Crossmodal Identification

Casey O'Callaghan

Washington University in St. Louis

Penultimate version. Please cite final published paper:

Casey O'Callaghan, "Crossmodal identification," In: *Sensory Individuals: Unimodal and Multimodal Perspectives*. Edited by: Aleksandra Mroczko-Wąsowicz and Rick Grush, Oxford University Press. © Oxford University Press 2023, 331-353. DOI: 10.1093/oso/9780198866305.003.0020

Abstract

In crossmodal identification, a subject token identifies an item perceived in one sensory modality with an item perceived in another sensory modality. Does crossmodal identification always occur in cognition, or does crossmodal identification sometimes take place in perception? This paper argues that crossmodal identification occurs in cognition and not perception. Nevertheless, multisensory perception is not unalive to crossmodal identity. Experimental evidence demonstrates that perception is differentially sensitive to the identity of individuals presented to distinct senses. Such sensitivity enhances recognition and improves action. This approach relies on distinguishing crossmodal identification from perceiving crossmodal identity. Perception registers crossmodal identity, but crossmodal identification as such belongs to thought.

Keywords

multisensory perception; object perception; binding; object files

1 Introduction

In crossmodal identification, a subject token identifies an individual thing or feature that is perceived in one sensory modality with an individual thing or feature that is perceived in a distinct sensory modality. You may surmise that the tennis ball you feel in your hand is the tennis ball you see. You may register that you hear and see it bounce. Sometimes, when you perceive something with each of two or more different senses, you can tell it is the same thing. Crossmodal identification takes place just in case a subject, s, perceives an individual, o_1, in a sensory modality, m; s perceives an individual, o_2, in a distinct sensory modality, n; and s token identifies o_1 with o_2.

This paper is about the nature of crossmodal identification. To simplify, it focuses on perceptible individuals and not attributes. It concerns the capacity to token identify an individual object or event that a subject perceives using distinct senses. Focusing on attributes perceived using distinct senses raises thorny additional issues.

What kind of psychological phenomenon is crossmodal identification? Which capacities does it comprise? In particular, does crossmodal identification always implicate thought or cognition, or is crossmodal identification sometimes wholly perceptual?

In the first place, crossmodal identification can be cognitive. One way to token identify an item perceived using distinct senses is to work it out. Using perceptual clues, a subject might infer that what is perceived through each of two separate senses is the same thing. Or a subject might just recognize it to be one thing, with past experience or stored knowledge as a guide. If crossmodal identification relies on inference or recognition, and each of these is extraperceptual, then so too is crossmodal identification.

It is tempting to think crossmodal identification can be perceptual. If so, a subject does not just recognize as one thing or identify in thought what is perceived using each of two distinct senses. Instead, one perceives the sameness, the numerical identity, of an individual which one perceives through distinct senses. Perceptibly, the thing you see is the thing you touch; you perceive the event you hear and the event you see to be one event. If so, one may conclude that crossmodal identification takes place in perception.[1]

So, we have two views about crossmodal identification. According to one, crossmodal identification always takes place in thought or extraperceptual cognition. According to the other, crossmodal identification sometimes occurs in perception. This paper negotiates the merits and drawbacks to each approach, and it offers a third way. It resists the claim that crossmodal identification takes place in perception. Strictly speaking, crossmodal identification is not a perceptual capacity because it is too demanding. Perception does not represent individuals perceived with one sense as being identical to individuals perceived with another sense. However, perception is not unalive to crossmodal identity. Perception in typical human subjects is differentially sensitive to the identity or sameness of individuals presented to distinct senses.

[1]See Bayne (2014); O'Callaghan (2014); Green (2021). Spence and Bayne (2015) are skeptical.

This relies on distinguishing crossmodal identification from perceiving crossmodal identity. Perception registers crossmodal identity but does not represent it as such.

2 Crossmodal identification in thought

The first option says crossmodal identification takes place outside perception. Philosophy has focused on token identification in thought. Testimony rather than night sky viewing taught me Hesperus is Phosphorus. Students understand that saying Hesperus is Hesperus differs in informativeness from saying Hesperus is Phosphorus. In the paradigms, an object is thought of in different ways, on multiple occasions, or using distinct words, and it is recognized, appreciated, or said to be one thing.

Suppose a subject perceives an object by means of one sense and also perceives it by means of another sense, such that one could continue to perceive it independently using each sense. Each way of perceiving enables thoughts about the object. Given differences between senses, a subject may not recognize or appreciate the identity. An uninitiated perceiver could think there are two distinct objects. Nevertheless, an experienced subject facing the scene could think there is one perceptible object. Tradition says the main difference is what initiated and uninitiated subjects think, not what they perceive.

Williamson (1990) suggests a model. In *Identity and Discrimination*, he says to discriminate is to activate knowledge of distinctness. When a subject discriminates *a* from *b*, the subject knows *a* is distinct from *b*, and that knowledge becomes active.

By analogy, identification may be understood as activating knowledge of identity or sameness. When a subject identifies *a* with *b*, the subject knows *a* is numerically identical with *b*, and that knowledge becomes active.

Notice two things. First, if knowledge requires belief and belief is extraperceptual, then both discrimination and identification are cognitive and extraperceptual.[2] Second, the approach just sketched treats identification as factive. Just as one cannot discriminate something from itself, it is not possible to token identify distinct items. This factive requirement may be waived

[2]If perception entails belief, this does not follow. However, if so, the perceptual account of crossmodal identification may collapse into the cognitive account.

to focus on a fallible mental counterpart. The target then is roughly the activation of warranted occurrent belief in identity.

In this picture, crossmodal identification takes place in cognition. A subject perceives an individual object or event with one sense, perceives it with another sense, and activates knowledge or belief that it is one and the same thing. A subject may see a tennis ball, touch that tennis ball, and think the item seen is identical with the item touched.

3 Appreciating identity

Crossmodal identification requires token identifying an item perceived using one sense with an item perceived using another sense. Token identification requires appreciating identity. Typical human subjects can appreciate crossmodal identity in thought. To determine whether crossmodal identification occurs in perception, it helps to say what it is to appreciate identity across senses.

Consider three types of cases. The first evidently involves just one presentation of an object, *a*. For instance, a subject may at once ascribe distinct attributes to an individual, as when observing that a perceptible object is yellow and spherical, or yellow and fuzzy. Using a term in language to refer to one thing on an occasion offers an example, as in, "Hesperus is a planet visible in evening." Regarding it once reflects that one appreciates the identity of that which bears distinct features.

The second involves regarding one thing presented in the same way or manner in distinct episodes as the same. Seeing one tennis ball on two occasions offers an example. You may think the tennis ball is yellow, then later think it is spherical, while thinking of it in the same particular visual way. This reflects a way to appreciate token identity. Recanati (2012) would say it involves a single mental file. Using distinct tokens of one proper name in language offers an example, as in, "Hesperus is visible" and "Hesperus is a planet." Fine (2007) calls this *coordination*, or *representing as the same*. He says:

> An object is represented as the same in a piece of discourse only if no one who understands the discourse can sensibly raise the question of whether it is the same. (40)

It is a constraint that the term "*a*" refers to the same thing in each use in a discourse. That is not true for "*a*" and "*b*." Not appreciating this betrays misunderstanding.

The third involves regarding something presented in differing ways or manners to be one thing. Using distinct Fregean modes of presentation to express an identity offers an example. So, too, can seeing and feeling a tennis ball. You may think what you see is yellow and think what you touch is fuzzy. Vision and touch enable you to think of it in differing ways. Perhaps each gives rise to a distinct mental file. Nonetheless, you may recognize that what you see is what you touch and thus appreciate the identity. In language, distinct proper names can serve to express an identity claim, as in, "Hesperus is Phosphorus." Fine calls this *representing as being the same*, which is incompatible with representing as the same (coordination). Representing as being the same, but not representing as the same, is informative:

> The idea of representing objects *as* the same is to be distinguished from the idea of representing the objects as *being* the same. The sentences "Cicero = Tully" and "Cicero = Cicero" both represent the objects as being the same but only the second represents them as the same. And, in general, one cannot informatively represent objects as being the same compatibly with representing them as the same. (40)

In general, it is informative to be told that *a* is *b* because, typically, it is not a constraint in a discourse that "*a*" and "*b*" refer to one thing.

Crossmodal identification belongs to the third type. Using Fine's terms, for one to token identify an individual perceived by means of distinct senses relies on representing it as being the same.

The argument proceeds by elimination. Crossmodal identification does not just belong to the first type of case, in which an object or event evidently is presented just once, in a given manner, while bearing one or more features (as in, *a is F and G*). In crossmodal identification, a subject perceives an item and its features by means of distinct senses. You see the tennis ball, and you touch the tennis ball. Moreover, these dissociate. You can see the tennis ball without touching it, and you can touch the tennis ball without seeing it. So, sight and touch each present the tennis ball, and neither presentation evidently depends on the other for its existence. Accordingly, to appreciate the identity across senses requires either that the item presented in sight and in touch is represented as the same (*a is F* and *a is G*), or that it is represented as being the same by means of an informative identity ($a = b$).

Crossmodal identification also does not just belong to the second type of case, in which an object or event evidently is presented more than once but as the same, or in the same manner. The appearance of identity across senses can be misleading or mistaken. You might just be wrong that the tennis ball you see is the tennis ball you touch. Fine says representing as the same does not allow such errors:

> Finally, what are in fact two distinct objects can be represented as being the same, as with the sentence "Cicero = Caesar." But two distinct objects cannot ever be represented as the same—or, at least, not without taking two names to be one or committing some other error of this kind. (40)

More to the point, it need not be evident that what you see is what you touch. A subject may fail to discern the identity. And one can sensibly raise the question whether what is seen is what is touched. Indeed, it can be rational for a subject who sees and touches one object to doubt the identity claim or ascribe incompatible features. So the identification is not trivial. It is informative to learn a crossmodal identity.[3]

Crossmodal identification thus belongs to the third type of case, in which an object or event presented with distinct senses in differing ways or manners is taken to be one thing. This means regarding what is perceived with one sense as being the same as what is perceived with another sense. Using Fine's terms, crossmodal identification is akin to representing as being the same.

Identification in such a case requires more than appreciating each of two distinct presentations. To see and to touch one thing does not suffice to appreciate its identity. One must take what is seen and what is touched to be the same. It must be regarded as being a single thing. Fine says:

> A further difference is that only a single sentence (such as "Cicero = Tully" can represent its objects as being the same but two different sentences (e.g., "Cicero is Roman," "Cicero is an orator") can represent their objects as the same. (40)

Maybe no explicit identity expression is required. What Campbell (1997, 59) calls *trading on identity* may suffice. If one appreciates that *a* is *F* and that *b* is *G*, appreciating that something is both *F* and *G* can be evidence one has traded on identity.

[3]For discussion of co-reference in distinct modalities, see Rescorla (2020).

4 Crossmodal identification in perception

A typical subject can think something seen and something touched is one thing. In thought, one can represent as being the same something that is perceived in each of two distinct sensory modalities. A lot of philosophy since Frege addresses how. Thus, assuming common sensibles, perception can activate knowledge of crossmodal identity.

The view that crossmodal identification takes place in perception faces obstacles. There are reasons to doubt that perception represents items perceived with distinct senses *as being the same*. And it is not clear perception appreciates crossmodal identity in the way required for token identification. Perception does not fit the model of crossmodal identification in thought.

Notice first that to perceive numerical identity across senses is not trivial. One cannot perceive the identity of what is seen with what is touched just by independently seeing and touching an item. To perceive crossmodal identity requires not only using two senses in parallel but also perceiving jointly with those senses. It relies on registering individuals, spatiotemporal characteristics, and relations among them across senses. This is a form of ineliminably multisensory perception.

Next consider what crossmodal identification requires. According to the previous section, crossmodal identification requires representing an item perceived using one sense as being identical with an item perceived using another sense. To represent an individual *a* (perceived with *m*) as being the same as an individual *b* (perceived with *n*) means representing *a*, representing *b*, and representing *a* to be identical with *b*. One must represent identity, as such.

It is not obvious contents structured this way figure elsewhere in synchronic perception. To see an object and its mirror image at once is not evidently perceptually to represent what is doubly perceived as being the same. Nor is seeing double perceptually representing as being the same.

How about diachronic reidentification? Consider finding your car in a lot or your eyeglasses in the dark. Clear cases in which a subject appreciates identity are not evidently ones in which the item is perceived in differing manners and represented perceptually as being the same. Instead, in visual or tactual reidentification, given how perception presents the item, it is plausible that what is perceived is represented as the same, or in the same manner. The high-level content, *being the same as*, does not obviously find a place in perception.

This is not conclusive. Consider an example involving perceptual indexicals. Suppose you see your car, see it later, and appreciate the identity. Both the indexical manners and their targets seem to match. But distinct uses of an indexical need not entail representing as the same. One can sensibly ask whether *here* is the same as *here*. Since you appreciate your car's identity, two uses of an indexical are compatible with representing a target as being the same. And this might occur within perception.

Of course, this presumes without argument that visual perception represents the identity as such. Still, even if we find no unisensory example of perceptually representing as being the same, crossmodal perceptual identification could be unique.

A perceptual approach to crossmodal identification faces additional obstacles. Crossmodal identification entails somehow appreciating the identity of what is perceived using distinct senses. However, it is always sensible to ask if what you perceive in one modality is identical to what you perceive in another. A subject rationally can doubt whether what is seen is the same as what is touched. Considering the question is not enough to settle it. Accordingly, to appreciate crossmodal identity requires further considerations beyond just perceiving one item using each of two senses. Crossmodal identification calls for evidence, warrant, or reasons supporting the identity.

It is questionable whether perception provides evidence by representing identity across senses. On one hand, perception may directly present a subject with crossmodal identity as such. If so, perception supplies the needed warrant by revealing crossmodal identity. One could then grasp immediately that a seen item is the same as a touched item. However, it is doubtful that perceiving subjects are directly presented with crossmodal identity *as such*. It is not evident phenomenologically that one perceives something seen and something touched as being numerically the same. Earnestly introspecting subjects may wonder. In contrast with *redness*, *texture*, or *motion*, there is ample room for skepticism (see Spence and Bayne 2015).

On the other hand, perception may present a subject with indirect evidence of crossmodal identity, such as spatiotemporal coincidence and correlation. This offers the support a subject needs to appreciate identity across senses. However, it enfeebles crossmodal identification in perception. Suppose crossmodal identification in thought finds perceptual support from features such as spatiotemporal relations. What does perceiving identity across senses add? Without an answer, crossmodal identification in perception is superfluous.

This challenge, too, is answerable. Multisensory perception may represent as such the identity of an item perceived with each of two senses. However, a subject may fail to appreciate it, or may appreciate it for other reasons. Whether and how a subject appreciates the identity in thought need not couple tightly with the fact that perception reveals or represents it.

Still, if crossmodal identification requires appreciating identity, a further obstacle remains. To perceptually represent something seen as being identical with something touched requires an appreciative response to the identity. Seeing a tennis ball and touching a tennis ball is compatible with seeing and touching distinct tennis balls. Crossmodal identification in perception thus calls for support from further features, such as spatiotemporal coincidence and correlations.

Suppose crossmodal identification requires appreciating such considerations—in effect, treating them as warranting identification. Appreciation, in this conception, entails reasons responsiveness. However, it is not clear perception responds appreciatively to considerations as favoring a proposition or to reasons as such. If perception does not respond rationally or appreciatively to factors supporting identity, crossmodal identification is not perceptual.

Put another way, crossmodal identification requires evidence that what is perceived using each of two senses is the same. That is because crossmodal identification belongs to the third type of case in which a subject appreciates identity. In such cases, an item is presented in differing ways that are compatible with distinctness. Appreciating identity requires responding to reasons for identity as such. If perception does not engage in what Siegel (2017) dubs *reckoning*, in which considerations are *taken* to support an upshot, as described by Boghossian (2014), then perception does not identify individuals across sensory modalities. If so, crossmodal identification does not take place in perception.

This argument sets the bar high for an appreciative response. However, perception could respond to factors that support identification even if a subject is not aware of its doing so. This echoes debates about whether perception involves inferences (Orlandi 2014; Siegel 2017). Transitions that are in several respects like inferences may occur in perception even if they are unlike inferences subjects perform. Accordingly, identification could take place in perception even if a subject does not identify *a* with *b*.

The clear upshot is that crossmodal identification in perception must differ in character from crossmodal identification in thought. That in principle is no problem. What matters is that

crossmodal identification requires representing something perceived with one sense as being the same as something perceived with another sense. Given differing presentations across senses, it is not obvious perception has the resources to do so.

Each obstacle could be answered, in the ways I have suggested. However, the lesson is that doing so takes work, and success is not guaranteed. Any account is bound to meet opposition. Some I have described, but others run deeper. For instance, many naïve realists will resist any proposal that requires perception constitutively to involve representing something perceived with one sense as being the same as something perceived with another sense. How to translate this idiom into friendlier terms is not evident. The hope was an account of which capacities crossmodal identification comprises, not a contentious partisan description of identification within perception. Given agreement that crossmodal identification occurs in thought, why go there?

5 Sensitivity to identity without crossmodal identification

Here is where things stand. Clear paradigms of token identification occur in thought. There are reasons to doubt that crossmodal identification takes place in perception. It is not evident what the cognitive approach leaves out. Thus, according to a reasonable line of thought, crossmodal identification is cognitive and not perceptual. Adapting Williamson's account of discrimination, we might say crossmodal identification is activating knowledge or warranted belief in identity across senses.

One response is to vindicate the perceptual approach to crossmodal identification. For instance, dig in and defend the claim that perceiving involves representing as being the same an item perceived using each of two distinct senses. I am not hopeful. Let us pursue another approach. I would like to consider the evidence that crossmodal identification takes place in perception and to offer an explanation that avoids the obstacles in the previous section.

According to the account I propose, perception is differentially sensitive to situations in which one thing is perceived using distinct senses. Perception is sensitive to sameness across senses. It registers crossmodal identity. My proposal does not say perception represents items perceived with distinct senses as being the same. It is compatible with saying that appreciating crossmodal identity requires extraperceptual cognition. Thus, it does not say crossmodal identification takes place in perception. This captures all we should ask from an account of

perception's role in crossmodal identification. Perceiving crossmodal identity does not entail crossmodal identification.

Before turning to the evidence, it helps to describe what it takes for perception to register identity across senses and consider an objection. In the first place, while perceiving often is a way of coming to know, perceiving need not activate knowledge or warranted belief. One might perceive *p* without believing *p*; one might perceive *p* without being justified in believing *p*, due to defeaters or lack of awareness; one might perceive *p* but lack concepts to think *p*; or one might seem to perceive *p* while *p* is false. More simply, one might perceive *p*, know *p*, but fail to activate one's knowledge that *p*. Therefore, perceptually registering the identity of an item presented to each of two senses does not entail activating knowledge or warranted belief in crossmodal identity. One could register perceptually the identity of something seen with something touched without cognitively appreciating it.

As I understand it, the bar for registering identity perceptually is low. It does not require thought. Indeed, perception can register the identity of *a* and *b* without representing *a* as being identical with *b*. And one can register identity perceptually without awareness of identity as such.

Perception therefore can register a crossmodal identity even if a subject fails to appreciate it. How a subject perceives can be sensitive to the fact that what is presented to one sense is the same as what is presented to another sense even if the subject does not recognize it. Most notably, perception can register crossmodal identity without representing an item perceived using one sense and an item perceived using another sense as being the same. Perceptually registering crossmodal identity does not imply perceptually representing anything *as being the same*. To insist that it does just renames the puzzle.

Which perceptual capacities does registering identity across senses comprise? In short, perception must be selectively sensitive to crossmodal identity. Perception at minimum ought to be differentially responsive to scenarios in which what is presented to one sense is token identical with what is presented to another sense.

To be differentially sensitive to a feature requires the capacity perceptually to detect it and to differentiate it from other features. Three conditions must be met. First, to detect it one must be capable of responding to the relevant feature. So, one must be affected by instances in which a single thing is perceived using each of two distinct senses. Since one may respond to diverse features in just the same way, mere detection does not suffice for selective sensitivity.

Second, one must be capable of differentiating the relevant feature from others. This means responding differently to it from how one responds to some range of other features. So, one must respond differently to crossmodal identity from how one responds to shapes, textures, or kinds. The most relevant contrast is between one's response to crossmodal identity and one's response to distinct items perceived with distinct senses. Without such a difference, there is little reason to think perception singles out crossmodal identity.[4] Finally, such differential sensitivity must be perceptual. Thus, one's responsiveness to crossmodal identity must rely on extracting information from sensory stimulation, and it must play the right role in one's psychology, especially with respect to cognition, belief, and action. If these conditions are met, perception is differentially sensitive to the token identity of an item perceived using each of two distinct senses. This is a way perceptually to register crossmodal identity.

Indiscriminable counterpart scenarios pose an objection. Suppose a subject perceives the same item using distinct senses. An indiscriminable counterpart is a scenario a subject cannot differentiate but which contains two distinct objects or events. For all a subject can tell, the scenarios match. But, in one case, two things are before the senses rather than one.

Being differentially sensitive to crossmodal identity requires responding differently to its presence and to the presence of other features. If one cannot tell the difference between crossmodal identity and its indiscriminable counterparts, why say perception registers crossmodal identity rather than some other feature, such as spatiotemporal coincidence?

This objection concerns the specificity or determinacy of sensitivity claims. These are some of the trickiest issues in philosophy of mind, perception, and psychosemantics, and I will not try to resolve them here. Instead, I hope to sidestep and to blunt their impact on this paper's project. Indeed, in three ways, sensitivity to a feature such as spatiotemporal coincidence helps support this paper's conclusions.

First, to appreciate crossmodal identity requires being sensitive to considerations favoring identity over distinctness. Perceptually registering crossmodal identity relies on cues to identity across senses. Spatial and temporal alignment are key evidence. Thus, being sensitive to spatiotemporal coincidence shows perception satisfies a necessary condition on registering crossmodal identity.

[4]Cf. Burge's (2010, 466) Principle of Relevant Representational Alternatives.

Second, being sensitive to either crossmodal identity or spatiotemporal coincidence differs from failing to be sensitive to any such relation between individuals perceived using distinct senses. Suppose perception is differentially sensitive to crossmodal coincidence but not identity. That remains a significant feature of perception, when compared with being sensitive only to spatiotemporal locations perceived using individual senses. Thus, in this project, perceptual sensitivity to spatiotemporal alignment might stand in for perceptual sensitivity to token identity across senses.

Third, spatiotemporal coincidence is being in the same place and time. Being sensitive to spatiotemporal coincidence is being sensitive to identity of place and time across senses. This may not imply perceptual sensitivity to the token identity of an individual perceived using distinct senses. Nonetheless, it is perceptual sensitivity to identity across senses.

6 Evidence for sensitivity to identity

This section describes evidence that perception is differentially sensitive to scenarios in which an item presented to one sense is token identical with an item presented to another sense. Where a subject perceives o_1 in m, and perceives o_2 in distinct n, sometimes the subject is perceptually sensitive to the oneness of o_1 with o_2. Results from psychophysics support the claim that perceptual systems are selectively responsive to cases in which the same item is before each of two distinct senses. This is less fancy than identification in thought. And it does not entail representing *as being the same*. Still, it is a way perceptually to register crossmodal identity.

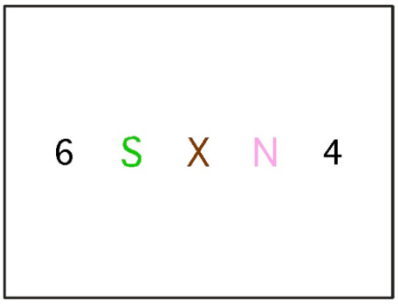
Evidence for sensitivity to identity stems from studies of *feature binding* and *object files*. Feature binding occurs when distinct features are treated perceptually as bundled or bound—to belong together—despite being registered separately by sensory systems. If visual systems extract separately information about shape and color, something accounts for a tennis ball's looking at once both yellow and round. Perceptually binding *yellow* and *round* reflects sensitivity to the token identity of that which bears each feature. Object files are temporary, episodic representations posited to explain how distinct features at a time are associated with a particular item and how perceivers keep track of an item over time. Tracking a tennis ball from moment to moment, even if it changes color, reflects sensitivity to its identity over time.[5]

[5]Green and Quilty-Dunn (forthcoming) is a state-of-the-art treatment of object files.

The *Many Properties Problem* illustrates what is at stake in feature binding (Jackson 1977). Perceivers differentiate seeing a red square beside a yellow circle from seeing a red circle beside a yellow square. But, in each case, just the same basic features are present. Something must account for the difference. What explains why *red*, *square*, and *left* go together in one case and not the other? Plausibly, a subset of features in a scene is bundled or bound. In this case, binding signals a common feature-bearing individual.

Translated to multisensory contexts, this argument does not offer such clear support for sensitivity to identity across senses. Consider how perceiving $color_1$ and $pitch_1$ on the left with $color_2$ and $pitch_2$ on the right differs from perceiving $color_1$ and $pitch_2$ on the left with $color_2$ and $pitch_1$ on the right. It is not evident $color_1$ and $pitch_1$ are bundled or bound together in one case but not the other. Color and pitch might belong to distinct individuals with similar locations. Moreover, it is not obvious introspectively that perception itself identifies locations across senses. Thus, it begs the question to appeal to the introspective exercise described in the Many Properties Problem to establish perceptual sensitivity to identity across senses.

6.1 Illusory conjunctions



Error type	*Example*	*Answers per trial*
Correct	Middle, brown X	0.44
Illusory conjunction	Left, green N	0.33
	Right, brown S	0.06
Single feature correct	Right pink	0.49
Single feature wrong	Something blue	0.06

Figure 1: Visual illusory conjunction, Treisman and Schmidt (1982) (Ptak 2008, Figure 14).

Illusory conjunctions outside focal attention are evidence for feature binding in unisensory contexts (Treisman and Schmidt 1982). If presented with a field of red circles and blue squares while doing another task, it is not uncommon to report a red square. Presented with a green "S" and pink "N," Treisman and Schmidt's participants sometimes perform as if they have seen a green "N" or pink "S" (Figure 1). Indeed, subjects have high confidence they have experienced such figures. According to Treisman and Schmidt, features are first coded separately then bound

together during object perception. The illusion reveals the capacity perceptually to bind distinct features that belong to the same object. Feature binding thus shows subjects are perceptually sensitive to the identity and distinctness of token feature bearers.

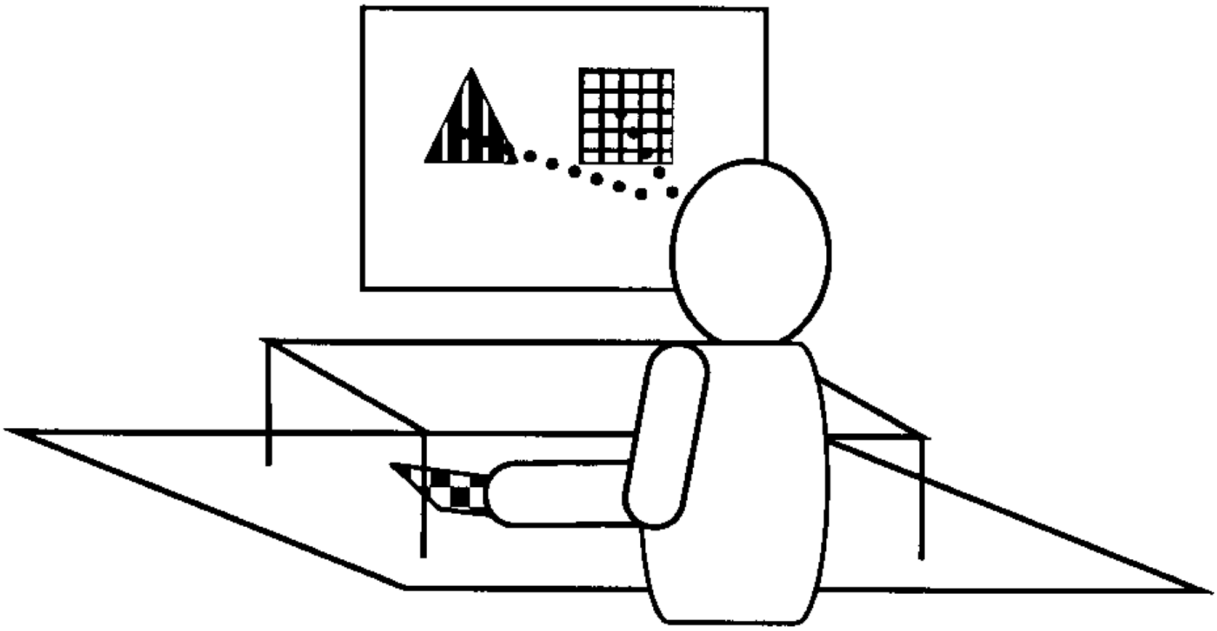


Figure 2: Crossmodal illusory conjunction, Cinel et al. (2002, Figure 1, 1245).

In a multisensory setting, Cinel et al. (2002) argue that illusory conjunctions occur between vision and touch. They first build objects with distinctive surface textures by gluing carpeting, Legos, fake fur, beans, and bubble wrap to 12 × 3 cm cardboard rectangles. They photograph each texture and use it to fill three visual figures: a square, a triangle, and a circle. The first experiment presents a tactual object, hidden beneath a screen, in one of two orientations, and two visual objects, for 75–300 ms (Figure 2). The task is to report the orientation of the tactual bar and the texture and shape of each visual object. When the tactual and visual textures differ, participants on average accurately report the visible texture 76% of the time. However, 8% of the time, participants report the felt texture as the visible texture. In illusory conjunction cases, subjects misattribute a texture they had felt to a shape they had seen.

Cinel et al. conduct further experiments using forced-choice matching, increased attention demands, memory loading, reversed vision and touch, and performance by a person with parietal damage. They argue that illusory conjunctions of visible and tactual features are

"perceptual in nature" (1261) and "preattentive" (1244) rather than effects of memory or cognition. Consistent with unisensory results, the effect persists in a target-matching task, and it is enhanced with greater memory demands, reduced attention, and same-hemifield stimuli. They conclude, "These results demonstrate that ICs are possible not only within the visual modality but also between two different modalities: vision and touch" (1245).

If a tactual feature can be ascribed illusorily to a visible figure, this suggests that in the good case perception is sensitive to the fact that what is touched is what is seen. In that case, the object seen does possess the tactual feature attributed to it. If so, one is selectively sensitive to token identity between vision and touch. If not, what explains attributing felt texture to a visible figure?

Here, there are reasons to pause. Studies supporting illusory conjunction rely on subjective reports, recall, and prior target matching, rather than more direct psychophysical measures. Indeed, Cinel et al. prompt reports and forced choices to probe feature conjunctions. This methodology is subject to extraperceptual interference and confounds, such as snap judgment, recognition, and recollection. So it is valuable to have further evidence.

Cinel et al.'s touch–vision paradigm raises a further doubt when compared with Treisman and Schmidt's unisensory paradigm. Treisman and Schmidt aim to test whether illusory conjunctions are part of visual phenomenology (especially Experiments I and II). Based on error patterns, confidence judgments, and their own impressions, Treisman and Schmidt report, "The phenomenal descriptions subjects gave made it clear that at least some of the errors were experienced as genuine illusions" (121). They say, "No conclusive answer can be given, but the evidence seems consistent with our own conviction, from the pilot experiment, that on some trials at least, illusory conjunctions have the character of perceptual experiences" (120).

Cinel et al. do not purport to examine perceptual experience. Their studies do not mention "experience," "phenomenology," or "confidence." Since their crossmodal paradigm does not test for experience, and it does not describe subjective reports or confidence judgments, it does not demonstrate that an illusory perceptual experience has occurred. So these results do not establish a true illusory conjunction.

There is a further disanalogy. In the unisensory paradigm, participants report visually experiencing color and shape jointly to qualify an apparent visible object. One figure appears

both “S”-shaped and pink, just as a true pink “S” looks. What seems pink seems also “S”-shaped. Evidently, one item appears to bear both visible features.

The touch–vision paradigm, however, does not show that a tactual feature and a visible feature are bound perceptually in this way as features of a single object. The felt texture does not evidently infuse visible shape like a color or fill it like a pattern. The visible figure looks to have visible shape and pattern, and the cardboard bar feels like it has texture and orientation. The tactual and visible features do not seem comingled or directly to be coinstantiated by one individual. The crossmodal paradigm does not demonstrate that felt texture (a texture as felt) seems to qualify any object that is seen. For all it shows, what is touched, which perceptibly has texture and orientation, may or may not appear to be one with what is seen, which perceptibly has shape and pattern.

Thus, in a key respect, the tactual–visual case is unlike the unisensory visual case. Because it does not demonstrate that tactual and visible features illusorily appear to belong to a single perceptible item, it does not show that some tactual feature bearer perceptibly is identical with a visual feature bearer. The crossmodal paradigm does not establish crossmodal identification in perception.

It does show that felt textures interfere with reporting and matching conflicting visual patterns. Consider a standard crossmodal illusion, such as ventriloquism, the McGurk effect, or the sound-induced flash. These illusions demonstrate that stimulating one sense can reshape experience associated with another. Each stems from attempting to resolve a conflict concerning features perceived using distinct senses, such as locations, phonemes, or quantities. Ventriloquism, for instance, reconciles a spatial discrepancy between sights and sounds.

In the crossmodal illusory conjunction paradigm, touch does not trigger an illusory visual pattern. Touch does not make the bubble wrap look like fur. So it is not a standard crossmodal illusion. But touch does affect performance in visual tasks. Roughly one in twelve times when they conflict, touch impacts a subject’s ability to report, recall, or match visual texture. This, however, does not require an illusory experience.

What then explains the interference? Handling conflict can reshape performance without causing illusion. What matters here is the performance, not the experience. The reason is that conflict presupposes a common subject matter. Handling conflict, being sensitive to it, means being sensitive to identity or sameness, for nothing prevents distinct things from having

conflicting features, including distinct textures. Cinel et al.'s crossmodal paradigm therefore illustrates that performance in perceptual tasks is differentially responsive to the fact that what bears felt texture typically also yields visible pattern. Felt texture sometimes overrides the visual pattern.

According to this interpretation, the crossmodal illusory conjunction paradigm shows typical human subjects are differentially sensitive to identity across senses. However, these results do not demonstrate crossmodal identification has taken place. Crossmodal conflict resolution does not require representing what is touched and what is seen as being the same. It does not require appreciating crossmodal identity.

6.2 Object-specific preview effects

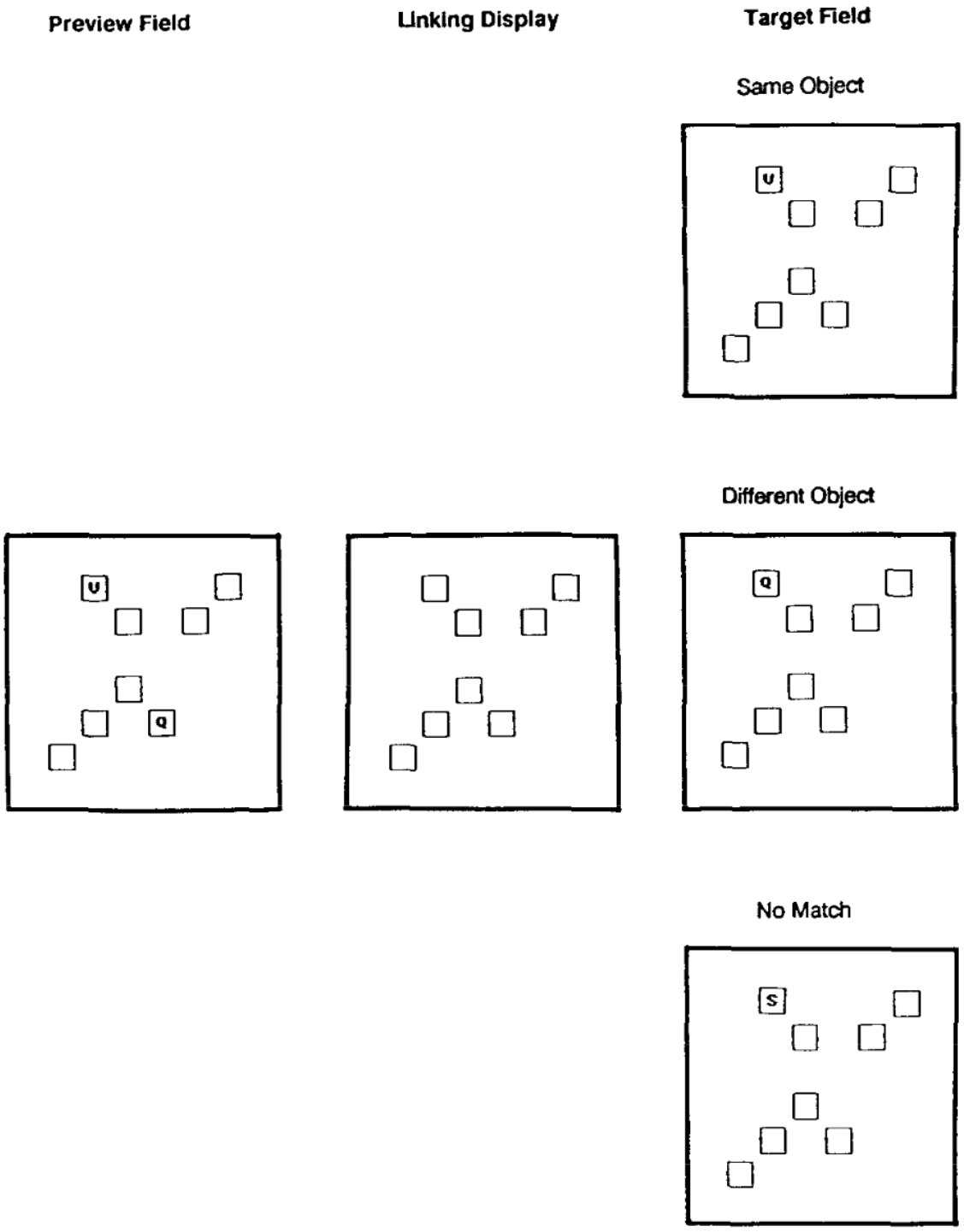


Figure 3: Visual object-specific preview effect, Kahneman et al. (1992, Figure 1, 181).

Turn to the *object-specific preview effect*, which is evidence for object files in visual perception (Kahneman et al. 1992). Suppose several objects at once are presented visually. Following a

brief interval, the same objects reappear. On the second viewing, subjects more quickly report a feature of one of those objects if that feature is unchanged from the preview to the target. Participants in one experiment (Figure 3) on average are 35 ms faster reporting which letter appears in a small box when it matches the letter that appeared earlier in that box (Same Object) than if it is an entirely new letter (No Match) or if that letter appeared earlier in a different box (Different Object)—roughly a 7% advantage for reappearing on the same object. Belonging to the same object, not simply having been viewed earlier, drives the speeded response. So the effect is object specific, unlike nonspecific feature priming.[6]

According to Kahneman et al., attention to the target triggers *reviewing*, in which one current object is selected and assigned to a corresponding object from the preview field. If the target and preview match, it facilitates recognition. If the target and preview mismatch, it interferes with recognizing the feature. The effect reveals the capacity to differentiate and assign features to distinct objects at a time and the capacity to keep track of an individual object viewed at distinct times. If so, the object-specific preview effect shows that subjects are perceptually sensitive to the situation in which a single object has distinct perceptible features and to the situation in which an object perceived at one time is token identical with an object perceived at another time.

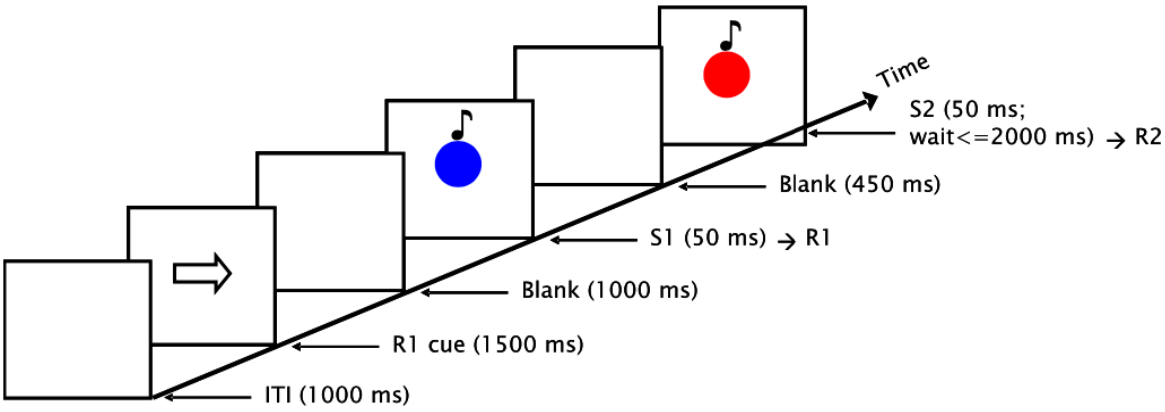


Figure 4: Audio–visual object-specific preview, Zmigrod et al. (2009, Figure 1, 675).

[6]In a study using whole words, the authors observed both nonspecific priming and object-specific preview effects (188).

Object-specific preview effects have been reported in multisensory settings. Zmigrod et al. (2009) describe audio–visual and audio–tactile experiments. First, they present a color–pitch preview, followed by an interval, then a color–pitch target (Figure 4). Participants in distinct trials report either the pitch or the color of the target. Subjects are quicker and more accurate reporting a target's pitch or its color when both pitch and color match the preview than when either pitch or color differs from the preview. Being paired with the same color facilitates one's response to a pitch, and vice versa. Similar results hold in a pitch–vibration study (Experiment 2).

This study is not a direct parallel to the unisensory visual paradigm introduced by Kahneman et al. (1992). It does not present multiple objects at once during the preview or target stage. And it collapses the distinction between No Match and Different Object conditions. So it does not distinguish an object-specific effect from nonspecific priming. To explain the effect does not require object files or same-object perception. The result does show that features perceived with distinct senses interact in a task for which one sense is not relevant. This means distinct features are associated across senses and compared over time. However, this experiment does not guarantee that color and pitch are perceived to belong to a single individual.

Nonetheless, if features perceived with distinct senses belong to one thing, this result makes sense. Objects tend not to change, so even task-irrelevant feature differences may slow you down. The study does not establish that the associated features are perceived to belong to one object. However, it does support perception's sensitivity to real world situations in which one thing has features perceived using distinct senses.

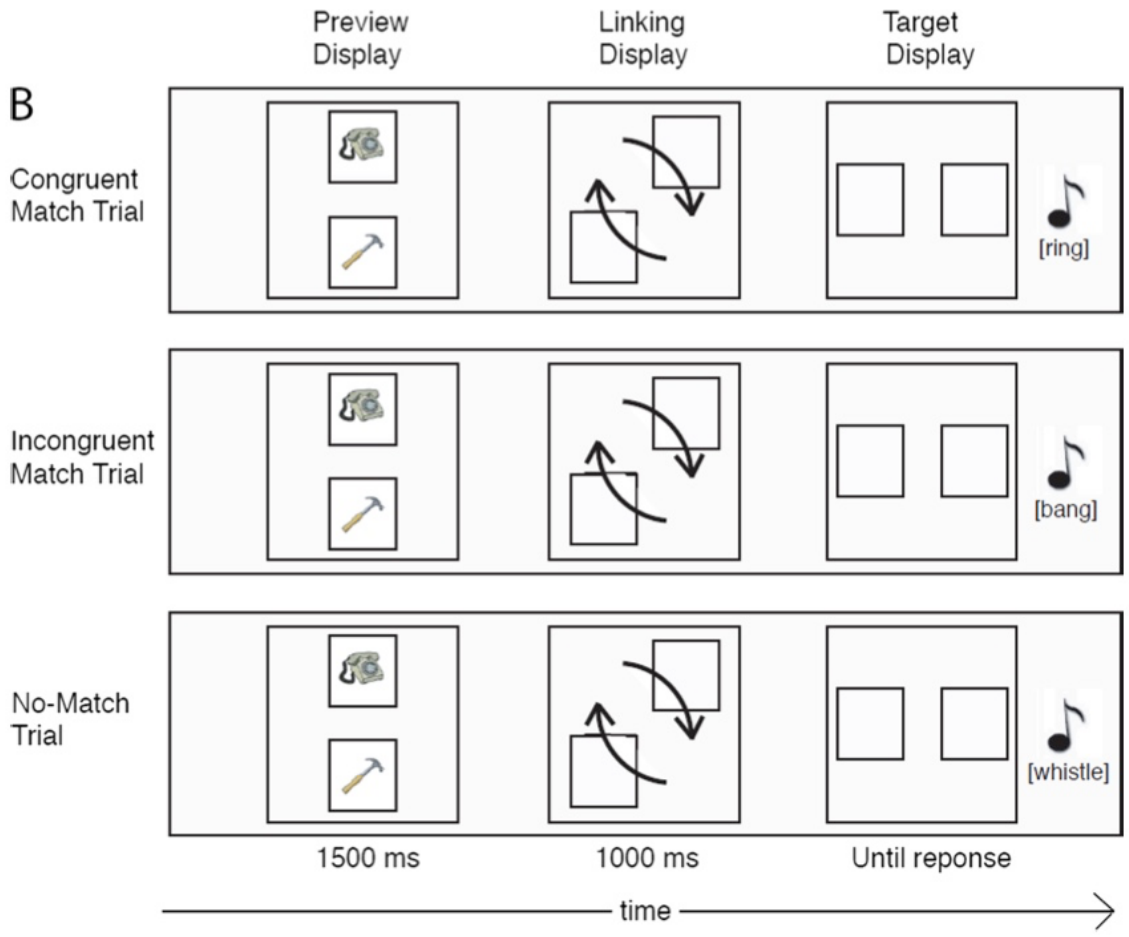


Figure 5: Multisensory object-specific preview effect, Jordan et al. (2010, Figure 1, 495).

Jordan et al. (2010) describe an intriguing variant to the original unisensory paradigm. This does demonstrate a multisensory object-specific preview effect. Jordan et al. present a brief preview picturing two objects framed in separate boxes (Figure 5). The pictures then disappear from the frames. After one second, the target display plays a sound at the location of one of the two frames. The sound either matches the object that appeared in that frame during preview, matches the object that appeared in the other frame during preview, or matches neither previewed object. The task is to report "match" if the sound fits *either* of the two previewed pictures and "no match" if not. Subjects on average are 18 ms faster in congruent than incongruent trials, a 2.5% advantage. They are 54 ms faster in incongruent than no match trials, reflecting a 7% general, nonspecific priming advantage even for the incongruent feature. While

the advantage is half that observed by Kahneman et al. in visual tasks, Jordan et al. nevertheless record a significant multisensory object-specific preview effect.

What follows? Notice first that this study does not present the same feature in the preview and the target. In that respect, it is unlike the visual paradigm. It displays a picture and later plays a corresponding sound. A phone is not the same feature as a ringing sound. Thus, the effect must not require perceptually identifying a feature seen with a feature heard.

Instead, the effect relies on discerning what makes the picture and sound congruent or incongruent. They are pictures and sounds of artifacts and animals—"a dog, whistle, train, hammer, piano, and phone" (496). But one may doubt that perception registers such kinds, natural or artifactual. Plausibly, perception reveals at most sense-specific gestalts. Thus, if the effect relies on recognizing a common kind shared by preview and target, that may signal an advantage in extraperceptual cognition rather than perception.

However congruence is explained, this multisensory preview advantage remains object specific. Object-specific preview effects require reidentifying something that persists through time. In the unisensory visual paradigm, this is a visible object, and it bears visible features. The object-specific advantage holds for features of that object that match—they are identical—between the preview and target. In Jordan et al.'s multisensory paradigm, the advantage also relies on reidentifying a visible frame in the target display as a visible frame from the preview. For the advantage to occur, the previewed picture and the target sound each must be associated with the same visible object. But this does not require representing what is seen and what is heard as being the same.

It follows that an audible feature must be associated with a visible object (frame). When a sound (ringing) and a congruent image (phone) each are associated with a single persisting visible object (or location), typically that means one thing in the environment is responsible for each. Sight and hearing then converge on one individual. The target display mirrors scenes in which one thing is responsible for the sights and the sounds. The multisensory object-specific preview effect therefore illustrates that perception is differentially sensitive to situations in which one thing is presented to vision and audition.

6.3 Attentional spreading

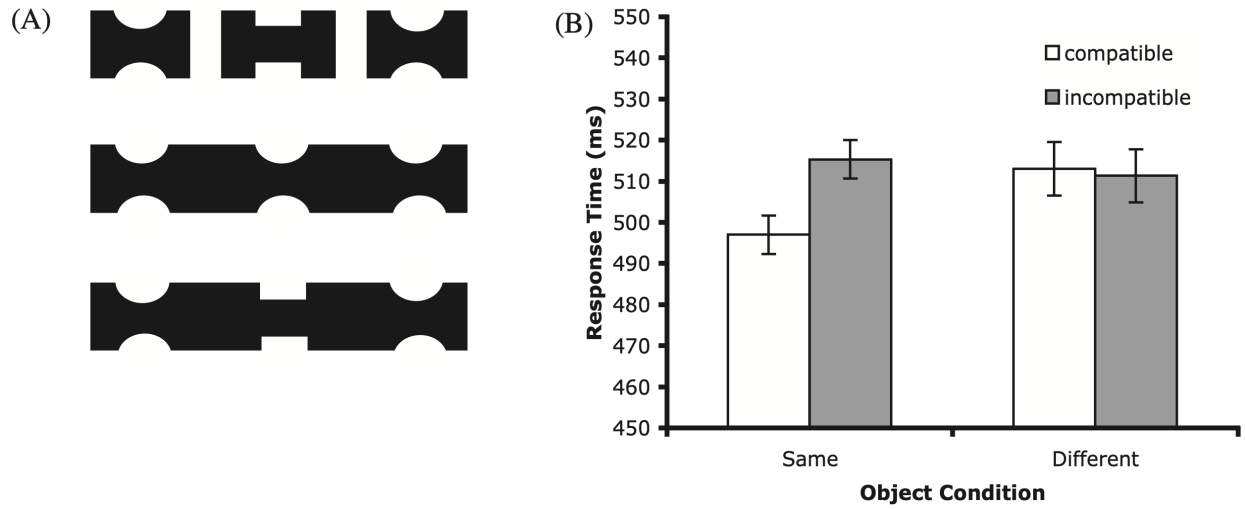


Figure 6: Visual object-based attentional spreading, Richard et al. (2008, Figure 1, 845).

Finally, consider *attentional spreading*, which supports sensitivity to identity across senses. Attending to one feature of an object enhances the salience of its other features. It does so more than it enhances salience for features of other objects. Unattended features do not have this effect. Attentional spreading relies on object-oriented attention, not simple spatial attention.

> An attentional spreading view of object-based attention proposes that the rate and efficiency of perceptual processes are improved by spreading attention through an attended object. This spreading likely enhances the representation of the attended object relative to unattended objects. (Richard et al. 2008, 843)

Richard et al. present features in fixed locations on distinct objects or the same object (Figure 6A). The target is the feature at the center of each bar—a circular or rectangular bite. Compatible flankers on the same object speed one's response to the target (middle stimulus, Figure 6A). Incompatible flankers interfere, slowing one's response (bottom stimulus). Flankers on distinct objects have no effect (top stimulus). Attentional spreading illustrates sensitivity to situations in which distinct visible features belong to one object.

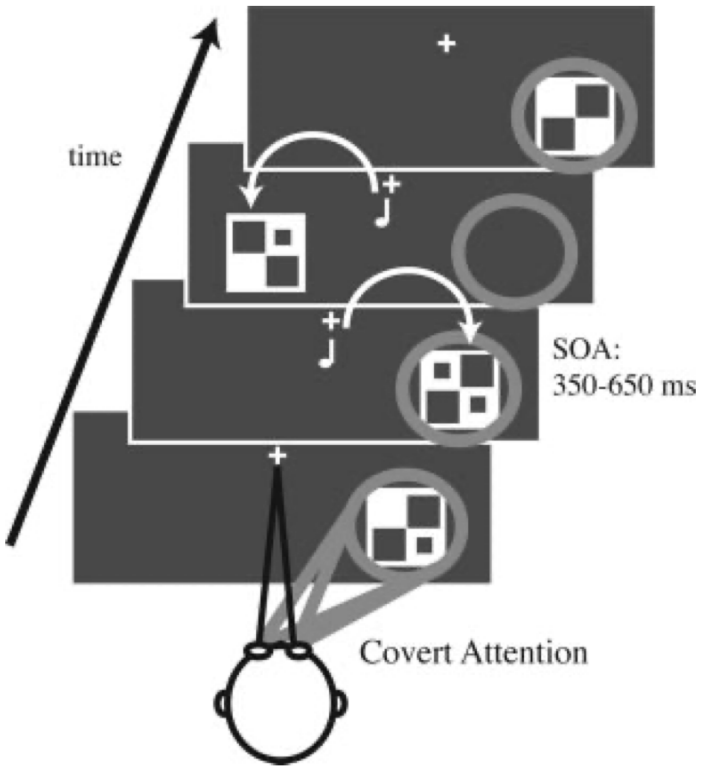


Figure 7: Audio–visual object-based attentional spreading, Busse et al. (2005, Figure 1, 18752)

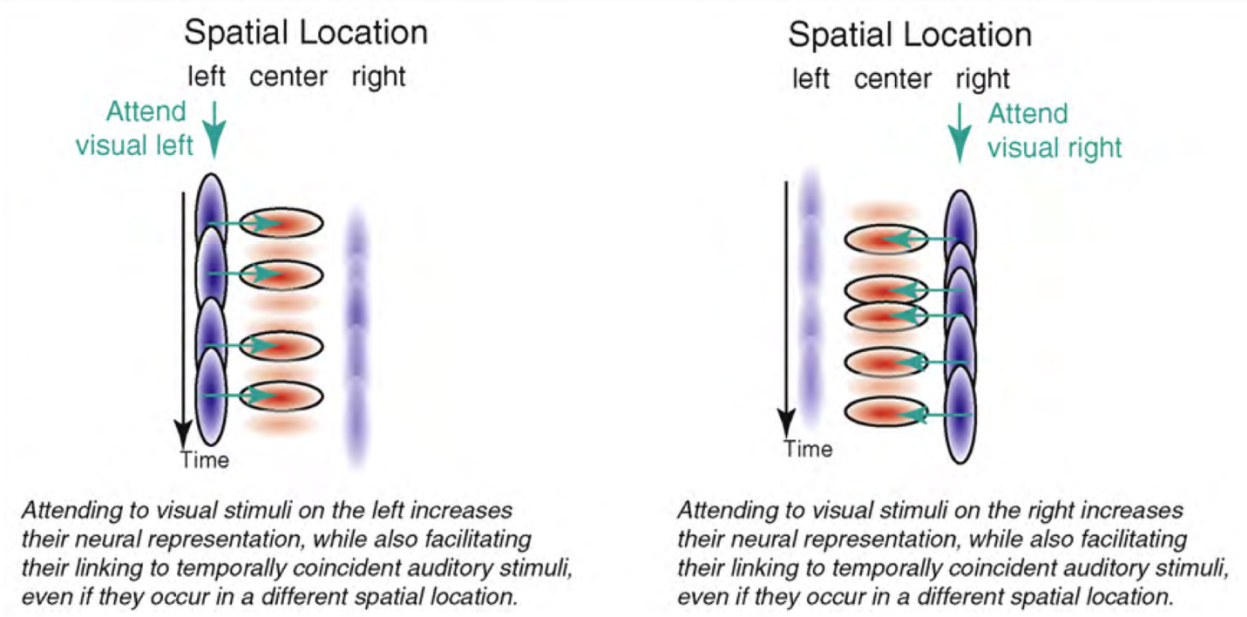


Figure 8: Audio-visual object-based attentional spreading, Talsma et al. (2010, Figure 1, 404)

Attentional spreading also occurs across senses. When a sound accompanies an image, attending to the image enhances not only the image's salience but also the sound's salience. In a study measuring event-related potentials associated with auditory perception, Busse et al. (2005) ask participants to focus on a central cross while covertly attending to one side. On either the attended or the unattended side, an image is shown (Figure 7). The task is to detect a target feature that appears on the image in roughly one in seven trials (in Figure 7, the two small squares). When a brief sound plays with an attended visual image, event-related potentials demonstrate enhanced auditory responsiveness to the sound itself, as compared to when the sound occurs with an unattended image. The sound does not affect accuracy or response time for the attended visual target. Subjects unsurprisingly are slower and less accurate detecting the visual feature on an unattended image. This study does not include behavioral measures regarding the sound.

In virtue of the parallel, it is tempting to conclude that this study reveals object-based attentional spreading. Attending the visual feature enhances one's response to the auditory feature. Since their locations differ, this is not just spatial attention. And the auditory feature is irrelevant, so it is not a task effect. According to this interpretation, attention spreads from visual feature to auditory feature because the two perceptibly belong to a single object (Figure 8).

This is too quick. Synchronous sights and sounds typically do stem from one source. So, crossmodal attentional spreading indicates sensitivity to situations in which one object or event is responsible for visible and audible features. However, to explain attentional spreading does not require that perception token identifies what is heard with what is seen. Perception need not represent an audible individual as being identical with a visible individual. It need not identify a single object or event as the bearer of both the visible feature (two small squares) and the audible feature (a tone). It is enough that synchronous visible and audible features are targets of special interest. What is seen and what is heard become associated because they occur at one time. Nevertheless, being associated or bound in this way does not require being token identified.

7 Binding without identification

The empirical evidence I have described shows that perception is differentially sensitive to circumstances in which one thing is perceptible by means of distinct senses. Features that

typically stem from a common source receive enhanced processing, facilitate speedier reporting, and sometimes are taken for one another. Perceptual responsiveness thus is selectively sensitive to crossmodal identity.

This operationalizes feature binding and object files across senses. But the evidence does not guarantee that crossmodal identification takes place within perception. Being differentially sensitive to crossmodal identity does not require co-ascribing features to a single individual or token identifying an individual perceived using one sense with an individual perceived using another. It does not entail representing crossmodal identity as such.

Feature binding is too permissive. It does not require a unified perceptual process or mechanism (cf. Treisman 2003, 98). And it does not specify a relation between things in the environment. As operationalized, feature binding is compatible with a variety of relations, not each of which entails perceptually identifying one feature bearer with another feature bearer. Consider some options. First, one might think feature binding encodes *co-instantiation*. This is plausible in certain unisensory examples. If a tennis ball looks yellow and round, one thing perceptibly bears both visible features at once. That item which one sees to be yellow, one sees also to be round. *Yellow* and *round* seem perceptibly to be co-instantiated by one individual.

However, co-instantiation is not plausible for other examples of feature binding. When one sees a red circle and hears a pitched sound (as in Zmigrod et al. 2009), it is not necessary that color, shape, pitch, and loudness perceptibly are or seem to be co-instantiated by one item. The visible individual that seems perceptibly to have color and shape need not also seem audibly to have pitch and loudness. The red circle is not the sound. No single item seen and heard evidently bears color, shape, pitch, and loudness at once.

Second, one might instead think feature binding indicates *co-constituency* or *co-parthood*. Touch conveys the fuzzy felt underside of a tennis ball while vision reveals its yellow upper hemisphere. The undersurface is a tangible part, the uppersurface is a visible part, and each perceptibly is a part of a whole. Binding in this understanding occurs partwise. One perceives bound tactual parts and visible parts to form a common whole. This suggests a model for crossmodal perceptual identification, if the whole that is felt is the whole that is seen. In that case, something perceptibly bears parts perceived with distinct senses.

The parthood model even illuminates audio–visual binding. Audible sounds and visible objects perceptibly belong to an encompassing event, such as a collision or explosion, that one

perceives multisensorily. This approach is especially promising, and I have developed it elsewhere (O'Callaghan 2016; Green 2019).

In the present context, however, there are two noteworthy caveats. One is that nothing in the studies described suggests binding across senses ought to be understood to specify co-constituency. In Zmigrod et al. (2009), what is the whole to which a red circle and a C-sharp tone perceptibly belong? In Jordan et al. (2010), is it plausible that subjects see a picture of a phone, represent it to be part of a broader event in which that phone is caused to ring, later hear a ringing sound, represent it to be part a broader event, and perceptually token identify the event seen with the event heard? These interpretations are not obvious or mandatory. The experimental evidence thus does not show that feature binding requires co-constituency. The theory of binding is one thing; a theory of perceptible objects is another.

Even so, there is another caveat. Perceptible co-constituency does not demonstrate crossmodal identification has taken place. Suppose one sees a red phone while hearing a loud ringing sound. And suppose binding occurs, such that the phone and the ringing perceptibly are parts of a whole. If so, a red visible object and a loud audible sound perceptibly are parts of some further object or event.

Nevertheless, the whole need not be perceptible with any single sense. Perceiving the whole, a composite, may require joint or cooperative use of multiple senses. If so, it is perceived only through multisensory perception. But such a novel perceptual act need not token identify what is seen with what is heard. The visible object may remain numerically distinct, according to multisensory perception, from the audible sound. The same can hold for each visible object and each audible event. If so, it need not be the case that for an individual, o_1, that *s* perceives with *m*, and individual, o_2, that *s* perceives with *n*, perception token identifies o_1 with o_2. Where o_1 and o_2 are distinct perceptible proper parts of o_3, and perceiving o_3 requires the joint use of both *m* and *n*, partwise binding can occur without token identifying what is visible with what is audible. Therefore, partwise multisensory binding does not entail crossmodal identification.

Third, one might think binding reflects numerical identity *across time*. For instance, the object-specific preview effect relies on tracking one thing between two times. Something heard may be bound temporally with something seen at another time, thus implicating crossmodal identity over time.

Current evidence does not show that token identification across times and senses occurs in perception. For instance, in Jordan et al.'s multisensory object-specific preview paradigm, subjects see a phone, follow the visible frame, then hear a ringing sound. The object-specific advantage relies on visual tracking plus synchronous binding between the frame and sound. The previewed phone and later sound need not be bound. Crossmodal reidentification may rely on extraperceptual resources, such as recognition, long-term memory, concepts, or inference.

Fourth, one might think binding conveys a common *causal source*. Rather than co-instantiation, co-constituency, or identity over time, binding could signal features perceived using distinct senses with the same causal origin. Felt heat and flickering light each come from the stove. Binding implicates their sharing a source. This is one way to interpret Bayesian causal source analyses proposed by empirical researchers (see Shams and Beierholm 2010).

Causal source analysis is a powerful way to explain crossmodal binding effects. But it is one thing to explain perceptual processes in Bayesian terms; it is another to say perception represents causal sources of features perceived using distinct senses as such. Perception may be sensitive to the token identity of causal sources in a multisensory setting thanks to Bayesian processes, but binding does not establish that perception token identifies the cause of a feature perceived using one sense with the cause of a feature perceived using another sense (see also Green 2019; Rescorla 2020).

Finally, one might think empirical evidence is compatible with treating binding across senses as a kind of *association* (Fulkerson 2011; Connolly 2014). Association is very permissive. Since it is so non-committal, association allows a range of differing relations in which audible and visible features may stand. Understood this way, binding could reflect perceptual sensitivity to identity across senses. However, even regular crossmodal association does not show perception token identifies an item perceived using one sense with an item perceived using another sense.

Perception is sensitive to situations in which one item is presented to distinct senses. However, studies of feature binding do not establish that perception involves token identification across senses. Binding nonetheless reflects a way to be sensitive to crossmodal identity.

8 Registering identity across senses

Empirical studies of binding and object files do not establish crossmodal identification in perception. The result of binding is compatible with a variety of relations in which perceptible items may stand. Binding as operationalized in such studies does not require a perceptual process in which an individual perceived in one sensory modality is token identified with an individual perceived in a distinct sensory modality.

Studies of binding and object files do provide evidence that perception is differentially sensitive to crossmodal identity. Illusory conjunctions, object-specific preview effects, and attentional spreading demonstrate that perception typically registers something presented at once to distinct senses. The lesson is that perception is selectively responsive to identity across senses, even if perception does not represent identity across senses as such.

This captures all we should ask from an account of perception's role in crossmodal identification. In the first place, perception's pattern of sensitivity is an apt response to identity across senses. Perception responds selectively to cues that are evidence for identity, such as spatial proximity and temporal coincidence. And the way it responds makes sense in the face of crossmodal identity. For instance, resolving conflict between felt texture and visible pattern, taking more time to report barking when it accompanies an object with a history as a phone, and devoting greater attention to sounds that are synchronous with items commanding visual attention are appropriate when one thing is presented to distinct senses.

Furthermore, perception's pattern of sensitivity to crossmodal identity redounds outside perception. For instance, it benefits perception-based recognition. The effects described reorganize perception in ways that enhance perceptual cues to crossmodal identity: conflict resolution reduces feature misalignment; same-object advantages prioritize continuity and consistency over time; attention spreading highlights multisensory clusters or configurations of features. Each positions perceivers better to recognize one thing that is presented to distinct senses.

Sensitivity to crossmodal identity also improves action, making it easier to coordinate complex behavior informed by distinct senses. With better perceptual information, one can act more fluently and more effectively. There is less to work out, less evidence to gather, less need to sort out whether to trust one's ears or one's eyes. Being differentially sensitive to crossmodal identity therefore is not idle or superfluous in relation to cognition and action.

One may conclude that typical human subjects perceive token identity across senses. That is alright. However, this sets the bar low for perceiving crossmodal identity. It does not require crossmodal identification in perception, representing an item perceived with one sense as being the same as an item perceived using another sense, or awareness of sameness. The existence of a rationalizing account appealing to identity across senses does not imply representing or appreciating crossmodal identity. Thus, it is a mistake to model perceiving crossmodal identity on crossmodal identification in thought.

This approach recognizes that multisensory perception is sensitive to identity in what is presented to distinct senses. But the evidence does not establish that perception token identifies what is perceived using one sense with what is perceived using another sense. So this approach does not claim that crossmodal identification takes place in perception. In this respect, it remains a skeptical approach. This is a Goldilocks position. It appreciates the evidence, but it does not overintellectualize perception. My contention is that this does all the work a theorist needs. We should not ask more of perception.

9 Conclusion

This paper describes two opposing views about crossmodal identification. According to the first, crossmodal identification occurs only in thought or extraperceptual cognition. According to the second, crossmodal identification also occurs in perception. Strictly speaking, it endorses the first and denies the second.

In this picture, crossmodal identification involves appreciating identity across senses. It may consist in activating knowledge or warranted belief in identity. However, existing evidence does not establish that perception itself token identifies an individual perceived in one sensory modality with an individual perceived in another sensory modality.

Perception is not unalive to crossmodal identity. Empirical studies show multisensory perception is differentially sensitive to situations in which distinct senses are presented with a single individual. Studies of feature binding and object files pioneered in vision research have been adapted to demonstrate multisensory effects that parallel unisensory ones. Illusory conjunctions, object-specific preview effects, and attentional spreading have been observed in multisensory settings. Multisensory perception responds, and it responds in ways that are appropriate, when a single item is presented to distinct senses. It resolves conflicts, speeds

responsiveness, and spreads attention. This, in turn, enhances one's capacity to recognize and to act on multisensory objects.

But binding is very permissive. Binding does not require a unified type of perceptual process, and it does not yield a single, uniform perceptual structure. The binding observed in multisensory contexts is compatible with a variety of relations in which perceptible features and objects may stand. Not each such relation requires a single object or feature bearer. Moreover, binding does not guarantee representing what is perceived using one sense as being identical with what is perceived using another sense. So, as operationalized in empirical studies of multisensory perception, binding can occur across senses even if crossmodal identification as such does not take place in perception.

In this approach, typical human perception is differentially sensitive to crossmodal identity. This is distinct from the capacity to token identify an individual across senses. But it is enough to serve the needs of recognition and action. Perception registers crossmodal identity, but crossmodal identification as such belongs to thought.

Acknowledgments

Many thanks to Austin Andrews, Kevin Connolly, E. J. Green, Jake Quilty-Dunn, Jeff Speaks, and an anonymous reviewer for generous written comments.

References

Bayne, T., 2014. 'The multisensory nature of perceptual consciousness'. In *Sensory Integration and the Unity of Consciousness*, edited by D. Bennett and C. Hill, 15–36. MIT Press.

Boghossian, P., 2014. 'What is inference?' *Philosophical Studies* 169: 1–18.

Burge, T. 2010. *Origins of Objectivity*, Oxford University Press.

Busse, L., Roberts, K., Crist, R., Weissman, D., and Woldorff, M., 2005. 'The spread of attention across modalities and space in a multisensory object'. *Proceedings of the National Academy of Science* 102: 18751–6.

Campbell, J., 1997. 'Sense, reference, and selective attention'. *Proceedings of the Aristotelian Society* 71: 55–74.

Cinel, C., Humphreys, G., and Poli, R., 2002. 'Cross-modal illusory conjunctions between vision and touch'. *Journal of Experimental Psychology: Human Perception and Performance* 28: 1243–66.

Connolly, K., 2014. 'Making sense of multiple senses'. In *Consciousness Inside and Out*, edited by R. Brown, 351–64. Springer.

Fine, K., 2007. *Semantic Relationism*. Blackwell.

Fulkerson, M., 2011. 'The unity of haptic touch'. *Philosophical Psychology* 24: 493–516.

Green, E. J., 2019. 'A theory of perceptual objects'. *Philosophy and Phenomenological Research* 99: 663–93.

Green, E. J., 2021. 'Binding and differentiation in multisensory object perception'. *Synthese* 198: 4457–91.

Green, E. J. and Quilty-Dunn, J., forthcoming. 'What is an object file?' *British Journal for the Philosophy of Science*, doi: 10.1093/bjps/axx055.

Jackson, F., 1977. *Perception: A Representative Theory*. Cambridge University Press.

Jordan, K., Clark, K., and Mitroff, S., 2010. 'See an object, hear an object file: Object correspondence transcends sensory modality'. *Visual Cognition* 18: 492–503.

Kahneman, D., Treisman, A., and Gibbs, B., 1992. 'The reviewing of object files: Object-specific integration of information'. *Cognitive Psychology* 24: 175–219.

O'Callaghan, C., 2014. 'Intermodal binding awareness'. In *Sensory Integration and the Unity of Consciousness*, edited by D. Bennett and C. Hill, 73–103. MIT Press.

O'Callaghan, C. 2016. Objects for multisensory perception. *Philosophical Studies*, 173: 1269–89.

Orlandi, N., 2014. *The Innocent Eye: Why Vision Is Not a Cognitive Process*. Oxford University Press.

Ptak, R. 2008. Saliency, Selection, and Spatial Neglect. Habilitation thesis.

Recanati, F., 2012. *Mental Files*. Oxford University Press.

Rescorla, M., 2020. 'Perceptual co-reference'. *Review of Philosophy and Psychology* 11: 569–89.

Richard, A., Lee, H., and Vecera, S., 2008. 'Attentional spreading in object-based attention'. *Journal of Experimental Psychology: Human Perception and Performance* 34: 842–53.

Shams, L. and Beierholm, U., 2010. ‘Causal inference in perception’. *Trends in Cognitive Sciences* 14: 425–32.

Siegel, S., 2017. *The Rationality of Perception*. Oxford University Press.

Spence, C. and Bayne, T., 2015. ‘Is consciousness multisensory?’ In *Perception and Its Modalities*, edited by D. Stokes, M. Matthen, and S. Biggs, 95–132. Oxford University Press.

Talsma, D., Senkowski, D., Soto-Faraco, S., and Woldorff, M., 2010. ‘The multifaceted interplay between attention and multisensory integration’. *Trends in Cognitive Sciences* 14: 400–10.

Treisman, A., 2003. ‘Consciousness and perceptual binding’. In *The Unity of Consciousness*, edited by A. Cleeremans, 95–113. Oxford University Press.

Treisman, A. and Schmidt, H., 1982. ‘Illusory conjunctions in the perception of objects.’ *Cognitive Psychology* 14: 107–41.

Williamson, T., 1990. *Identity and Discrimination*. Basil Blackwell.

Zmigrod, S., Spapé, M., and Hommel, B., 2009. ‘Intermodal event files: Integrating features across vision, audition, taction, and action’. *Psychological Research* 73: 674–84.