*Figure 8*. Mean response times for valid and invalid trials in *Target Pilot 1*. Error bars represent within-subjects standard errors (Morey, 2008).

*Target Pilot 2.* Because participant accuracy in *Target Pilot 1* remained higher than intended, even for the lowest contrast targets, I conducted a second pilot study to further increase the difficulty of the task. In it, I presented target gratings with a constant contrast (30.2%), retained from the lowest contrast level in the previous pilot. Targets were presented at five levels of orientation: rotated 1°, 3°, 5°, 7° and 9° to either side of vertical (Figure 9).

Sixteen students (11 female, 5 male; $M_{age}$ = 18.56, $SD_{age}$ = 1.31) enrolled in a 100-level Psychology paper, participated for course credit. Participants completed five blocks of 100 trials each. Each block contained 80 valid trials and 20 invalid trials, and comprised a complete counterbalance for all conditions. As in the previous pilot, mean target-discrimination accuracy and response times were subjected to separate 2 × 2 repeated measures ANOVAs, this time with cue validity (valid, invalid) and target orientation (1°, 3°, 5° 7°, 9°) as within-subjects variables.

The ANOVA for response accuracy (Figure 10) revealed main effects of cue validity, $F(1, 15) = 10.394$, $p = .006$, $\eta^2_p = .409$, and orientation, $F(2.569, 38.539) = 116.068$, $p < .001$, $\eta^2_p = .886$. The interaction was not significant. Participants responded more accurately following a valid cue ($M$ = .850, $SD$ = .044) than an invalid cue ($M$ = .801, $SD$ = .070). This accuracy advantage suggests that spatial cueing remained effective at all levels of target orientation. Post-hoc comparisons revealed that accuracy for 1° targets ($M$ = .635, $SD$ = .054) was significantly lower than for all other targets (all $p$s < .001). Additionally, accuracy for 3° targets ($M$ = .818, $SD$ = .055) was significantly reduced (all $p$s < .001) compared to all remaining targets (all $p$s < .001); and

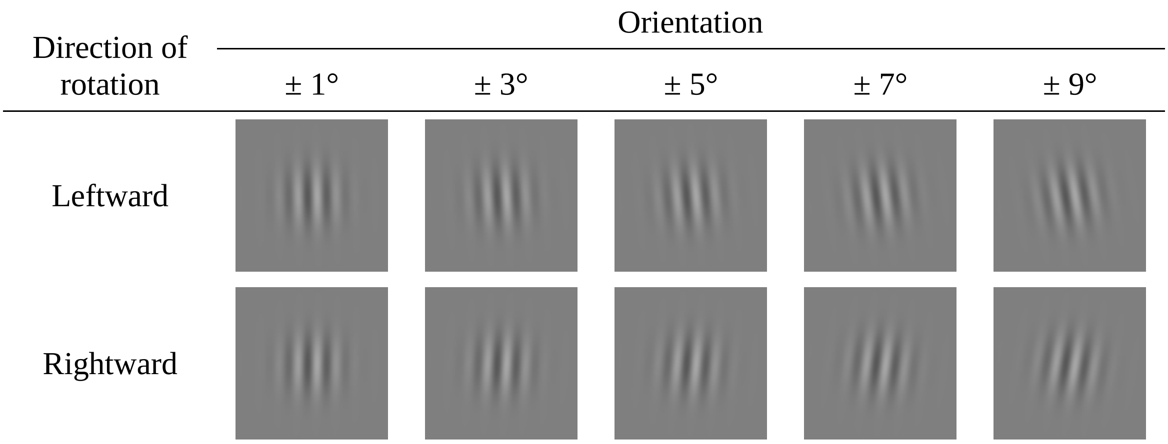


*Figure 9.* Target gratings for *Target Pilot 2.*

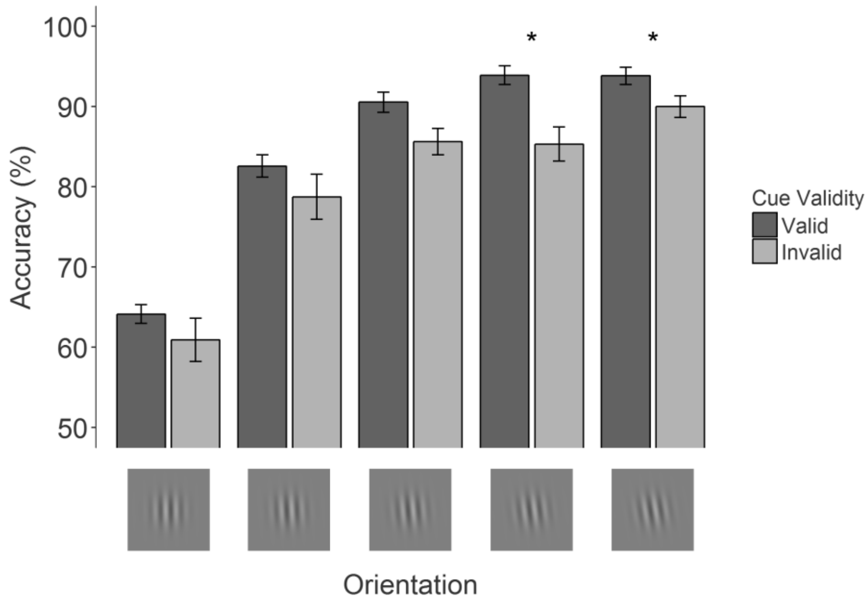


*Figure 10*. Mean target-discrimination accuracy for valid and invalid trials in *Target Pilot 2*. Error bars represent within-subjects standard errors (Morey, 2008).

accuracy for 5° targets (*M* = .896, *SD* = .053) was lower than for 9° targets (*M* = .931, *SD* = .051; *p* = .015), while neither differed from 7° targets (*M* = .922, *SD* = .052). Thus, target-discrimination accuracy was lowest for target gratings oriented ± 1° from vertical. Importantly, accuracy for 1°-orientation targets approached the desired mean (~65%).

The ANOVA for mean response times (Figure 11) revealed significant main effects of cue validity, $F(1, 15) = 7.957$, $p = .013$, $\eta^2_p = .347$, and target orientation, $F(1.272, 19.073) = 11.425$, $p = .002$, $\eta^2_p = .432$. The interaction was not significant. Participants responded faster to validly cued targets (*M* = 666 ms, *SD* = 132 ms) than invalidly cued targets (*M* = 716 ms, *SD* = 160 ms), suggesting that spatial cueing occurred for all target orientations. Pairwise comparisons for target orientation showed that participants took longer to respond to 1° target gratings (*M* = 739 ms, *SD* = 164 ms) than to all other targets (all *p*s < .025). Additionally, responses to 3° targets (*M* = 698 ms, *SD* = 138 ms) were slower than responses to 7° targets (*M* = 652 ms, *SD* = 134 ms; *p* =.002) or 9° targets (*M* = 644 ms, *SD* = 133 ms, *p* < .001); and responses to 5° targets (*M* = 672 ms, *SD* = 130 ms) were slower than responses to 9° targets (*p* < .001). Thus, response time data also suggested that target discrimination was more difficult for targets with smaller orientation angles.

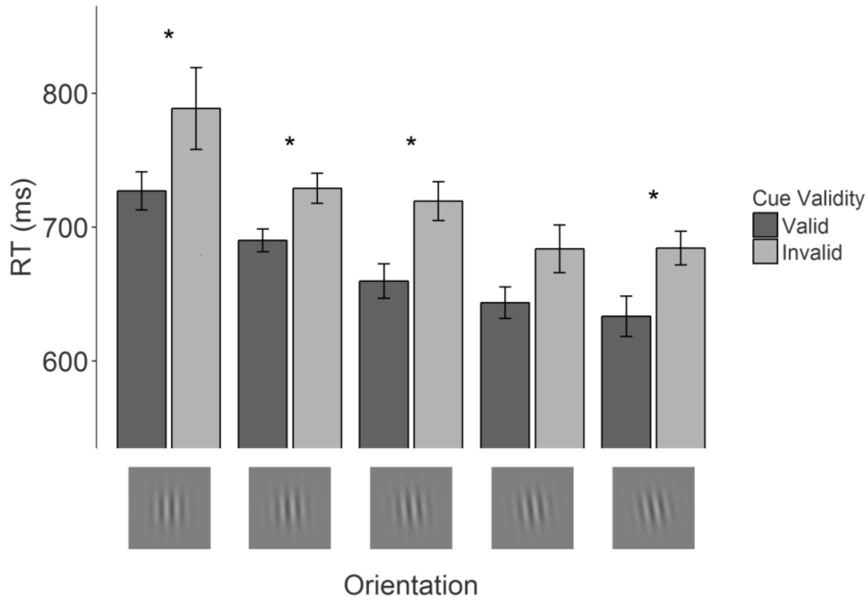


*Figure 11*. Mean response times for valid and invalid trials in *Target Pilot 2*. Error bars represent within-subjects standard errors (Morey, 2008).

Critically, participants' mean accuracy dropped to the desired level of ~65% when targets had an orientation ± 1° from vertical and 30.2% contrast. These figures provided a rough estimate of the mean accuracy I could expect for participants in the congruence-cued spatial attention task (*Experiment 1*), based on the difficulty of discriminating the chosen targets. Of course, differences in the task conditions between the visual pilot studies and the congruence-cued spatial attention task might allow mean accuracy to rise or fall in the crossmodal task. Additionally, individual participants' visual sensitivity—and thus their ability to discriminate each target set—was still likely to vary between participants.

For the spatial attention experiment (*Experiment 1*), I selected ± 1° target orientations as the highest-difficulty targets to be included in the calibration task, and ± 5° targets as the lowest-difficulty targets. I expected that this range would account for any differences in task difficulty between the pilot and main experiment paradigms, and for individual difference in visual sensitivity.

**Procedure**

*Experiment 1* participants sat in a dimly lit room. Visual stimuli were displayed on a 23-inch Alienware AW2310 computer monitor (16:9 aspect ratio, 120 Hz refresh rate).

Left and right visual fields were separated by a mirror stereoscope, mounted at a 63 cm viewing distance from the screen. A textured border was displayed on both visual fields throughout the experiment to facilitate stable binocular convergence. All visual stimuli appeared on a background of approximately 39% grey (RGB: 100, 100, 100). Auditory stimuli were presented through Manhattan noise-cancelling headphones (model number 176163). All stimuli were presented, and responses recorded, in MATLAB 2016a. Participant responses were made by button press using a Cedrus RB-730 response box, positioned immediately under the mount for the stereoscope and directly in front of the participant. Participants were provided verbal and written instructions at the beginning of the experimental session, and a visual demonstration of the response keys to use for each combination of target location and orientation (Figure 12).

Each experimental session progressed through three or four phases, depending on awareness condition. In order, aware participants completed a practice phase (1 block, 12 trials), a calibration phase (1 block, 100 trials), and an experimental phase (4 blocks, 80 trials per block). Unaware participants completed the same three stages, plus a post-test control phase (1 block, 64 trials). Trials in each phase followed the same structure (Figure 13), consisting of a cueing, target and response stage (in that order). This basic structure

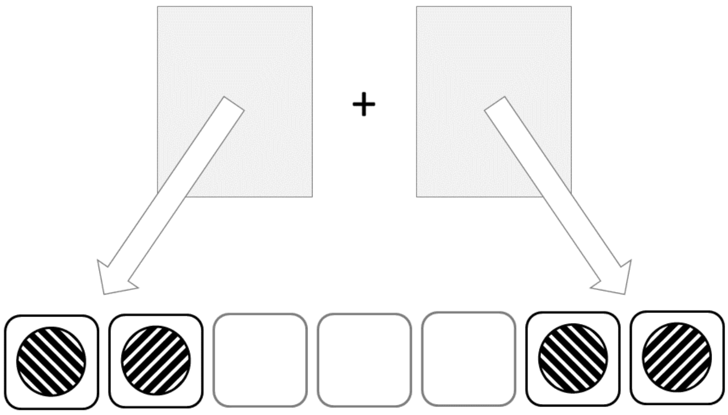


*Figure 12*. Response key mapping for target discrimination in *Experiments 1* and *2*. Participants responded by button press on the outer two keys at the left and right of the response box. Participants used the left hand to indicate a target on the left of fixation, and the right hand to indicate targets on the right. The first two fingers of each hand were used to report target orientation.

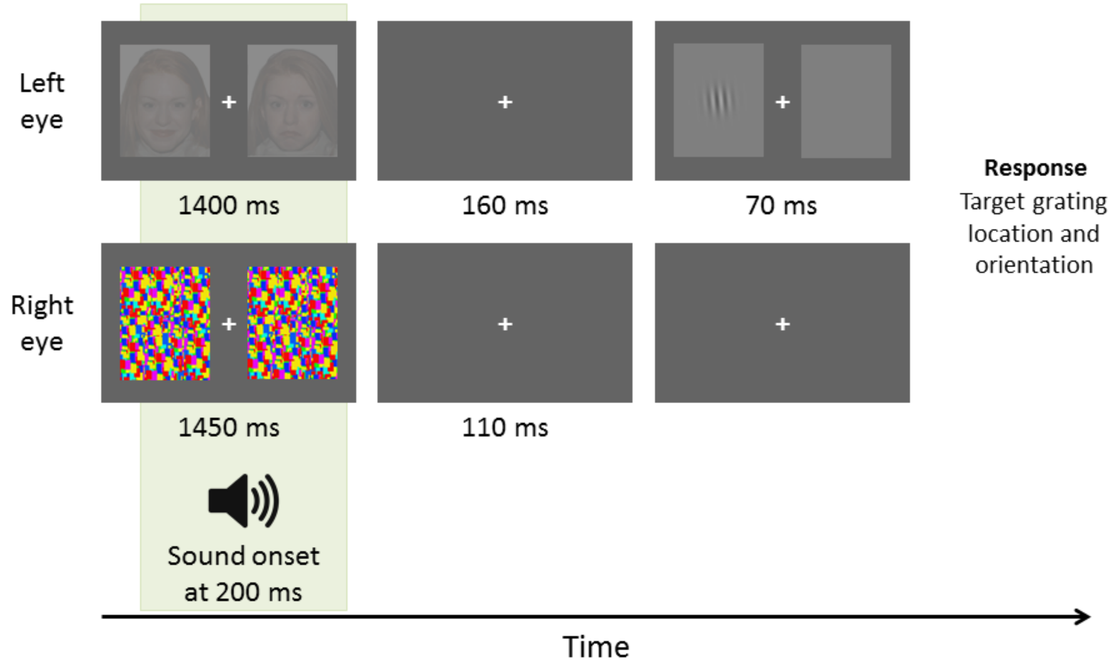


*Figure 13*. Trial diagram for *Experiment 1*. The left eye monocularly viewed cue images and targets (both awareness conditions), while the right eye viewed the CFS mask (unaware condition only, otherwise blank). Cue images variously presented intact emotional faces, scrambled emotional faces, and neutral objects (see description of experimental phases in main text). Presentation to each eye was randomised across trials.

is described in the following paragraphs, and specific features for each experimental phase are provided in the following sections.

During cueing, two images (e.g., emotional faces) were displayed monocularly, appearing either side of a central fixation cross (0.2° visual angle, appearing to both eyes). Images subtended 3.2° × 4.1° of visual angle, and were offset 4.4° from the fixation cross (centre to centre). Presentation eye (left or right) was determined at random, with the constraint that equal trial numbers were presented to each eye within each block. Images displayed for 1400 ms, including 200 ms of linear contrast ramping from zero (0%) to full (10%) contrast intended to facilitate effective suppression. For participants in the unaware condition, two dynamic CFS masks displayed concurrently to the opposite eye for 1450 ms (persisting 50 ms after the cue images offset to help prevent breakthrough). Masks occupied the same area of visual space as the cue images, thus suppressing the images from conscious awareness.

Cue images were accompanied by concurrent emotional voice cues. Emotional voice tokens onset 200 ms after cue images, so that voice onset coincided with the images reaching full contrast. Voice duration varied between 700 and 1200 ms, so as to offset no later than the offset of the cue images. In experimental- and control-block trials, emotional voice tokens were paired with emotional face images, such that the emotional congruence between faces and voices cued the most likely location of upcoming targets. This cueing contingency is described in full in the *Experimental trials* section below.

The target stage followed a short ISI that separated the cue images and target gratings by 160 ms (both conditions), and separated the CFS mask and target gratings by 110 ms (unaware condition only). On each trial, a single grating appeared either to the left or right of fixation, in the same location as one of the preceding scrambles. Gratings were oriented to the left or right of vertical and subtended ~ 1.5° of visual angle. All gratings displayed on a rectangular background of 50% grey, subtending 3.2° × 4.1° of visual angle. A matching 50% grey rectangle (*foil*) displayed in the opposite location to prevent exogenous orienting to the appearance of a single target. Gratings and foils were displayed monocularly, to the same eye as the earlier cue images for a duration of 70 ms.

In the response stage, participants indicated the location (left or right) and orientation (leftward or rightward) of the target grating by a single button press (as illustrated in Figure 12). For targets that appeared on the right, participants responded with the right hand, using the index finger to indicate leftward and middle finger to indicate rightward orientations. For targets on the left, the left index finger indicated a rightward orientation, and left middle finger indicated a leftward orientation. A response window opened at target onset (0 ms response time) and remained open until a response was received. Participants' responses and response times were recorded for each trial.

Participants in the unaware condition were also required to indicate their subjective awareness of the masked images on each trial. Following the receipt of a target response, a written awareness probe was presented binocularly. Unaware participants had been told that other images (i.e., the masked images) *may* appear during a trial, but did not know that these images were actually present on every trial. To maintain this impression, the question was phrased to ask whether "other images" had been shown during the current trial, without referencing either suppression or breakthrough. Participants rated breakthrough on a scale of 1–4, by completing the stem sentence "On this trial, before the target, I…" with one of four responses, mapped to the same response keys as the target-discrimination task. Responses included (1) "definitely only saw colours", (2) "may have only seen colours", (3) "may have seen other images", and (4) "definitely saw other images". Responses of 1 or 2 were scored as *no-breakthrough* trials; responses of 3 or 4 were scored as *breakthrough* trials. A second awareness probe was included in control-block trials only and is described in the *Control trials* section below.

**Practice trials.** The practice block was included to familiarise participants with the trial procedure and response keys. Practice trials presented uninformative cues, comprising two scrambled face images and no emotional voice tokens, so that participants would not learn the cueing contingency ahead of the experimental phase of the session. The practice block presented 12 trials, including three repetitions for each combination of target location and orientation (2 locations × 2 orientation × 3 repetitions = 12 trials). During practice trials, participants received written feedback on their accuracy. After a response was made, the words "CORRECT" or "INCORRECT" briefly appeared in place of the fixation cross.

**Calibration trials.** The calibration block was included to assess each participants' visual sensitivity for a range of target gratings (oriented 1–5° either side of vertical) and identify the target orientation that would maintain accuracy above chance and below ceiling. To assess visual sensitivity, the calibration block included 20 trials for each level of target orientation (100 trials total), for which target-discrimination accuracy was calculated separately. Accuracy for each orientation level was compared to an a priori standard for desired accuracy (65%), representing accuracy midway between floor and ceiling. Whichever target set showed accuracy closest to 65% (either above or below) was selected for use in later phases of the experiment for that participant. For example, a participant whose accuracy for targets oriented 1–5° to the left or right of vertical was 40%, 45%, 60%, 75% and 90% respectively, would be assigned targets with a ± 3° orientation for all experimental and control block trials because their accuracy for those targets was closest to 65%. In the case of a tie between two or more target orientations, the smallest angle rotation from vertical was preferred.

Calibration trials were designed to approximate the viewing conditions of later experimental trials (i.e., including audiovisual cueing and intact images) while still omitting the cueing contingency. To this end, targets were calibrated without informative cues. Instead, in the cueing stage of each trial, two images of neutral objects were displayed either side of fixation, paired with a single emotional voice token. Because neutral images did not indicate emotional valence, neither image was congruent with the emotional voice token. Thus, for calibration trials, the audiovisual "cue" stimuli, did not provide a cue to the target location.

Calibration trials were randomised, so that trials for each level of target orientation were intermixed throughout the calibration block. For each target orientation, trials included five repetitions of each combination of target location and orientation (2 locations × 2 orientations × 5 repetitions = 20 trials).

**Experimental trials.** The experimental phase comprised 320 trials, presented in four blocks of 80 trials each. In experimental trials, cues comprised two emotional face images and one emotional voice token, thus introducing cueing information for the first time. On each trial, the emotional prosody of the voice token matched the emotion expressed by one of the two face images, thus defining a congruent relationship with one image and an incongruent relationship with the other. This emotional congruence predicted the most likely location of the upcoming target grating (but not its orientation). Specifically, the target grating appeared in the same position as the emotionally congruent face on 80% trials in each block. So, for those trials in which the emotional voice expressed happiness, the target would be most likely to appear in the location of the happy face (*valid* cue; 80% of trials) and less likely to appear in the location of the sad face (*invalid* cue; 20% of trials).

Experimental trials were counterbalanced across sets of eight participants in each awareness condition. The full counterbalance includes all possible combinations of cue and target stimuli: 8 face identities × 2 image orders (happy–sad or sad–happy) × 2 voice prosodies (happy or sad) × 8 nonsense words × 2 target locations × 2 target orientations = 1024 unique trials. Cue validity is not included in this calculation, because the validity of the cue is already defined by the conjunction of these other variables. Thus, half of these trials are validly cued ($N_{valid}$ = 512) and half are invalidly cued ($N_{invalid}$ = 512). Therefore, in order to achieve 80% cue validity, each valid trial must be repeated four times (4 × 512 unique valid trials = 2048 valid trials). A full counterbalance therefore comprises 2560 trials (2048 valid trials + 512 invalid trials). This full counterbalance was divided into exactly eight non-overlapping lists of 320 trials each. Each participant completed experimental trials from one of these lists. In each awareness condition, three participants were assigned each of the possible lists (3 participants × 8 lists = 24 participants per condition).

Each list maintained a strict 4:1 ratio of valid and invalid trials (i.e., 80% cue validity), and remained balanced with respect to face identities, configurations, voice prosodies, target locations and target orientations. Thus, each participants' list was internally balanced with respect to all variables. For each participant, trials were randomly assigned to four 80-trial experimental blocks, with the constraint that each block contained 64 valid trials and 16 invalid trials.

**Control trials.** Unaware participants completed a final block of 64 control trials that included an objective awareness probe. Cues comprised two emotional face images and one emotional voice token, however on 50% of control trials both emotional faces were replaced by their scrambled versions. The cueing contingency remained in place—that is, 80% of targets appeared in the same position as the emotionally congruent face image—even in those trials using scrambled face images.

Participants responded to the target (indicating location and orientation) and the subjective awareness probe (indicating breakthrough), as usual. Following the subjective awareness probe, an additional objective probe was presented binocularly. For the objective probe, participants were asked to complete the stem sentence "On this trial, if there were other images, they were…" with either "faces" or "scrambles". Responses were mapped to the same response keys as the target-discrimination and subjective-awareness responses.

## Data analysis

Data was pre-processed to identify participants for exclusions (suppression analyses) and to prepare data for hypothesis testing analyses. The primary analyses were completed using JASP and R.

**Suppression checks.** Suppression check analyses were conducted to ensure that unaware-condition data included only those trials and participants for which suppression was successful. To this end, each participant was evaluated on their proportion of reported breakthrough trials (subjective awareness check) and on their face–scramble discrimination accuracy during control-block trials (objective awareness check). Participants who reported high rates of subjective breakthrough, or who were able to reliably discriminate faces and scrambles despite reporting no breakthrough, were excluded and replaced.

*Subjective awareness check.* For each unaware-group participant, the proportion of breakthrough trials was separately calculated for valid experimental trials, invalid experimental trials, and all control-block trials. To ensure that sufficient valid and invalid experimental trials were retained for hypothesis-testing analyses, participants who reported more than 75% breakthrough in either cueing condition were excluded and replaced. Additionally, to ensure that each participant had sufficient no-breakthrough trials for the objective awareness check, participants who reported more than 75% breakthrough in the control block (regardless of cue validity) were excluded and replaced.

*Objective awareness check.* Subjective awareness checks are limited by individual differences in detection criteria (i.e., how much breakthrough must occur before a participant is willing to report seeing an image), working memory and response biases. For participants' trial-by-trial reports of breakthrough to be a useful measure of awareness, it is important to ensure that participants are able to reliably report awareness of the masked images.

To evaluate participants' self-report accuracy, each participant's face–scramble discrimination accuracy was separately calculated for breakthrough and no-breakthrough trials. I expected that if participants' self-reported breakthrough accurately tracked awareness of the masked stimuli, they should be able to discriminate faces and scrambles with greater-than-chance accuracy for breakthrough trials (because their self-report indicated at least partial awareness of the masked images). Conversely, I expected that participants should be unable to discriminate faces and scrambles on trials where they reported no breakthrough. If they could make this judgement with above-chance accuracy having reported no awareness of the masked images, this would suggest that their self-reported breakthrough ratings were underestimating their true awareness—and therefore could not be taken as a reliable indication of suppression.

For no-breakthrough control trials, participants' face–scramble accuracy was compared to chance (50%) on a single-sample *z*-test. Participants who obtained greater-than-chance accuracy on no-breakthrough trials were excluded and replaced.

**Hypothesis testing.** Hypothesis testing analyses were carried out for experimental block trials only. To avoid including breakthrough trials in analyses for the unaware group, all trials that included a report of subjective breakthrough were excluded. Mean accuracy and response time were then calculated separately for each participant's valid and invalid trials. Accuracy was defined as the proportion of responses that correctly identified both location and orientation of the target grating. Mean response times were calculated for experimental block trials in which participants correctly identified the target location and orientation (both awareness groups) and indicated no breakthrough (unaware group only). Outliers more than two standard deviations above a participants' mean response time were also removed, as were all trials with a response time less than 200 ms (which were more likely to be anticipations than genuine responses).

For one participant, the MATLAB script failed to correctly load the target grating on 28 trials. As this was only discovered during data analysis, the participant was not replaced. Instead, the affected trials were omitted from both accuracy and response time analyses.

## Results

The aim of this study was to test whether the conjunction of emotionally congruent facial expressions and vocal prosodies cue spatial attention under conditions of (a) full perceptual awareness (aware group), and (b) visual unawareness (unaware group). If a significant spatial-cueing effect were observed for the aware group, this would support the hypothesis that the congruence-cued spatial attention task provides indirect evidence of multisensory integration. If a significant spatial-cueing effect were also observed for the unaware group, this would support the hypothesis that multisensory integration of unconscious perceptual stimuli can occur, at least within the same sensory modality as the unconscious stimulus.

In order to test these hypotheses, I calibrated the congruence-cued spatial attention task to constrain target-discrimination accuracy midway between chance and ceiling, conducted subjective and objective awareness checks for unaware-group participants, and recorded target-discrimination accuracy and response times for all participants. Results for suppression, calibration and spatial cueing are reported separately below.

## Suppression checks

Nine unaware-group participants were excluded, with replacement, based on either subjective (*n* = 7) or objective (*n* = 2) awareness checks. Thus, masking achieved successful suppression for 72.72% of unaware-group participants. Retained participants reported no breakthough (i.e., successful suppression) for 78.35% of experimental trials (*SD* = 23.79%). All experimental trials in which subject breakthrough was reported were excluded from the spatial cueing analyses reported below.

In the control task, retained participants reported no breakthrough for 73.83% of trials (*SD* = 23.49%). Across all retained participants, face-scramble accuracy for breakthrough trials was 73.88%, compared to 45.33% for no-breakthrough trials. These findings are consistent with the prediction that unaware-group participants should only be able to discriminate faces and scrambles when they become subjectively aware of the masked images.

## Calibration and target-discrimination accuracy

Following the calibration task, participants were assigned targets with mean orientation of 1.72° (*SD* = 1.14°) either side of vertical (see *Table 2* for counts of participants assigned each target orientation). Participants' target-discrimination accuracy for their assigned target orientations during the calibration task converged on the desired mean (*M* = 65.41%, *SD* = 7.91%, range: 50–90%; collapsing across awareness conditions). Importantly, assigned target orientations did not differ between awareness conditions ($p = .583$, Fisher's exact test), suggesting that task difficulty remained constant between groups.

Mean accuracy for experimental trials, collapsing across awareness and cueing conditions, closely approximated the expected accuracy (*M* = 65.26%, *SD* = 10.97%). Accuracy for each participant significantly exceeded chance (i.e., 25%), and fell short of ceiling (i.e, 100%) on independent single-sample *z*-tests (all *p*s < .001). Thus, calibration successfully constrained target-discrimination around the desired mean, and prevented floor and ceiling effects.

Finally, an exploratory analysis contrasting accuracy for target location with accuracy for target orientation showed that participants identified the target location (*M* = 98.85%, *SD* = 1.20%) with greater accuracy than target orientation (*M* = 65.92%,

Table 2

*Assigned target orientations following calibration for* Experiment 1

| Awareness condition | Target orientation | | | | |
|---|---|---|---|---|---|
| | ± 1° | ± 2° | ± 3° | ± 4° | ± 5° |
| Aware | 17 | 2 | 4 | 0 | 1 |
| Unaware | 13 | 5 | 4 | 0 | 2 |

*Note:* Participants were assigned the target orientation for which their target-discrimination accuracy was closest to 65% during calibration trials. In the case of a tie between two orientation levels, participants were assigned the orientation with smallest rotation angle. Calibration data for excluded participants are not shown.

*SD* = 10.73%), $t(47) = 22.026$, $p < .001$, $d = 3.179$. Indeed, participants made almost no errors for target location, suggesting that target-discrimination accuracy was primarily driven by the discrimination of orientation and not location. Thus, the chance threshold of 25% used in this experiment—and by Palmer and Ramsey (2012)—most likely underestimates true chance performance. A more conservative threshold of 50% would likely provide a better estimate of chance accuracy.

Applying this more stringent criterion for above-chance accuracy would have resulted in the replacement of eight participants who are retained in this experiment (three unaware participants, five aware participants). While inclusion of these participants may reduce the statistical power to detect an effect in this experiment, as these participants may effectively be guessing the orientation of the target gratings, excluding them does not change the significance of for any of the effects reported below.

**Spatial cueing**

Participants' mean target-discrimination accuracy and response times for experimental trials were analysed for evidence of spatial cueing, with accuracy considered the primary dependent measure. Both accuracy and response time data were submitted to separate 2 × 2 mixed measures ANOVAs, with a within-subjects variable of cue validity (two levels: valid, invalid) and a between-subjects variable of group (two levels: aware, unaware).

If multisensory integration is responsible for the congruence-cueing effects reported by Palmer and Ramsey (2012), aware participants should show greater accuracy—and possibly faster correct-response times—for valid trials, compared to invalid trials. Spatial cueing during perceptual awareness would support the hypothesis that spatial attention can be used as an effective, indirect assessment of multisensory integration, validating the use the congruence-cued spatial attention task to probe unconscious integration.

If multisensory integration occurs in the absence of conscious awareness (and congruence cueing is believed to indicate integration), then unaware participants should also show greater accuracy—and possibly faster correct-response times—for valid trials, compared to invalid trials. Spatial cueing for unaware participants, in addition to aware participants, would constitute a conceptual replication of Palmer and Ramsey (2012) and support the hypothesis that integration occurs in the absence of conscious awareness.

**Accuracy.** The ANOVA for accuracy showed no main effect for either awareness condition, $F(1, 46) = .421$, $p = .520$, $\eta^2_p = .009$, or cue validity, $F(1, 46) = .756$, $p = .389$, $\eta^2_p = .016$, and no interaction, $F(1, 46) = .006$ $p = .941$, $\eta^2_p < .001$ (Figure 14). This

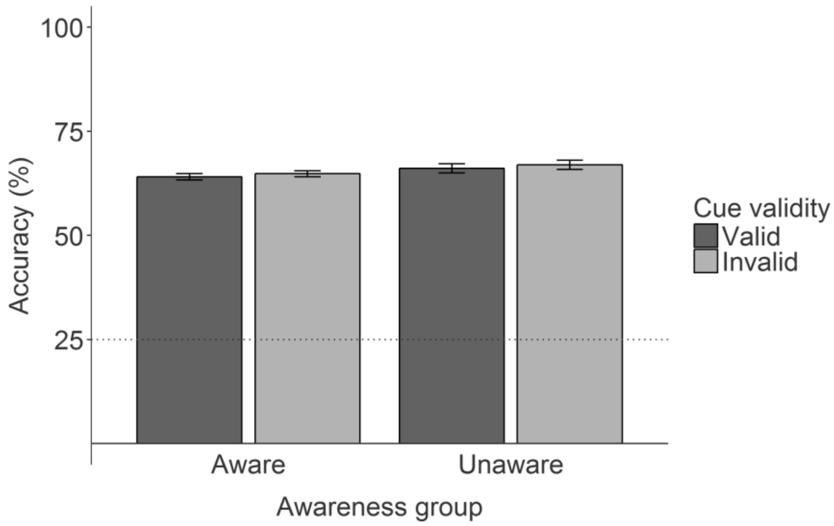


*Figure 14.* Target discrimination accuracy by awareness condition (*Experiment 1*). No significant effects of awareness or cue validity, and no significant interaction, were observed. Dotted line represents chance accuracy. Error bars represent within-subjects standard errors (Morey, 2008).

suggests that neither aware (valid: *M* = .641, *SD* = .119; invalid: *M* = .648, *SD* = .118) nor unaware (valid: *M* = .661, *SD* = .104; invalid: *M* = .670, *SD* = .124) participants were more accurate following valid than invalid cues—and therefore that neither group learned to allocate spatial attention in response to the face–voice congruence cues.

Thus, the claim that multisensory integration occurs outside of conscious awareness is not supported. Additionally, the failure to observe a spatial-cueing effect even for aware participants suggests that a congruence-cued spatial attention task may not be a valid measure of multisensory integration at all.

**Response times.** The ANOVA for response times revealed a significant main effect of awareness condition, $F(1, 46) = 11.99$, $p < .001$, $\eta^2_p = .207$ (Figure 15). Aware participants (valid: *M* = 722 ms, *SD* = 133 ms; invalid: *M* = 729 ms, *SD* = 148 ms) responded significantly faster than unaware participants (valid: *M* = 911 ms, *SD* = 223 ms; invalid: *M* = 919 ms, *SD* = 238 ms). This response time advantage for aware participants may have arisen due to either reduced cognitive load in that condition (aware participants responded to targets only, whereas unaware participants responded to targets and to subjective awareness probes), or to increased interstimulus offset in that condition

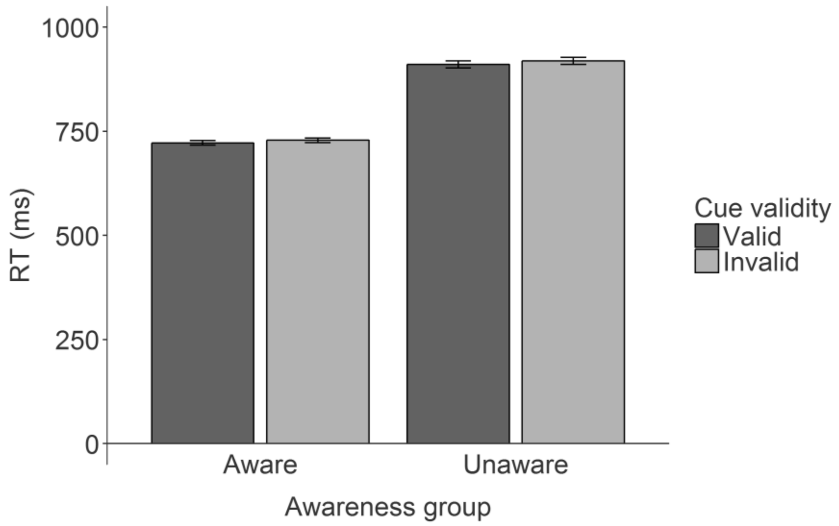


*Figure 15.* Response times for correct responses by awareness condition (*Experiment 1*). Aware participants respond faster than unaware participants. No effect of cue validity, and no interaction. Error bars represent within-subjects standard errors (Morey, 2008).

(the CFS masked persisted for 50 ms after emotional face images were removed). However, as this comparison has no implication for the hypotheses being tested, no further exploration of this finding was attempted.

No main effect was observed for cue validity, $F(1, 46) = 1.054$, $p = .310$, $\eta^2_p = .022$, suggesting that neither group were faster to respond to validly cued targets. The interaction was also not significant, $F(1, 46) = .013$, $p = .908$, $\eta^2_p < .001$. Thus, response times also provide no support for spatial cueing by emotional congruence, in either aware or unaware conditions.

## Discussion

*Experiment 1* shows no spatial-cueing effect for either dependent variable. Thus, the experiment provides no support for the hypothesis that spatial cueing by cross-modal congruence provides a valid indication of multisensory integration; nor for the hypothesis that multisensory integration can occur in the absence of perceptual awareness.

The one significant finding—that aware-group participants responded faster than unaware-group participants—is most likely an artefact of the task itself. There are two plausible explanations for this effect. First, aware participants completed a single task of target discrimination on each trial, whereas unaware participants completed both the target-discrimination task and provided a subjective report of breakthrough. Although the awareness check occurred after the target response was received, the conjunction of two tasks per trial may have increased the working-memory load for unaware-group participants relative to the aware group, and thus contributed to slower response times for unaware participants. Second, cueing-phase stimuli for aware participants (i.e., emotional faces) offset 160 ms before target gratings appeared, whereas cueing-phase stimuli for unaware participants (i.e., emotional faces and CFS masks) offset 110 ms before targets displayed. This difference arose due to the later offset of the CFS mask, which was intended to aid suppression. However, this explanations is less likely, as the SOAs for audiovisual cues and targets grating did not differ between awareness conditions—thus preparatory time, including both the cue and ISI remained the same for all participants.

More importantly, the lack of a significant spatial-cueing effect, even in aware participants, suggests that further investigation of the congruence-cued spatial attention task itself is required. Broadly, there are three plausible explanations for the failure to replicate the expected spatial-cueing effect.

First, it may be the case that the congruence-cued spatial attention task is not a suitable test for inferring multisensory integration. If this were so, the results reported by Palmer and Ramsey (2012) may simply be spurious, as the authors' use of small samples in each of their experiments would have increased the risk of false positives (Button et al., 2013). However, their replication of the same spatial-cueing effect across three separate experiments argues against this explanation. It may also be the case that the reported findings are real, but reflect some other cognitive process and not multisensory integration. Yet, if this were the case, it would be surprising that the same cognitive process did not contribute to significant cueing effects in the current experiment, regardless of the underlying mechanisms.

Second, congruence-cued spatial attention may genuinely imply multisensory integration, but of a kind that does not extend to the integration of emotional expressions. The audiovisual cues presented by Palmer and Ramsey (2012) represented *temporally* congruent or incongruent speech events, whereas the audiovisual cues presented in this experiment were *semantically* congruent or incongruent (i.e., they meaningfully expressed happiness or sadness). While we might expect that emotional faces and voices temporally co-occur in the real world, the presentation of static images of emotional information may mean that they do not have the same synchrony in this experiment (for the importance of synchronous presentation in multisensory integration see Koelewijn et al., 2010). If the cross-modal integration of temporally and semantically congruent information depends on different cognitive processes or neural pathways, that may explain why the use of emotionally congruent cues did not produce the same spatial-cueing effects as the conjunction of auditory and visual speech events.

Finally, there may be a critical methodological difference between the original paradigm and my replication attempt. While I had made several methodological changes for this experiment, most of these were unlikely to account for the observed null results. Yet one change—the removal of the requirement that participants explicitly identify the cue-relevant features of the voice cue—may plausibly account for the current findings. In their discussion, Palmer and Ramsey (2012) suggested that unconsciously multisensory integration may occur only when participants consciously the stimuli in the influencing modality (i.e., the voice cues). However, the null results reported here may suggest that participants must do more than simply be aware of the voice cues—they may have to explicitly identify the cue-relevant features of the voice in order to respond to the congruence-based spatial cues.

Having identified these three possible explanations for the failure to replicate the expected spatial-cueing effects, I designed *Experiment 2* to investigate the latter, methodological, explanation. Specifically, I manipulated the perceived task-relevance of the emotional voice tokens, by assigning participants to one of two emotion-discrimination conditions. On each trial, participants in one condition (*with-report*) indicated both the location and orientation of the target, and the emotional prosody of the voice; while those in the other condition (*no-report*) indicated only the location and orientation of the target. Participants in both conditions remained perceptually aware of all stimuli (i.e., there was no masking manipulation), because the finding of an aware spatial-cueing effect would help to establish task validity of the congruence-cued spatial attention task, and is a necessary prerequisite to a robust investigation of unconscious multisensory integration using this task. Thus, *Experiment 2* tests the hypothesis that spatial cueing by cross-modal congruence occurs only when the cue-relevant features of the cue in the influencing domain are explicitly identified.

## Experiment 2

Having found no significant spatial-cueing effects in either awareness condition in *Experiment 1,* I conducted a further study to investigate the possibility that participants had failed to attend to the cue-relevant information of emotion in the voice recordings. If this were the case, then making the emotional prosody explicitly task-relevant should ensure that the relevant emotional information is attended, potentially allowing participants to respond to the spatially informative cue provided by emotionally congruent faces and voices.

In the original congruence-cued spatial attention study, Palmer and Ramsey (2012) concluded that unconscious multisensory integration may only occur when the *influencing modality* is consciously perceived. According to this account, based on global workspace theories of consciousness, consciously perceived sounds are globally broadcast, and are therefore available for integration with unconsciously perceived visual information within vision-specific cognitive processes. However, the finding of *Experiment 1* may suggest a more stringent requirement: that participants must explicitly attend to the cue-relevant features of the stimulus in the influencing domain—in this case, to the emotional prosody of the voice.[5]

In their original study, Palmer and Ramsey (2012) had participants repeat the last syllable they heard aloud at the end of every trial. While participants' verbal responses were not analysed, the manipulation ensured that participants experienced—and indeed, attended to—the cue-relevant features of the voice cue (in this case, the spoken syllables).

To investigate whether explicit discrimination of cue-relevant features in the influencing modality is necessary for spatial cueing by cross-modal congruence, I assigned participants to one of two emotion-discrimination groups. One group of participants explicitly identified the emotional prosody on each trial (*with-report* group), while the other group did not (*no-report* group). Both groups were still required to report the location/orientation of the target grating. Aside from the addition of the emotion-discrimination task in the with-report condition, experimental trials followed the same procedure as aware-group trials in *Experiment 1*.

[5] While other forms of associative learning can occur without explicit processing of context (Esteves, Parra, Dimberg, & Öhman, 1994), cueing by cross-modal congruence has not been well studied and may well require explicit identification of cue-relevant features of the stimuli presented in the influencing modality.

Perceptual awareness was not manipulated in this experiment: both faces and voices were presented without masking to all participants. The rationale for conducting this experiment with aware participants only, was twofold. First, if the congruence-cued spatial attention task does assess multisensory integration, then aware spatial-cueing effects should be obtainable regardless of whether multisensory integration occurs in the absence of visual awareness. A failure to obtain aware spatial-cueing effects would therefore suggest the unconscious cueing effects are either not real, or not due to multisensory integration. Thus, finding an aware spatial-cueing effect is a necessary prerequisite to interpreting any unconscious spatial-cueing effects, should they be found in subsequent attempts at replication. Second, conducting the experiment with aware participants removes the need for visual masking, subjective and objective awareness checks, and for substantial replacement of participants for whom suppression was unsuccessful. Thus, conducting *Experiment 2* with aware participants only allowed me to investigate the critical question—whether cross-modal congruence can cue spatial attention—using a simpler experimental procedure.

In *Experiment 2,* I employed similar experimental design to the previous experiment. Specifically, I used a 2 × 2 mixed factorial design, in which emotion discrimination was manipulated between subjects (two levels: with-report, no-report) and cue validity was manipulated within-subjects (two levels: valid, invalid). As in the previous experiment, the emotional congruence of vocal prosody and facial expression predicted the most likely location of the upcoming target grating with 80% validity. Target-discrimination accuracy and response times were recorded for all participants, and emotion-discrimination responses were also recorded for participants in the with-report group. Target-discrimination accuracy remained the primary dependent measure for hypothesis testing (while response times were analysed as secondary measure). Emotion-discrimination accuracy was analysed to ensure that all participants in the with-report group were able to reliably discriminate *happy* and *sad* vocal prosodies, but was not used for hypothesis testing.

Finally, the procedure included four minor procedural changes, to streamline the experimental process and aid in the interpretation of any significant results. First, chance accuracy was defined as 50%, increased from the 25% threshold used in *Experiment 1*. The target-discrimination task involves two separate judgements of target location and target orientation. Participants in the previous experiment made almost no errors for target location (placing them essentially at ceiling), whereas they made multiple errors for target orientation. These findings suggest that the critical forced-choice decision is the discrimination of target orientation and not location—thus the chance of accurately *guessing* is 50%. By increasing the threshold for chance accuracy, participants that identify the target location accurately but guess at the target orientation can be identified and replaced. Second, the calibration block was removed. Previously, the majority of participants (62.5%) were assigned target gratings oriented ± 1° from vertical. Rather than calibrating targets for each participant, I elected to run the task with ± 1° targets only. This had the advantage of shortening the length of the experiment, and avoiding the potential for participant fatigue, but ran the risk that some number of participants would be excluded for having target-discrimination accuracy indistinguishable from chance. Third, the practice block was expanded to 40 trials and altered to introduce features of the experimental trials in stages. This provided a more comprehensive training process prior to participants beginning the experimental task. Finally, a post-test questionnaire was developed to investigate whether participants had explicitly identified the cueing contingency during the task. The inclusion of the questionnaire allowed me to conduct an additional analysis of whether the explicit identification of the cueing contingency might account for any observed spatial-cueing effects. Together, these changes were expected to streamline the experimental sessions without fundamentally altering the task, and were therefore not expected to influence the results in any systematic way.

Thus, *Experiment 2* aimed to test the hypothesis that spatial cueing by cross-modal congruence occurs only when cue-relevant features in the influencing modality are explicitly attended. The prediction implied by this hypothesis, is that a significant spatial-cueing effect—that is greater target-discrimination accuracy for valid trials as compared to invalid trials—should be observed for the with-report group, but not the no-report group. A spatial-cueing effect for with-report participants would additionally provide support for the validity of the congruence-cued spatial attention task as an indicator for multisensory integration. The lack of spatial-cueing effect for no-report participants would replicate the null findings for aware-group participants in *Experiment 1*.

## Method

### Participants

Seventy-four individuals were recruited via on-campus posters, social media advertising, and email invitations addressed to participants of previous experiments conducted by our lab. None of the recruited participants had participated in *Experiment 1*, or in the pilot studies used to select the cue or target stimuli. All participants reported normal or corrected-to-normal vision and hearing. Participants received a voucher for either groceries or cinema admission to thank them for participating.

The sample size of 24 participants per condition (after exclusions for chance target-discrimination accuracy) was retained from *Experiment 1*.[6] Thus, 48 participants were randomly assigned to either the with-report or no-report emotional task conditions. To ensure participants were able to discriminate the low-contrast targets, each participant's target-discrimination accuracy was compared to chance (50%) on a single-sample *z*-test. Participants whose accuracy did not significantly exceed chance were excluded with replacement. This process required the replacement of 26 participants (16 with-report participants; 10 no-report participants). These exclusions can be attributed to the removal of the calibration task used in *Experiment 1*, and to the increased threshold for chance performance used to assess participants' response accuracy.

**Stimuli**

All emotional cue stimuli (i.e., emotional face images and voice recordings) were retained from *Experiment 1*. Target gratings with orientations of ± 1° from vertical were also retained.

Additionally, a novel post-test questionnaire was developed to investigate whether participants explicitly identified the cueing contingency during the experiment. To allow participants to spontaneously report the cueing contingency, the questionnaire presented five open-ended, but increasingly specific, questions about the relationship between emotional voices, emotional faces and target gratings. Each question appeared on a separate sheet of paper. All five questions are presented in *Table 3*.

---

[6] Interestingly, this sample of size was calculated based on a conservative estimate of the true effect size for the unconscious spatial-cueing effects reported by Palmer and Ramsey (2012). Because the current experiment includes two aware conditions (instead of one aware, one unaware), retaining the same sample size actually *increases* its power to detect a spatial-cueing effect for aware participants when collapsing across emotional task conditions.

Table 3

*Post-test contingency awareness questionnaire (Experiment 2)*

| Question number | Question |
|---|---|
| 1 | Did you discover or implement any strategy that helped you to detect the target grating? If so, what was it? |
| 2 | Did you notice any relationship between the cue stimuli (i.e., faces images and voice recordings) in your condition? If so, what was it? |
| 3 | Did you notice any relationship between the cues (i.e., faces and voices) and targets (i.e., gratings) in your condition? If so, what was it? |
| 4 | Did you notice that the cues predicted the location of the target? If so, please describe the relationship. |
| 5 | Did you notice that the target typically appeared in the same location as the *congruent* face (i.e., the face that expressed the same emotion as the concurrent voice recording)? |

*Note*. Each question appeared on a separate sheet of paper. Participants were instructed to answer each question in order, without reading ahead or returning to amend previous answers.

**Procedure**

Participants completed the congruence-cued spatial attention task alone, under the same viewing conditions as in the previous experiment. The experimental task presented a single practice block (40 trials) and four experimental blocks (320 trials, 80 per block) for all participants. Following the experimental task, participants completed the post-test awareness questionnaire. They were instructed to complete each question in order, without reading ahead or returning to amend previous answers.

**Trials.** Trials presented emotional face–voice cues and target gratings with the same layout and display timings as in the previous experiment, and retained the same ratio of valid-to-invalid trials (i.e., 80:20) in each experimental block. Unlike the previous experiment, the response window for participants' target-discrimination responses was limited to 1500 ms (obviating the need for the removal of trials based on response-time outliers).

Experimental trials for no-report participants were identical to aware-group trials in the previous experiment (with the exception of the reduced response window). Experimental trials for with-report participants were the same, aside from the inclusion of the emotion-discrimination judgement at the end of each trial. For these participants, a written question (“Was the voice HAPPY or SAD?”) displayed onscreen immediately after a target-discrimination response was received (or the response window had ended with no response). Participants responded with either *Happy* or *Sad* (two-alternative, forced-choice), using the left-most and right-most keys on their response box. The key bindings for each emotion were consistent for each participant (i.e., happy and sad response buttons remained unchanged during the experiment) and counterbalanced between participants (i.e., half of the participants in each group reported happy voices with left hand and sad with the right, and half had the response buttons reversed). Response windows for emotion-discrimination were open-ended.

The practice block adopted a new training routine, in which participants completed four 10-trial mini-blocks, in which features of the task were introduced consecutively. Unlike the previous experiment, practice trials presented both emotional face and voice cues, such that they contained informative cues with the same cueing contingency as experimental trials. Each mini-block maintained a consistent ratio of valid (80%) and invalid (20%) cues. Feedback was provided onscreen following each target response, with either “CORRECT” (in green text) or “INCORRECT” (in red text) briefly replacing the fixation cross in the centre of the display.

For participants in the with-report group, practice mini-blocks progressed through the following stages: (1) 10 trials including the target-discrimination task only, with an increased target duration (140 ms) and open-ended target response windows; (2) 10 trials including both target- and emotion-discrimination tasks, and retaining the increased target duration and open-ended target response windows; (3) 10 trials including normal target durations (70 ms), but retaining open-ended target response windows; (4) 10 trials matching the procedure for experimental trials (i.e., target- and emotion-discrimination tasks, 70 ms target durations, 1500 ms target response windows). For participants in the no-report group, practice followed the same mini-block structure with the exception that the emotion-discrimination task was never included.

## Data analysis

Hypothesis-testing analyses were carried out for experimental block trials only. Mean target-discrimination accuracy and response times were calculated for each participant, and submitted to separate mixed-model ANOVAs. Consistent with the previous experiment, accuracy is considered the primary measure for the purpose of hypothesis testing, while response times are analysed as a potential source of supporting evidence. All hypothesis-testing analyses were preregistered. The preregistration document is publically available (https://aspredicted.org/bx6g5.pdf), and a copy is included in the *Appendix*.

Mean accuracy was defined as the proportion of correct responses, in which participants accurately identified both target location and orientation. Mean response times were calculated for trials in which participants provided correct target responses, excluding trials in which a response was recorded before 200 ms (as these were more likely to be anticipations rather than genuine responses to a target grating).

Unplanned trial exclusions were necessitated by technical issues that occurred during experimental trials. Specifically, the MATLAB script failed to load the target gratings for some trials, resulting in their unintentional omission. This issue occurred for 127 trials, across all retained participants (representing 0.83% of all experimental trials). Nine participants were affected (4 with-report, 5 no-report). Affected trials were omitted from all analyses.

Data was pre-processed to identify and replace participants that did not meet accuracy criteria. The primary analyses were completed using JASP and R.

# Results

## Target-discrimination accuracy

Each participant's mean target-discrimination accuracy, collapsing across cue validities, was compared to chance (50%) and ceiling (100%) on separate single-sample *z*-tests. Where accuracy did not significantly exceed chance, or differ from ceiling, participants were replaced. Ten emotion-unattended, and 16 emotion-attended participants were excluded due to their accuracy being indistinguishable from chance. No participants were at ceiling.

For participants in the with-report group, each participant's mean emotion-discrimination accuracy was compared to chance (50%) on separate single-sample *z*-tests. This check was included to ensure that with-report participants were able to reliably discriminate—and had therefore attended to—the emotional prosody expressed in each voice token. All participants in the emotion-attended group identified the emotional prosody of the voice cues with greater-than-chance accuracy ($M = .943$, $SD = .044$, all $p$s $< .05$).

**Spatial cueing**

Consistent with the previous experiment, participants' mean target-discrimination accuracy was considered the primary dependent measure. Both accuracy and response times were submitted to separate 2 × 2 mixed measures ANOVAs, with a within-subjects variable of cue validity (two levels: valid, invalid) and a between-subjects variable of emotion discrimination (two levels: with-report, no-report).

**Accuracy.** The ANOVA for accuracy (Figure 16) showed no main effects for emotion discrimination, $F(1, 46) = .069$, $p = .793$, $\eta^2_p = .002$, or cue validity, $F(1, 46) = 1.255$, $p = .268$, $\eta^2_p = .027$, but the interaction of emotion discrimination and cue validity was significant, $F(1, 46) = 4.275$, $p = .044$, $\eta^2_p = .085$. The interaction was explained by a significant difference in target-discrimination accuracy for valid and invalid trials in the with-report condition, $t(23) = 2.471$, $p = .021$, $d = .504$, but not in the no-report condition, $t(23) = .620$, $p = .541$, $d = .127$. Specifically, with-report participants showed greater target-discrimination accuracy for valid trials ($M = .660$, $SD = .048$) than for invalid trials ($M = .631$, $SD = .067$); whereas no-report participants showed no difference in accuracy between valid ($M = .637$, $SD = .068$) and invalid ($M = .646$, $SD = .075$) trials. These findings support the hypothesis that spatial cueing by cross-modal congruence occurs only when cue-relevant features in the influencing modality are explicitly attended.

**Response times.** The ANOVA for response times (Figure 17) revealed significant main effects of cue validity, $F(1, 46) = 5.726$, $p = .021$, $\eta^2_p = .111$, and emotion discrimination, $F(1, 46) = 12.22$, $p = .001$, $\eta^2_p = .210$. The interaction was not significant,

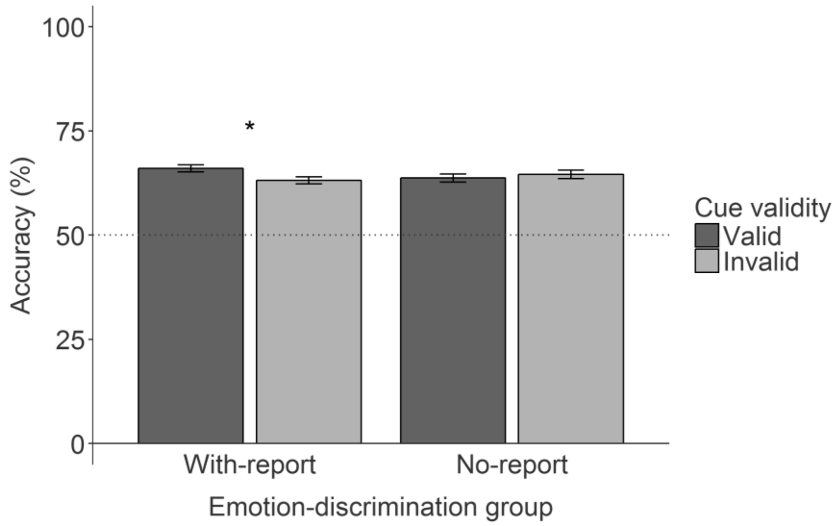


*Figure 16.* Target discrimination accuracy by emotion-discrimination group (*Experiment 2*). When emotional prosody was reported, participants' target-discrimination accuracy was greater for valid trials than invalid trials. When prosody was not reported, no difference in accuracy was observed between valid and invalid trials. Dotted line represents chance accuracy. Error bars represent within-subjects standard errors (Morey, 2008).

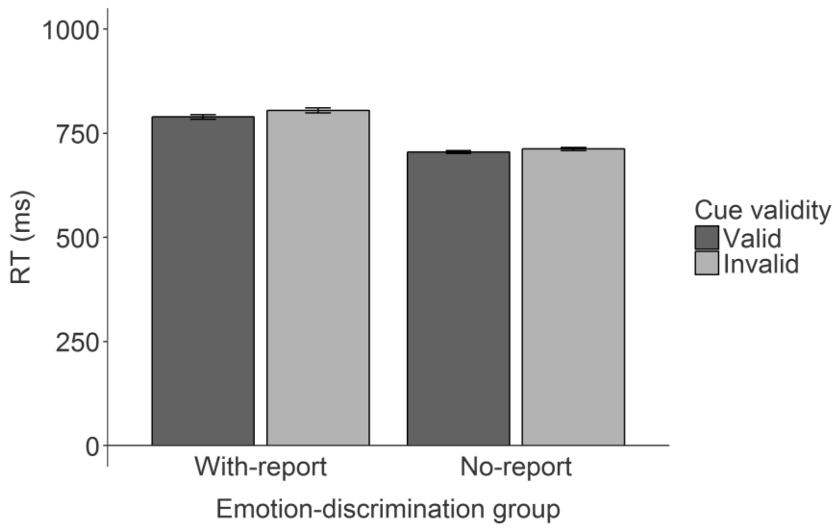


*Figure 17.* Response times for correct responses by emotion-discrimination group (*Experiment 2*). Error bars represent within-subjects standard errors (Morey, 2008).

$F(1, 46) = .774$, $p = .393$, $\eta^2_p = .016$.

The main effect of cue validity reflected faster response times for valid trials (emotion-attended: *M* = 789 ms, *SD* = 69 ms; emotion-unattended: *M* = 705 ms, *SD* = 99 ms) than invalid trials (emotion-attended: *M* = 805 ms, *SD* = 79 ms; emotion-unattended: *M* = 712 ms, *SD* = 106 ms). Although this main effect suggests that both with- and no-report groups show a spatial-cueing effect for response times, this conclusion was not supported by the data for accuracy (which was considered the more informative measure for spatial cueing). To explore this dissociation, exploratory comparisons were conducted separately for each group. These comparisons revealed that the response-time advantage for validly cued trials was not significant for either group alone (with-report: $t(23) = 1.924$, $p = .067$, $d = .393$; no-report: $t(23) = 1.436$, $p = .165$, $d = .293$). Thus, the main effect of cue validity provides some support for an overall spatial-cueing effect, but provides no support for the primary hypothesis of the experiment. Given that response time data is likely the weaker measure of spatial attention in this paradigm (as the task was optimised for accuracy), the finding of an overall spatial-cueing effect for response times is unlikely to be meaningful.

The main effect of emotion discrimination reflected faster correct response times for no-report participants than for with-report participants, regardless of cue validity. This difference likely reflects greater cognitive load in the with-report group, due to the additional task of discriminating emotional prosody. Although participants in that group indicated the emotional prosody after indicating the target location and orientation, the task required that they maintain their emotion-discrimination judgement in working memory throughout the trial. Regardless of whether this explanation is correct, the effect has no bearing the hypothesis being tested.

**Contingency awareness**

Participants who reported awareness of the cueing contingency (i.e., that targets were more likely to appear in the location of the congruent emotional face) at any point in the post-test questionnaire were considered *contingency aware*; remaining participants were considered *contingency unaware*. Surprisingly, the proportion of participants that reported noticing the cueing contingency during the experiment did not differ between with-report (10 of 24; 41.67%) and no-report (12 of 24; 50.00%) groups, $\chi^2(1) = .337$,

$p$ = .438. Because explicit awareness of the cueing contingency entails that participants attended to the emotional prosody, this finding suggests that at least half of the no-report participants attended to the emotional prosody even in the absence of instructions to do so. Additionally, this finding suggests that explicitly discriminating the emotional prosody does not necessarily imply awareness of the cueing contingency. Finally, because a significant spatial-cueing effect for accuracy was found in the with-report condition only, these observations suggest that explicit awareness of the cueing contingency is not necessary for spatial cueing.

Yet, it is still possible that contingency awareness may contribute to spatial cueing. To investigate whether awareness of the cueing contingency contributed to spatial cueing across both emotional task conditions, I conducted an exploratory three-way ANOVA for target-discrimination accuracy, with contingency awareness (aware, unaware) and emotion discrimination (with-report, no-report) as between-subjects variables, and cue validity as a within-subjects variable (valid, invalid). If contingency awareness contributes to cross-modal congruence cueing, then contingency-aware participants should show greater accuracy for valid trials than invalid trials; whereas contingency-unaware participants should show no advantage for valid trials. This would be reflected in a significant two-way interaction of contingency awareness and cue validity, or three-way interaction of all independent variables.

The ANOVA revealed no significant main effects (contingency awareness: $F(1, 44) = .003$, $p = .953$, $\eta^2_p < .001$; cue relevance: $F(1, 44) = 046$, $p = .831$, $\eta^2_p = .001$; cue validity: $F(1, 44) = 1.830$, $p = .183$, $\eta^2_p = .040$; see Figure 18). Critically, neither of the predicted interactions were significant (contingency awareness × cue validity: $F(1, 44) = 2.186$, $p = .146$, $\eta^2_p = .047$; contingency awareness × emotion discrimination × cue validity: $F(1, 44) = 1.407$, $p = .242$, $\eta^2_p = .031$). Finally, the interaction of emotion discrimination and cue validity was significant, $F(1, 44) = 5.333$, $p = .026$, $\eta^2_p = .108$. However, this interaction reflects the spatial-cueing effect already reported for target-discrimination accuracy in the two-way ANOVA above. These results suggest that contingency awareness does not meaningfully contribute to the observed spatial-cueing effects for target-discrimination accuracy.

Individual accuracy scores for valid and invalid trials, grouped by contingency awareness, are presented in a Brinley plot (Figure 19). Participants whose target-discrimination accuracy was greatest for valid trials are shown above the diagonal, while those whose accuracy was greatest for invalid trials are shown below. Thus, the spatial-

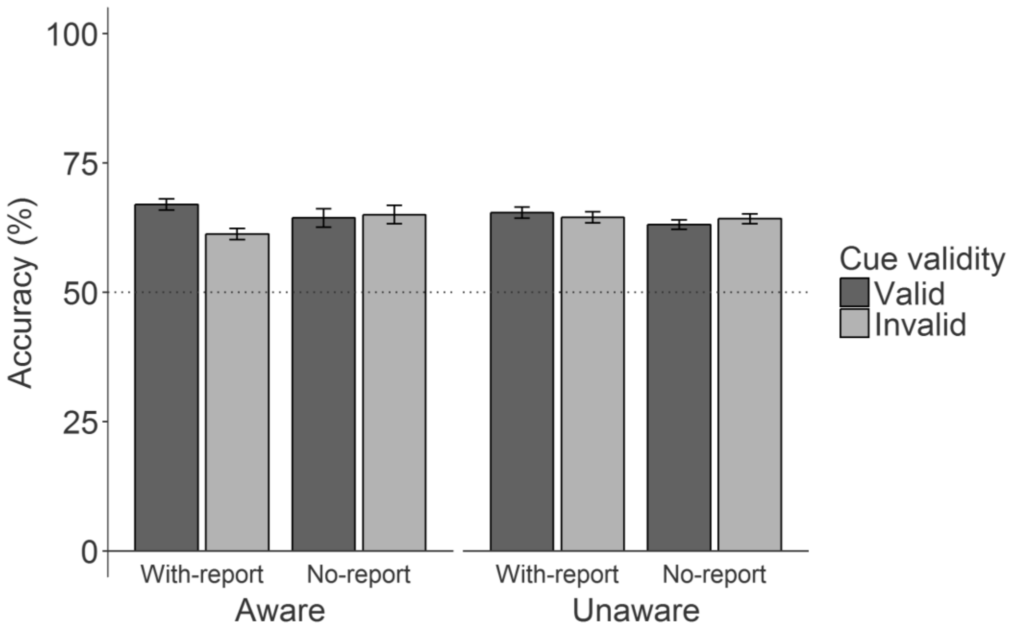


*Figure 18.* Target-discrimination accuracy for contingency-aware and contingency-unaware participants. Dotted line represents chance accuracy. Error bars represent within-subjects standard errors (Morey, 2008).

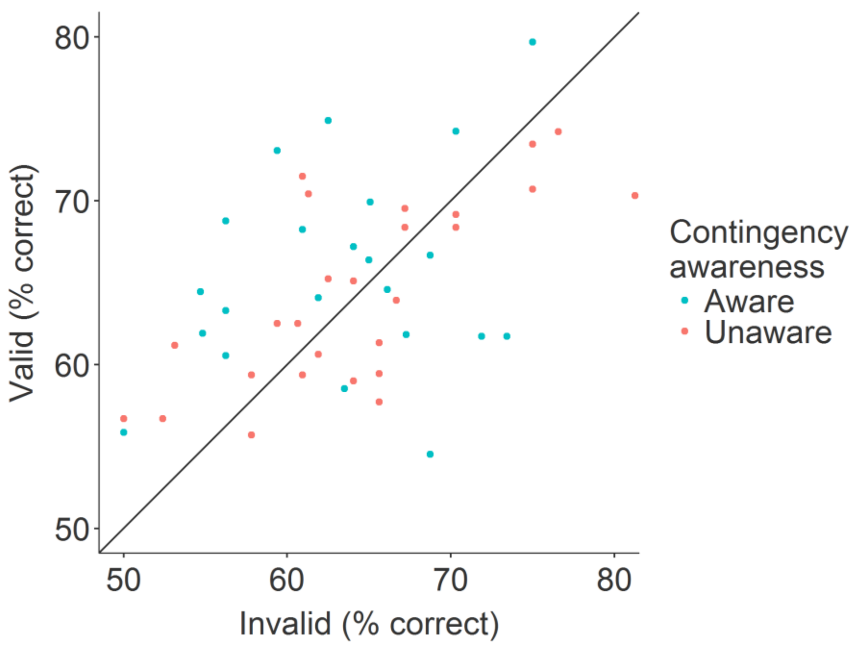


*Figure 19.* Brinley plot comparing participants' relative accuracy for valid and invalid trials, grouped by self-reported contingency-awareness. Points above the line represent participants whose accuracy was greater for valid trials than invalid trials.

cueing effect is primarily driven by those participants shown above the diagonal. Because both contingency-aware and contingency-unaware participants are similarly distributed above and below the diagonal, this plot illustrates that participants who explicitly identified the cueing contingency in the post-test questionnaire were individually no more likely to show a valid cueing effect than those who did not.

## Discussion

In this experiment, I investigated the hypothesis that multisensory integration occurs only when cue-relevant information in the influencing modality is explicitly attended. Previously, I had found that perceptually aware participants were not sensitive to spatial cueing by cross-modal congruence when using emotional face–voice cues (*Experiment 1*). Here, I replicated that finding for participants who received no instructions regarding the emotional voice cues (no-report condition), while obtaining the previously anticipated spatial-cueing effect for participants that were instructed to attend to and report the vocal prosody (with-report condition), thus supporting the hypothesis.

Although *Experiment 2* did not seek to test whether audiovisual integration can occur in the absence of visual awareness, these findings have implications for attempts to answer that question using the congruence-cued spatial attention task. Specifically, these findings provide evidence that spatial attention can be cued by congruent information in two different sensory domains—at least in the case of full perceptual awareness. If this were not the case, that would suggest either that the previous findings reported by Palmer and Ramsey (2012) were spurious, or that those findings do not assess multisensory integration. So while the current findings cannot answer the question directly, they are consistent with the hypothesis that the congruence-cued spatial attention task validly assesses multisensory integration—though other possible explanations for the observed cueing effects have not been tested. A more comprehensive investigation of the task validity would therefore seek to rule out alternative explanations for the observed effects.

Assuming that the task does validly assess multisensory integration, the finding that cross-modal congruence can effectively cue spatial attention suggests a way forward for testing the original hypothesis (i.e., that multisensory integration can occur for unconscious perceptual information). The next attempt at replicating Palmer and Ramsey (2012) should combine the updated awareness manipulation used in *Experiment 1* (including the use of both subjective and objective awareness checks, and the inclusion of a perceptually aware control group), with the emotion-discrimination task introduced in the with-report condition of *Experiment 2*. Based on the results reported in this thesis, I would predict that a significant spatial-cueing effect would be observed for aware participants in the proposed experiment, replicating the findings for with-report participants in *Experiment 2*. If unaware participants in the proposed experiment *also* showed a significant cueing effect, after controlling for subjective breakthrough, that would replicate the original authors' findings of spatial cueing without visual awareness—and support the hypothesis of unconscious multisensory integration.

Finally, this experiment rules out the possibility that aware participants may only respond to congruence cues when they have explicitly identified the cueing contingency. With-report and no-report groups did not differ in reported contingency awareness. Additionally, participants that did identify the cueing contingency, collapsing across emotion-discrimination conditions, did not show a significant spatial-cueing effect.

## General Discussion

In this thesis, I first asked whether multisensory integration of facial and vocal expressions of emotion occurred in the absence of visual awareness. In *Experiment 1*, I compared the target-discrimination accuracy[7] of visually aware and unaware participants for an adapted version of the congruence-cued spatial attention task (Palmer & Ramsey, 2012). The task involved the presentation of valid spatial cues, defined by the emotional congruence of facial expressions and vocal prosody (i.e., *cross-modal congruence*), for 80% trials and invalid spatial cues for the remaining 20% of trials. Following the logic of the classic spatial cueing paradigm (Posner, 1980), an accuracy advantage for valid over invalid trials would suggest that participants were using cross-modal emotional congruence to allocate spatial attention to the cued location. According to Palmer and Ramsey (2012), such an effect would also imply multisensory integration for conscious auditory and unconscious visual information, as it was assumed that cross-modal congruence could only be identified by a process of integration. However, I found no evidence of spatial cueing for either aware or unaware participants using the adapted paradigm. Thus, my findings did not support the hypothesis that multisensory integration occurs in the absence of perceptual awareness.

In *Experiment 2*, I addressed a concern that the null results of the previous experiment may be due to procedural changes that I had made while adapting the task. Specifically, participants in the original study repeated the voice cues aloud at the end of each trial (Palmer & Ramsey, 2012), which ensured that they explicitly attended to the syllables being spoken—and therefore, to the cue-relevant features of the stimuli presented in the *influencing domain*. In contrast, participants in *Experiment 1* were not asked to attend to the emotional voice cues or make any explicit judgement or report about them. Thus, for *Experiment 2*, I hypothesised that cross-modal congruence may only cue spatial attention if participants attend to the congruence-relevant features of the voices. Participants in this experiment viewed all cue and target stimuli without visual masking (as per the aware condition in *Experiment 1*). Half of the participants (with-report group) identified the target location/orientation *and* emotional prosody on every

[7] While I also recorded and analysed response time data, this was intended to support the primary analysis of accuracy scores, given that the task design was optimised for accuracy. In this chapter, I focus on the findings for accuracy, as these are the most pertinent data for the purposes of hypothesis testing and interpreting the results.

trial; while the remaining participants (no-report group) identified only the target location/orientation. Consistent with the hypothesis, I found a significant spatial-cueing effect for participants in the with-report group only. These findings suggest that participants must attend to the cue-relevant features of the cue stimulus presented in the influencing modality (i.e., emotional prosody of the voice cue) in order to use the audiovisual cue to allocate spatial attention. Furthermore, they suggest that the emotion-discrimination judgement (or similar, depending on the chosen cueing stimuli) is a critical feature of the congruence-cued spatial attention task. Thus, its omission from the procedure in *Experiment 1* likely accounts for the lack of significant cueing effects.

While I had initially set out to investigate whether previous findings of unconscious multisensory integration (Faivre et al., 2014; Palmer & Ramsey, 2012) would extend to the integration of facial and vocal emotional expressions, my findings do not clearly answer that question. Instead, it became clear through this investigation that we require a better understanding of the tasks that are used to infer unconscious multisensory integration.

In order to investigate unconscious multisensory integration using behavioural tasks, researchers have had to develop indirect means of probing integration. Traditionally, behavioural measures of multisensory integration have relied on participants' reports of subjective perceptual experiences—but such reports are not possible for the integration of unconscious stimuli, as unaware participants have no objective experience to report. Thus, researchers have had to infer the occurrence of multisensory integration indirectly, using downstream effects on other cognitive processes, such as the allocation of spatial attention. For this inference to be valid, we must understand what cognitive processes these tasks actually assess. Specifically, we should be able to say whether such tasks genuinely assess multisensory integration. To explore these issues, I have used this chapter to discuss the implications of my experiments for the findings of Palmer and Ramsey (2012), for integration and awareness more generally, and for the theoretical models of consciousness that motivate much of this research in the first place.

## Does Unconscious Multisensory Integration Occur?

In attempting to extend the unconscious multisensory integration effects reported by Palmer and Ramsey (2012) to the audiovisual integration of emotional expressions, I made a number of methodological improvements to the congruence-cued spatial attention task. Two changes, in particular, were adopted to allow for easier interpretation of any potential spatial-cueing effects. First, the inclusion of trial-by-trial subjective awareness probes and a post-test objective awareness control task, allowed me to control for subjective breakthrough during continuous flash suppression (CFS; Tsuchiya & Koch, 2005) and ensure that unaware-group participants remained genuinely unaware of the masked emotional faces. Second, the inclusion of a visually aware control group allowed me to compare spatial cueing with and without visual awareness—and, incidentally, to test whether spatial cueing occurred for aware participants, as implicitly predicted by the hypothesis that the congruence-cued spatial attention task probes multisensory integration.

Using the adapted spatial attention task, I found no evidence of spatial cueing for either aware or unaware participants in *Experiment 1* (the only test of unconscious multisensory integration included in this thesis). Assuming that spatial cueing by cross-modal congruence is possible, and genuinely implies multisensory integration, we should expect aware participants to allocate spatial attention in response informative cues regardless of whether integration occurs unconsciously. As such, the lack of spatial cueing in both awareness conditions was initially surprising, however the results of *Experiment 2* suggest that having participants explicitly attend to the cue-relevant features of the emotional voice is a critical feature of the task. Thus, the null results of *Experiment 1* are most likely due to the omission of the emotion-discrimination task and should not be taken as evidence against the hypothesis of unconscious multisensory integration. A more valid test for unconscious multisensory integration would therefore include the emotion-discrimination task (from *Experiment 2*) alongside the awareness manipulation and suppression checks (from *Experiment 1*) for both aware and unaware participants groups. Unfortunately, it was not possible to complete this additional study within the constraints of a one-year Master's thesis.

If the proposed experiment was undertaken, we should expect to find a significant spatial-cueing effect for aware participants, as observed in *Experiment 2* and predicted by the hypothesis that congruence-cueing implies multisensory integration. The critical test for the hypothesis of unconscious multisensory integration would then be whether a significant spatial-cueing effect was also observed for the unaware condition, after adequately controlling for subjective awareness. If spatial cueing *persisted* in the absence of awareness, that would provide a conceptual replication of Palmer and Ramsey (2012), extension of their findings to emotional face–voice integration, and stronger evidence of multisensory integration in the absence of visual awareness. Alternatively, if spatial cueing occurred *only* in the case of visual awareness, that would provide evidence against unconscious multisensory integration—at least for the emotional qualities of faces and voices.

## Does Cross-modal Congruence Cueing Imply Multisensory Integration?

Clearly, to investigate unconscious multisensory integration we need some indirect measure of integration. To this end, researchers have adapted a number of tasks to use cross-modal congruence as a cue for some behavioural task, such as the allocation of spatial attention (e.g., Palmer & Ramsey, 2012) or for later congruence judgements (e.g., Faivre et al., 2014). Yet, if such tasks are to provide compelling evidence of unconscious multisensory integration, it must be empirically shown that they do, in fact, assess the multisensory integration.

In the case of the congruence-cued spatial attention task, it is not yet clear that the task does assess multisensory integration. While some prior research has investigated cross-modal attention effects, that work has focussed on whether unimodal perceptual cues can cue attention in another unimodal perceptual domain (e.g., Butter et al., 1989; Spence et al., 1998), or whether the conjunction of multimodal stimuli can exogenously capture attention (e.g., Spence & Santangelo, 2009; Van der Burg et al., 2008). As far I was able to determine, the use of cross-modal congruence to informatively cue spatial attention has only been attempted by Palmer and Ramsey (2012), and now by me. As such, there is no prior evidence to suggest that the task validly probes multisensory integration. Thus, even if spatial attention can be cued by the congruence of conscious auditory and unconscious visual information, it is not necessarily true that that would imply multisensory integration.

The assumption that congruence cueing implies multisensory integration has thus far been justified by the observation that congruence-based cues are only informative if auditory and visual cues are processed in conjunction (e.g., Faivre et al., 2014; Palmer & Ramsey, 2012). In the case of emotional congruence, there can be no congruent facial expression—and thus no cue—in the absence of vocal prosody. As such, cueing by cross-modal congruence implies that participants have processed the information in each modality in conjunction, in order to determine their emotional or temporal congruence, rather than separately. However, it is a leap to assume that processing cross-modal information in conjunction implies multisensory integration.

Critically, the assumption that cross-modal conjunctions are processed by multisensory integration implies that *any* multi-modal perceptual information processed together must be integrated. Yet, this suggestion is at odds with the perceptual literature on multisensory integration, which has investigated integration in very specific conjunctions of stimuli (e.g., De Gelder & Vroomen, 2000; McGurk & MacDonald, 1976). *Multisensory integration*, in this literature, usually describes the processing of particular co-occurring stimuli that are *bound* together to produce a unified percept (Treisman, 1996). Not all co-occurring stimuli meet this criterion, suggesting that not all conjunctions are processed via multisensory integration. An alternative cognitive process that does rely on the conjunction of stimuli, sometimes in multiple modalities, is *associative learning*. Importantly, associative learning can occur even for stimuli that are not known to produce multisensory integration effects, as in classical conditioning (Pavlov, 1927). Thus, while the use of cross-modal congruence *may* rely on the processes of multisensory integration, it may equally rely on other cognitive processes, such as associative learning.

The underlying mechanism involved in congruence-cueing tasks must therefore be investigated empirically, before we can confidently suggest that such cueing effects imply multisensory integration. While *Experiment 2* provides some support for the hypothesis that the congruence-cued spatial attention task assesses multisensory integration (i.e., by verifying the positive prediction that perceptually aware participants should be cued by emotional face–voice congruence), neither of the current experiments were intended to contrast that hypothesis with alternative explanations. Stronger support for the hypothesis might be provided by attempts to rule out other possible mechanisms, such as associative learning.

Here, I provide to proposals for testing the multisensory-integration hypothesis. First, the hypothesis that multisensory integration is responsible for congruence cueing implies that cueing effects should be observed for cue stimuli that are readily integrated (e.g., audiovisual speech events, audiovisual emotional expressions), but not for stimuli that are not.

In *Experiment 2*, I tested the positive prediction by extending the congruence-cued spatial attention task to emotional face and voice cues, which are already known to produce aware multisensory integration effects (De Gelder & Vroomen, 2000). Yet, this test is not particularly informative, because the associative-learning hypothesis implies the same prediction that spatial cueing should occur.

The more interesting test, then, is of the negative prediction that spatial cueing should not occur in the absence of multisensory integration. This could be tested in a similar fashion by replacing the emotional face/voice cues for another set of audiovisual stimuli with arbitrary, rather than canonical, congruence relationships. For instance, if the informative spatial cues were defined by the arbitrary congruence relationship that happy voices match images of houses, while sad voices match images of cars, there should be no reason to assume that participants would automatically integrate these stimuli. In fact, even though the images and voices would co-occur, participants would have no reason to integrate happy voices with houses, rather than cars, until *after* they had learned the cue (at least implicitly). Under these conditions, the multisensory-integration and associative-learning hypotheses imply different predictions. If multisensory integration is responsible for congruence cueing, the participants in the proposed experiment should *not* be cued to the most likely target location, because they cannot integrate the audiovisual cues according to the arbitrary "congruence" relationships. Conversely, if associative learning is responsible for congruence cueing, participants *should* be cued to the most likely target location, because they should be able to learn from the informative conjunction of auditory and visual cues regardless of whether their features are integrated.

My second proposal is that the two hypotheses could be placed in competition by varying the stimulus-onset asynchrony (SOA) for auditory and visual cues. In this thesis, as in Palmer and Ramsey (2012), the face and voice cues were presented together in order to allow for multisensory integration. However, because multisensory integration only occurs when stimuli appear in close temporal proximity (Koelewijn et al., 2010), the multisensory-integration hypothesis suggests that spatial cueing should *not* occur if faces and voices were presented consecutively rather than concurrently. Conversely, associative learning does not require such strict temporal alignment, thus the associative-learning hypothesis suggests that spatial cueing should persist.

Either of the proposed experiments could—potentially—support the conclusion that the congruence-cued spatial attention task validly assesses multisensory integration. For now though, in the absence of empirical evidence for this conclusion, we cannot assume that the congruence cueing of spatial attention implies multisensory integration. Only once this assumption has been empirically supported, can we confidently use the congruence-cued spatial attention task to investigate unconscious multisensory integration.

## What is the Relationship between Multisensory Integration and Consciousness?

A final question that was raised in the *Introduction*, asked what relationship exists between integration and conscious awareness. While various possible relationships have been proposed, there are two leading hypotheses that motivate the investigation of unconscious multisensory integration. One hypothesis suggests that integration is antecedent to awareness, and that the cognitive processes involved in integration—including multisensory integration—are necessary prerequisites to awareness (e.g., Kanwisher, 2001). On this view, multisensory integration occurs prior to awareness, and thus should occur regardless of whether the integrated perceptual information is later experienced. Thus, unconscious multisensory integration would support this hypothesis. Alternatively, conscious awareness may be necessary for integration, as suggested by global workspace theories (e.g., Baars, 1999; Baars, 2002; Changeux & Dehaene, 2008; Dehaene & Changeux, 2003, 2005, 2011; Dehaene et al., 2006; Dehaene et al., 2014; Dehaene et al., 1998; Dehaene & Naccache, 2001; Sergent & Dehaene, 2004). On this account, awareness arises when information is globally broadcast between functionally and spatially distinct cognitive modules. If information is not globally broadcast, then it is not available for integration. Thus, a typical prediction of this hypothesis is that multisensory integration should not occur in the absence of awareness.

Yet these standard predictions may not adequately capture the theories from which they are derived. Although Palmer and Ramsey (2012) reported a potential unconscious multisensory integration effects, they plausibly argued that their findings are actually consistent with global workspace theories of consciousness. Specifically, they have suggested that unconscious perceptual information may be restricted to unimodal cognitive modules—as global workspace theories suggest—but could still be integrated within the particular modality in which it arises. On this account, unconsciously perceived visual information (e.g., visual speech events) would not be broadcast, and thus would only be available to unimodal visual processes; whereas consciously perceived auditory information (e.g., auditory speech events) would be globally broadcast and be available for integration outside of the unimodal auditory networks. Thus, both visual and auditory information would be available to unimodal visual processes, allowing for the possibility of unconscious integration within the visual domain. For Palmer and Ramsey (2012) this explanation accounts for both their reported spatial attention effects (which were assessed in the visual domain) and their failure to influence subjective auditory experiences with the same stimuli (which was assessed in the auditory domain). Thus, they argued that unconscious multisensory integration is consistent with global workspace theories, provided that the stimuli in the influencing modality are consciously perceived.

A consequence of this argument is that a successful replication of Palmer and Ramsey (2012) would be compatible with both global-workspace and integration-first models of consciousness. Thus, the more important question for understanding the relationship between awareness and integration is whether an unconscious stimulus is available for integration *outside of* its immediate unimodal domain.

The current evidence relevant to this revised question is limited and inconclusive. The failure to obtain a McGurk illusion with masked visual-speech videos (Palmer & Ramsey, 2012) argues against this kind of unconscious multisensory integration. However, Faivre et al. (2014) have reported unconscious multisensory integration effects that seem incompatible with global workspace theories. Their study presented a response priming task, in which participants saw a written numeral (e.g., *8*) and heard a concurrent number word (e.g., *two*), followed by a written letter (e.g., *m*) and spoken letter name (e.g., *em*). Participants were instructed to report the congruence or incongruence of the audiovisual letter pair only. If participants responded faster to letter pairs that shared a congruence relationship with the preceding number pair (i.e., both congruent or both incongruent), than to those that did not, the experimenters could conclude that their responses were primed by the preceding cue (Kiesel, Kunde, & Hoffmann, 2007). In three experiments, Faivre et al. (2014) masked the written numeral only, the spoken number only, and both the written and spoken number from conscious awareness. They reported significant response priming effects under all three masking conditions, suggesting that multisensory integration for audiovisual cues occurrs even in the complete absence of perceptual awareness.

Thus, while both papers report unconscious multisensory integration effects, the findings of Palmer and Ramsey (2012) provide support for global workspace theories of consciousness, while those of Faivre et al. (2014) support integration-first models of consciousness. Critically though, the latter findings are also subject to the caveats I have identified for the congruence-cued spatial attention task: they must be replicable, demonstrably occur in the absence of perceptual awareness, and be shown to genuinely assess multisensory integration. If, and only if, these requirements are met, would a conclusion of unconscious multisensory integration be supported.

## Conclusions

The major contributions of this thesis are the methodological improvements that I have made to the congruence-cued spatial attention paradigm, and the proposal that any indirect method of assessing multisensory integration must be empirically validated. The key methodological changes, from a theoretical perspective, were the inclusion of subjective and objective suppression checks, the inclusion of a perceptually aware control group, and the extension of the task to another class of multimodal stimuli.

Suppression checks included trial-by-trial self-reports for subjective awareness and a post-test control task that assessed the reliability of each participant's self-reports. These checks improve on the original paradigm (Palmer & Ramsey, 2012) by providing a more robust check for subjective awareness, and by allowing for the exclusion of individual breakthrough trials for participants that remain mostly unaware of the masked stimuli. Thus, data obtained for unaware-group participants using these suppression checks are far less likely to be influenced by breakthrough trials, and therefore represent a more genuine state of unawareness.

The inclusion of an aware condition is also important for interpreting the results of a congruence-cued spatial attention study. In order to conclude that a known cognitive process persists in the absence of perceptual awareness, it is necessary to show that the same task produces the expected effect in both aware and unaware conditions. The hypothesis that the congruence-cued spatial attention task assesses multisensory integration, for instance, leads to the prediction that spatial cueing should occur for consciously perceived audiovisual stimuli (regardless of whether unconscious multisensory integration occurs). By presenting cues without visual masking, I was able to show that spatial-cueing effects are obtainable for aware participants (*Experiment 2*).

Finally, by presenting a novel set of perceptual stimuli (i.e., emotional faces and voices), I was able to extend the reported findings to another class of multisensory integration (at least for aware participants, as reported in *Experiment 2*). If the congruence-cued spatial attention task genuinely assesses multisensory integration, then we should expect that any cue set, for which multisensory integration effects are known to occur, should produce significant spatial-cueing effects. Thus, the significant cueing effect observed in *Experiment 2* is consistent with the hypothesis that the task genuinely assesses the cognitive processes involved in multisensory integration (though the thesis as a whole did not pursue this question further by testing other possible explanations for the effect).

In this thesis, I found no support for spatial cueing by audiovisual congruence in the absence of visual awareness (*Experiment 1*), but did observe spatial cueing for perceptually aware participants (*Experiment 2*). While these findings suggest that cross-modal congruence can cue spatial attention—at least when consciously perceived—I have suggested that there is insufficient evidence to conclude that such an effect implies multisensory integration.

Instead, I have suggested that the investigation of unconscious multisensory integration faces three specific challenges that must be met in order to adequately test the current theories of consciousness and integration. First, researchers that develop tasks in which participants respond to the congruence of unconsciously perceived, cross-modal stimulus pairs cannot simply assume that successful responding implies integration; they must rule out other cognitive explanations for the observed behaviours. Second, for any study that claims to show unconscious multisensory integration to be persuasive, it must rule out partial awareness for the masked stimuli by including suitably robust awareness checks. Third, once suitable tasks and awareness controls are developed, we must test both the positive and negative predictions of the leading theories of consciousness in order to decide between them or further refine them.

**References**

Adolphs, R., Tranel, D., & Damasio, A. R. (2003). Dissociable neural systems for recognizing emotions. *Brain and cognition, 52*(1), 61-69. doi:https://doi.org/10.1016/S0278-2626(03)00009-5

Almeida, J., Mahon, B. Z., & Caramazza, A. (2010). The role of the dorsal visual processing stream in tool identification. *Psychological Science, 21*(6), 772-778. doi:https://doi.org/10.1177/0956797610371343

Almeida, J., Mahon, B. Z., Nakayama, K., & Caramazza, A. (2008). Unconscious processing dissociates along categorical lines. *Proceedings of the National Academy of Sciences, 105*(39), 15214-15218. doi:https://doi.org/10.1073/pnas.0805867105

Almeida, J., Pajtas, P. E., Mahon, B. Z., Nakayama, K., & Caramazza, A. (2013). Affect of the unconscious: visually suppressed angry faces modulate our decisions. *Cognitive, Affective, & Behavioral Neuroscience, 13*(1), 94-101. doi:https://doi.org/10.3758/s13415-012-0133-7

Alsius, A., & Munhall, K. G. (2013). Detection of audiovisual speech correspondences without visual awareness. *Psychological Science, 24*(4), 423-431. doi:https://doi.org/10.1177/0956797612457378

Baars, B. J. (1999). Attention versus consciousness in the visual brain: differences in conception, phenomenology, behavior, neuroanatomy, and physiology. *The Journal of general psychology, 126*(3), 224-233. doi:https://doi.org/10.1080/00221309909595364

Baars, B. J. (2002). The conscious access hypothesis: origins and recent evidence. *Trends in Cognitive Sciences, 6*(1), 47-52. doi:https://doi.org/10.1016/s1364-6613(00)01819-2

Balduzzi, D., & Tononi, G. (2008). Integrated information in discrete dynamical systems: motivation and theoretical framework. *PLoS computational biology, 4*(6), e1000091. doi:https://doi.org/10.1371/journal.pcbi.1000091

Barraclough, N. E., Xiao, D., Baker, C. I., Oram, M. W., & Perrett, D. I. (2005). Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. *Journal of Cognitive Neuroscience, 17*(3), 377-391. doi:https://doi.org/10.1162/0898929053279586

Biderman, N., & Mudrik, L. (2017). Evidence for Implicit—But Not Unconscious—Processing of Object-Scene Relations. *Psychological Science*, 0956797617735745. doi:https://doi.org/10.1177/0956797617735745

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews Neuroscience, 3*(1), 13. doi:https://doi.org/10.1038/nrn701

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision, 10*, 433-436. doi:https://doi.org/10.1163/156856897x00357

Breitmeyer, B. G. (2015). Psychophysical "blinding" methods reveal a functional hierarchy of unconscious visual processing. *Consciousness and Cognition, 35*, 234-250. doi:https://doi.org/10.1016/j.concog.2015.01.012

Bushara, K. O., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., & Hallett, M. (2003). Neural correlates of cross-modal binding. *Nature neuroscience, 6*(2), 190. doi:https://doi.org/10.1038/nn993

Butter, C. M., Buchtel, H. A., & Santucci, R. (1989). Spatial attentional shifts: Further evidence for the role of polysensory mechanisms using visual and tactile stimuli. *Neuropsychologia, 27*(10), 1231-1240. doi:https://doi.org/10.1016/0028-3932(89)90035-3

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

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current biology, 10*(11), 649-657. doi:https://doi.org/10.1016/s0960-9822(00)00513-3

Calvert, G. A., & Thesen, T. (2004). Multisensory integration: methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris, 98*(1-3), 191-205. doi:https://doi.org/10.1016/j.jphysparis.2004.03.018

Carmel, D., Arcaro, M., Kastner, S., & Hasson, U. (2010). How to create and use binocular rivalry. *Journal of visualized experiments: JoVE*(45). doi:https://doi.org/10.3791/2030

Cerullo, M. A. (2015). The problem with phi: a critique of integrated information theory. *PLoS computational biology, 11*(9), e1004286. doi:https://doi.org/10.1371/journal.pcbi.1004286

Changeux, J.-P., & Dehaene, S. (2008). The neuronal workspace model: conscious processing and learning. In *Learning and Memory: A Comprehensive Reference* (pp. 729-757).

Cheesman, J., & Merikle, P. M. (1986). Distinguishing conscious from unconscious perceptual processes. *Canadian Journal of Psychology/Revue canadienne de psychologie, 40*(4), 343. doi:https://doi.org/10.1037/h0080103

Chen, Y.-C., & Spence, C. (2010). When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures. *Cognition, 114*(3), 389-404. doi:https://doi.org/10.1016/j.cognition.2009.10.012

Chen, Y.-C., & Spence, C. (2011). Crossmodal semantic priming by naturalistic sounds and spoken words enhances visual sensitivity. *Journal of Experimental Psychology: Human Perception and Performance, 37*(5), 1554. doi:https://doi.org/10.1037/a0024329

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The journal of the acoustical society of america, 25*(5), 975-979. doi:https://doi.org/10.1121/1.1907229

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain research, 1242*, 126-135. doi:https://doi.org/10.1016/j.brainres.2008.04.023

De Gelder, B., Böcker, K. B., Tuomainen, J., Hensen, M., & Vroomen, J. (1999). The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses. *Neuroscience letters, 260*(2), 133-136. doi:https://doi.org/10.1016/s0304-3940(98)00963-x

De Gelder, B., Pourtois, G., & Weiskrantz, L. (2002). Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures. *Proceedings of the National Academy of Sciences, 99*(6), 4121-4126. doi:https://doi.org/10.1073/pnas.062018499

De Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition & Emotion, 14*(3), 289-311. doi:https://doi.org/10.1007/978-1-4614-3585-3_13

De Gelder, B., Vroomen, J., Pourtois, G., & Weiskrantz, L. (1999). Non-conscious recognition of affect in the absence of striate cortex. *Neuroreport, 10*(18), 3759-3763. doi:https://doi.org/10.1097/00001756-199912160-00007

Dehaene, S., & Changeux, J.-P. (2003). Neural mechanisms for access to consciousness. *The cognitive neurosciences III*.

Dehaene, S., & Changeux, J.-P. (2005). Ongoing spontaneous activity controls access to consciousness: a neuronal model for inattentional blindness. *PLoS biology, 3*(5), e141. doi:https://doi.org/10.1371/journal.pbio.0030141

Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron, 70*(2), 200-227. doi:https://doi.org/10.1016/j.neuron.2011.03.018

Dehaene, S., Changeux, J.-P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: a testable taxonomy. *Trends in Cognitive Sciences, 10*(5), 204-211. doi:https://doi.org/10.1016/j.tics.2006.03.007

Dehaene, S., Charles, L., King, J.-R., & Marti, S. (2014). Toward a computational theory of conscious processing. *Current opinion in neurobiology, 25,* 76-84. doi:https://doi.org/10.1016/j.conb.2013.12.005

Dehaene, S., Kerszberg, M., & Changeux, J.-P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National Academy of Sciences, 95*(24), 14529-14534. doi:https://doi.org/10.1073/pnas.95.24.14529

Dehaene, S., & Naccache, L. (2001). Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition, 79*(1-2), 1-37. doi:https://doi.org/10.1016/s0010-0277(00)00123-2

Diano, M., Celeghin, A., Bagnis, A., & Tamietto, M. (2017). Amygdala response to emotional stimuli without awareness: facts and interpretations. *Frontiers in psychology, 7,* 2029. doi:https://doi.org/10.3389/fpsyg.2016.02029

Dimberg, U., Thunberg, M., & Elmehed, K. (2000). Unconscious facial reactions to emotional facial expressions. *Psychological Science, 11*(1), 86-89. doi:https://doi.org/10.1111/1467-9280.00221

Driver, J., & Spence, C. (2000). Multisensory perception: beyond modularity and convergence. *Current biology, 10*(20), R731-R735. doi:https://doi.org/10.1016/s0960-9822(00)00740-5

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion, 6*(3-4), 169-200.

Engel, A. K., Fries, P., König, P., Brecht, M., & Singer, W. (1999). Temporal binding, binocular rivalry, and consciousness. *Consciousness and Cognition, 8*(2), 128-151. doi:https://doi.org/10.1006/ccog.1999.0389

Esteves, F., Parra, C., Dimberg, U., & Öhman, A. (1994). Nonconscious associative learning: Pavlovian conditioning of skin conductance responses to masked fear-relevant facial stimuli. *Psychophysiology, 31*(4), 375-385. doi:https://doi.org/10.1111/j.1469-8986.1994.tb02446.x

Faivre, N., Berthet, V., & Kouider, S. (2012). Nonconscious influences from emotional faces: a comparison of visual crowding, masking, and continuous flash suppression. *Frontiers in psychology, 3*, 129. doi:https://doi.org/10.3389/fpsyg.2012.00129

Faivre, N., Mudrik, L., Schwartz, N., & Koch, C. (2014). Multisensory integration in complete unawareness: evidence from audiovisual congruency priming. *Psychological Science, 25*(11), 2006-2016. doi:https://doi.org/10.1177/0956797614547916

Fang, F., & He, S. (2005). Cortical responses to invisible objects in the human dorsal and ventral pathways. *Nature neuroscience, 8*(10), 1380. doi:https://doi.org/10.1038/nn1537

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods, 39*(2), 175-191. doi:https://doi.org/10.3758/bf03193146

Föcker, J., Gondan, M., & Röder, B. (2011). Preattentive processing of audio-visual emotional signals. *Acta psychologica, 137*(1), 36-47. doi:https://doi.org/10.1016/j.actpsy.2011.02.004

Gerdes, A., Wieser, M. J., & Alpers, G. W. (2014). Emotional pictures and sounds: a review of multimodal interactions of emotion cues in multiple domains. *Frontiers in psychology, 5*, 1351. doi:https://doi.org/10.3389/fpsyg.2014.01351

Gray, J., & Wedderburn, A. (1960). Shorter articles and notes grouping strategies with simultaneous stimuli. *Quarterly Journal of Experimental Psychology, 12*(3), 180-184. doi:https://doi.org/10.1080/17470216008416722

Hedger, N., Adams, W. J., & Garner, M. (2015). Autonomic arousal and attentional orienting to visual threat are predicted by awareness. *Journal of Experimental Psychology: Human Perception and Performance, 41*(3), 798. doi:https://doi.org/10.1037/xhp0000051

Holender, D. (1986). Semantic activation without conscious identification in dichotic listening, parafoveal vision, and visual masking: A survey and appraisal.

*Behavioral and brain Sciences, 9*(1), 1-23. doi:https://doi.org/10.1017/s0140525x00021269

Hubel, D. H., & Wiesel, T. N. (1959). Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology, 148*(3), 574-591. doi:https://doi.org/10.1113/jphysiol.1959.sp006308

Hubel, D. H., & Wiesel, T. N. (1977). Ferrier lecture-Functional architecture of macaque monkey visual cortex. *Proc. R. Soc. Lond. B, 198*(1130), 1-59. doi:https://doi.org/10.1098/rspb.1977.0085

Izatt, G., Dubois, J., Faivre, N., & Koch, C. (2014). A direct comparison of unconscious face processing under masking and interocular suppression. *Frontiers in psychology, 5*, 659. doi:https://doi.org/10.3389/fpsyg.2014.00659

Jiang, Y., Costello, P., Fang, F., Huang, M., & He, S. (2006). A gender-and sexual orientation-dependent spatial attentional effect of invisible images. *Proceedings of the National Academy of Sciences, 103*(45), 17048-17052. doi:https://doi.org/10.1073/pnas.0605678103

Kanwisher, N. (2001). Neural events and perceptual awareness. *Cognition, 79*(1-2), 89-113. doi:https://doi.org/10.1016/s0010-0277(00)00125-6

Kiesel, A., Kunde, W., & Hoffmann, J. (2007). Mechanisms of subliminal response priming. *Advances in Cognitive Psychology, 3*(1-2), 307. doi:https://doi.org/10.2478/v10053-008-0032-1

Koelewijn, T., Bronkhorst, A., & Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: A review of audiovisual studies. *Acta psychologica, 134*(3), 372-384. doi:https://doi.org/10.1016/j.actpsy.2010.03.010

Kouider, S., & Dehaene, S. (2007). Levels of processing during non-conscious perception: a critical review of visual masking. *Philosophical Transactions of the Royal Society of London B: Biological Sciences, 362*(1481), 857-875. doi:https://doi.org/10.1098/rstb.2007.2093

Lakens, D., & Evers, E. R. (2014). Sailing from the seas of chaos into the corridor of stability: Practical recommendations to increase the informational value of studies. *Perspectives on Psychological Science, 9*(3), 278-292. doi:https://doi.org/10.1177%2F1745691614528520

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8*. University of Florida, Gainesville, FL.

Leo, F., Bertini, C., Di Pellegrino, G., & Làdavas, E. (2008). Multisensory integration for orienting responses in humans requires the activation of the superior colliculus. *Experimental Brain Research, 186*(1), 67-77. doi:https://doi.org/10.1007/s00221-007-1204-9

Leopold, D. A., & Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry. *Nature, 379*(6565), 549. doi:https://doi.org/10.1038/379549a0

Lerner, Y., Singer, N., Gonen, T., Weintraub, Y., Cohen, O., Rubin, N., . . . Hendler, T. (2012). Feeling without seeing? Engagement of ventral, but not dorsal, amygdala during unaware exposure to emotional faces. *Journal of Cognitive Neuroscience, 24*(3), 531-542. doi:https://doi.org/10.1162/jocn_a_00165

Liddell, B. J., Brown, K. J., Kemp, A. H., Barton, M. J., Das, P., Peduto, A., . . . Williams, L. M. (2005). A direct brainstem–amygdala–cortical 'alarm'system for subliminal signals of fear. *Neuroimage, 24*(1), 235-243. doi:https://doi.org/10.1016/j.neuroimage.2004.08.016

Lin, Z., & He, S. (2009). Seeing the invisible: The scope and limits of unconscious processing in binocular rivalry. *Progress in neurobiology, 87*(4), 195-211. doi:https://doi.org/10.1016/j.pneurobio.2008.09.002

Logothetis, N. K., & Schall, J. D. (1989). Neuronal correlates of subjective visual perception. *Science, 245*(4919), 761-763. doi:https://doi.org/10.1126/science.2772635

Macaluso, E., & Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends in neurosciences, 28*(5), 264-271. doi:https://doi.org/10.1016/j.tins.2005.03.008

Macaluso, E., Frith, C. D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science, 289*(5482), 1206-1208. doi:https://doi.org/10.1126/science.289.5482.1206

Marcel, A. J. (1983). Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive psychology, 15*(2), 238-300. doi:https://doi.org/10.1016/0010-0285(83)90010-5

Massaro, D. W. (1987). *Speech perception by ear and eye*. New York: Psychology Press.

Massaro, D. W., & Egan, P. B. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review, 3*(2), 215-221. doi:https://doi.org/10.3758/bf03212421

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*(5588), 746. doi:https://doi.org/10.1038/264746a0

Meredith, M. A. (2002). On the neuronal basis for multisensory convergence: a brief overview. *Cognitive brain research, 14*(1), 31-40. doi:https://doi.org/10.1016/s0926-6410(02)00059-9

Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of neurophysiology, 56*(3), 640-662. doi:https://doi.org/10.1152/jn.1986.56.3.640

Meredith, M. A., & Stein, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus neurons. *Journal of neurophysiology, 75*(5), 1843-1857. doi:https://doi.org/10.1152/jn.1996.75.5.1843

Mindt, G. (2017). The Problem with the'Information'in Integrated Information Theory. *Journal of Consciousness Studies, 24*(7-8), 130-154. doi:https://doi.org/10.1371/journal.pcbi.1004286

Moors, P., Boelens, D., van Overwalle, J., & Wagemans, J. (2016). Scene integration without awareness: no conclusive evidence for processing scene congruency during continuous flash suppression. *Psychological Science, 27*(7), 945-956. doi:https://doi.org/10.1177/0956797616642525

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *reason, 4*(2), 61-64. doi:https://doi.org/10.20982/tqmp.04.2.p061

Morris, J. S., Öhman, A., & Dolan, R. J. (1999). A subcortical pathway to the right amygdala mediating “unseen” fear. *Proceedings of the National Academy of Sciences, 96*(4), 1680-1685. doi:https://doi.org/10.1073/pnas.96.4.1680

Mudrik, L., Breska, A., Lamy, D., & Deouell, L. Y. (2011). Integration without awareness: expanding the limits of unconscious processing. *Psychological Science, 22*(6), 764-770. doi:https://doi.org/10.1167/11.11.1138

Mudrik, L., Faivre, N., & Koch, C. (2014). Information integration without awareness. *Trends in Cognitive Sciences, 18*(9), 488-496. doi:https://doi.org/10.1016/j.tics.2014.04.009

Mudrik, L., & Koch, C. (2013). Differential processing of invisible congruent and incongruent scenes: A case for unconscious integration. *Journal of Vision, 13*(13), 24-24. doi:https://doi.org/10.1167/13.13.24

Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance, 15*(2), 315. doi:https://doi.org/10.1037//0096-1523.15.2.315

Noesselt, T., Bergmann, D., Hake, M., Heinze, H.-J., & Fendrich, R. (2008). Sound increases the saliency of visual events. *Brain research, 1220*, 157-163. doi:https://doi.org/10.1016/j.brainres.2007.12.060

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: detecting the snake in the grass. *Journal of experimental psychology: general, 130*(3), 466. doi:https://doi.org/10.1037//0096-3445.130.3.466

Palmer, T. D., & Ramsey, A. K. (2012). The function of consciousness in multisensory integration. *Cognition, 125*(3), 353-364. doi:https://doi.org/10.1016/j.cognition.2012.08.003

Pavlov, I. P. (1927). *Conditional reflexes: an investigation of the physiological activity of the cerebral cortex*. Oxford, England: Oxford University Press.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*, *162*(1-2), 8-13. doi:https://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce, J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in neuroinformatics, 2*, 10. doi:https://doi.org/10.3389/neuro.11.010.2008

Pessoa, L. (2005). To what extent are emotional visual stimuli processed without attention and awareness? *Current opinion in neurobiology*, *15*(2), 188-196. doi:https://doi.org/10.1016/j.conb.2005.03.002

Pockett, S. (2014). Problems with theories that equate consciousness with information or information processing. *Frontiers in systems neuroscience, 8*, 225. doi:https://doi.org/10.3389/fnsys.2014.00225

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

Rohenkohl, G., Gould, I. C., Pessoa, J., & Nobre, A. C. (2014). Combining spatial and temporal expectations to improve visual perception. *Journal of Vision, 14*(4), 8-8. doi:https://doi.org/10.1167/14.4.8

Schiller, P. H., Finlay, B. L., & Volman, S. F. (1976). Quantitative studies of single-cell properties in monkey striate cortex. I. Spatiotemporal organization of receptive fields. *Journal of neurophysiology, 39*(6), 1288-1319. doi:https://doi.org/10.1152/jn.1976.39.6.1288

Schirmer, A., & Adolphs, R. (2017). Emotion perception from face, voice, and touch: comparisons and convergence. *Trends in Cognitive Sciences, 21*(3), 216-228. doi:https://doi.org/10.1016/j.tics.2017.01.001

Searle, J. R. (2013). Can information theory explain consciousness. *New York Review of Books, 1*(10).

Senkowski, D., Talsma, D., Grigutsch, M., Herrmann, C. S., & Woldorff, M. G. (2007). Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations. *Neuropsychologia, 45*(3), 561-571. doi:https://doi.org/10.1016/j.neuropsychologia.2006.01.013

Sergent, C., & Dehaene, S. (2004). Neural processes underlying conscious perception: experimental findings and a global neuronal workspace framework. *Journal of Physiology-Paris*, *98*(4-6), 374-384. doi:https://doi.org/10.1016/j.jphysparis.2005.09.006

Sheinberg, D. L., & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy of Sciences, 94*(7), 3408-3413. doi:https://doi.org/10.1073/pnas.94.7.3408

Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, *12*(1), 7-10. doi:https://doi.org/10.1097/00001756-200101220-00009

Soto-Faraco, S., Navarra, J., & Alsius, A. (2004). Assessing automaticity in audiovisual speech integration: evidence from the speeded classification task. *Cognition, 92*(3), B13-B23. doi:https://doi.org/10.1016/j.cognition.2003.10.005

Spence, C., & Driver, J. (2000). Attracting attention to the illusory location of a sound: reflexive crossmodal orienting and ventriloquism. *Neuroreport, 11*(9), 2057-2061. doi:https://doi.org/10.1097/00001756-200006260-00049

Spence, C., Nicholls, M. E., Gillespie, N., & Driver, J. (1998). Cross-modal links in exogenous covert spatial orienting between touch, audition, and vision. *Perception & psychophysics, 60*(4), 544-557. doi:https://doi.org/10.3758/bf03206045

Spence, C., & Santangelo, V. (2009). Capturing spatial attention with multisensory cues: A review. *Hearing research, 258*(1-2), 134-142. doi:https://doi.org/10.1016/j.heares.2009.04.015

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews Neuroscience, 9*(4), 255. doi:https://doi.org/10.1038/nrn2377

Stein, B. E., Stanford, T. R., Wallace, M. T., Vaughan, J. W., & Jiang, W. (2004). Crossmodal spatial interactions in subcortical and cortical circuits. *Crossmodal space and crossmodal attention*, 25-50. doi:https://doi.org/10.1093/acprof:oso/9780198524861.003.0002

Stoerig, P., & Cowey, A. (1992). Wavelength discrimination in blindsight. *Brain, 115*(2), 425-444. doi:https://doi.org/10.1093/brain/115.2.425

Stoerig, P., & Cowey, A. (1995). Visual perception and phenomenal consciousness. *Behavioural brain research, 71*(1-2), 147-156. doi:https://doi.org/10.1016/0166-4328(95)00050-x

Stoerig, P., & Cowey, A. (1997). Blindsight in man and monkey. *Brain: a journal of neurology, 120*(3), 535-559. doi:https://doi.org/10.1093/brain/120.3.535

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. *The journal of the acoustical society of america*, *26*(2), 212-215. doi:https://doi.org/10.1121/1.1907309

Tamietto, M., Castelli, L., Vighetti, S., Perozzo, P., Geminiani, G., Weiskrantz, L., & de Gelder, B. (2009). Unseen facial and bodily expressions trigger fast emotional reactions. *Proceedings of the National Academy of Sciences, 106*(42), 17661-17666. doi:https://doi.org/10.1073/pnas.0908994106

Thurlow, W. R., & Jack, C. E. (1973). Certain determinants of the "ventriloquism effect". *Perceptual and motor skills, 36*(3_suppl), 1171-1184. doi:https://doi.org/10.2466/pms.1973.36.3c.1171

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron, 21*(4), 753-759. doi:https://doi.org/10.1016/s0896-6273(00)80592-9

Tononi, G. (2008). Consciousness as integrated information: a provisional manifesto. *The Biological Bulletin, 215*(3), 216-242. doi:https://doi.org/10.2307/25470707

Tononi, G., & Edelman, G. M. (1998). Consciousness and complexity. *Science, 282*(5395), 1846-1851. doi:https://doi.org/10.1126/science.282.5395.1846

Tononi, G., & Koch, C. (2015). Consciousness: here, there and everywhere? *Phil. Trans. R. Soc. B, 370*(1668), 20140167. doi:https://doi.org/10.1098/rstb.2014.0167

Tooley, M. D., Carmel, D., Chapman, A., & Grimshaw, G. M. (2017). Dissociating the physiological components of unconscious emotional responses. *Neuroscience of Consciousness, 3*(1). doi:https://doi.org/10.1093/nc/nix021

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., . . . Nelson, C. (2009). The NimStim set of facial expressions: judgments from untrained research participants. *Psychiatry research, 168*(3), 242-249. doi:https://doi.org/10.1016/j.psychres.2008.05.006

Treisman, A. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology, 12*(4), 242-248. doi:https://doi.org/10.1080/17470216008416732

Treisman, A. (1996). The binding problem. *Current opinion in neurobiology, 6*(2), 171-178. doi:https://doi.org/10.1016/s0959-4388(96)80070-5

Tsuchiya, N., & Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature neuroscience, 8*(8), 1096. doi:https://doi.org/10.1038/nn1500

Van der Burg, E., Olivers, C. N., Bronkhorst, A. W., & Theeuwes, J. (2008). Pip and pop: nonspatial auditory signals improve spatial visual search. *Journal of Experimental Psychology: Human Perception and Performance, 34*(5), 1053. doi:https://doi.org/10.1037/0096-1523.34.5.1053

Van Opstal, F., Gevers, W., Osman, M., & Verguts, T. (2010). Unconscious task application. *Consciousness and Cognition, 19*(4), 999-1006. doi:https://doi.org/10.1016/j.concog.2010.05.002

Vroomen, J., Bertelson, P., & De Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & psychophysics, 63*(4), 651-659. doi:https://doi.org/10.3758/bf03194427

Wallace, M. T., & Stein, B. E. (1997). Development of multisensory neurons and multisensory integration in cat superior colliculus. *Journal of Neuroscience, 17*(7), 2429-2444. doi:https://doi.org/10.1523/jneurosci.17-07-02429.1997

Wallace, M. T., & Stein, B. E. (2001). Sensory and multisensory responses in the newborn monkey superior colliculus. *Journal of Neuroscience, 21*(22), 8886-8894. doi:https://doi.org/10.1523/jneurosci.21-22-08886.2001

Wambacq, I. J., Shea-Miller, K. J., & Abubakr, A. (2004). Non-voluntary and voluntary processing of emotional prosody: an event-related potentials study. *Neuroreport, 15*(3), 555-559. doi:https://doi.org/10.1097/00001756-200403010-00034

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience, 18*(1), 411-418. doi:https://doi.org/10.1523/jneurosci.18-01-00411.1998

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior research methods, 42*(3), 671-684. doi:https://doi.org/10.1167/10.7.653

Yang, E., Brascamp, J., Kang, M.-S., & Blake, R. (2014). On the use of continuous flash suppression for the study of visual processing outside of awareness. *Frontiers in psychology, 5*, 724. doi:https://doi.org/10.3389/fpsyg.2014.00724

## Appendix

## Preregistration Document

The experimental design and primary analyses of *Experiment 2* were preregistered on AsPredicted (aspredicted.org) on 18 January 2018. A publically available PDF can be accessed at https://aspredicted.org/bx6g5.pdf. The text of the preregistration document is reproduced below:

### Multisensory perceptual awareness - attentional cueing (#7875)

Created: 01/17/2018 11:44 AM (PT)
Public: 04/29/2018 06:26 PM (PT)

Author(s) Daniel Jenkins (Victoria University of Wellington) - daniel.jenkins@vuw.ac.nz
Gina Grimshaw (Victoria University of Wellington) - gina.grimshaw@vuw.ac.nz

**1) Have any data been collected for this study already?**

No, no data have been collected for this study yet.

**2) What's the main question being asked or hypothesis being tested in this study?**

This study investigates the hypothesis that people learn predictive relationships between crossmodal cues only when both modalities are attended.

**3) Describe the key dependent variable(s) specifying how they will be measured.**

Participants will complete a spatial cueing task in which congruence between concurrent auditory and visual cue stimuli predicts the location (left/right) of a subsequent visual target (low-contrast Gabor grating). Cues consist of two photographs of emotional human faces (one happy, one sad) presented either side of fixation, and one voice recording (happy or sad). The emotional valence of the voice matches exactly one visual cue image (congruent image), which predicts the location of the upcoming target. The target location is preceded by the congruent image in 80% of trials (valid cue) and by the incongruent image in 20% (invalid cue). Participants indicate the location (left/right) and rotation (left/right) of each target, with accuracy (proportion correct) and response time (correct responses only) as DVs. Correct trials are those in which the participant accurately reports both location and rotation of the target within a constrained response window (1500 ms). Participants in the attended-emotion condition also indicate whether the voice was "happy" or "sad". Although accuracy and response time will be recorded, these inform exclusion decisions and are not DVs in our primary analyses.

**4) How many and which conditions will participants be assigned to?**

Participants will be assigned to attended-emotion and control conditions by random assignment. Control participants will perform only the target detection task. Attended-emotion participants will perform both the target identification task and the emotion discrimination task, identifying the target first, and then the emotional valence of the voice. A within-subjects variable of cue validity (two levels: valid, invalid) will also be included, with validly cued trials making up 80% of trials in each block, and invalidly cued trials making up the remaining 20%.

**5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.**

Target detection accuracy and response times will be subjected to separate 2 × 2 mixed measures ANOVAs, with a within-subjects variable of cue validity (two levels: valid, invalid) and a between-subjects variable of condition (two levels: emotion-attended, control). Response time for each participant is calculated as the average response time recorded for correct target detection responses only.

**6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.**

Participants whose target detection accuracy does not significantly exceed chance (50%) or differ from ceiling (100%) on a single-sample z-test will be excluded. (Although four responses are possible, and only one is correct, pilot studies suggest participants are near perfect at identifying location, making the task essentially a decision between two rotation options. For this reason, we assume chance performance is 50%, not 25%.) Attended-emotion participants whose emotion discrimination does not significantly exceed chance (50%) on a single-sample z-test will be excluded. No other outlier or exclusion criteria will be applied.

**7) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.**

We will collect data for 48 participants (24 per condition). Sample size is based on a power analysis using conservative estimates of effect sizes reported by Palmer and Ramsey (2012)—the only paper, to our knowledge, that has used cross-modal congruence as a cue in a spatial cueing paradigm.

**8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)**

*N/A*