




The Different Bases of the Meaning and of the Seeing-in Experiences

Fabrizio Calzavarini¹ · Alberto Voltolini¹ 

Accepted: 23 January 2023
© The Author(s) 2023

Abstract

There are some complex experiences, such as the experiences that allow us to understand linguistic expressions and pictures respectively, which seem to be very similar. For they are stratified experiences in which, on top of grasping certain low-level properties, one also grasps some high-level semantic-like properties. Yet first of all, those similarities notwithstanding, a phenomenologically-based reflection shows that such experiences are different. For a meaning experience has a high-level fold, in which one grasps the relevant expression's meaning, which is not perceptual, but is only based on a low-level perceptual fold that merely grasps that expression in its acoustically relevant properties. While a pictorial experience, a seeing-in experience, has two folds, the configurational and the recognitional fold, in which one respectively grasps the physical basis of a picture, its vehicle, and what the picture presents, its subject, that are both perceptual, insofar as they are intimately connected. For unlike a meaning experience, in a seeing-in experience one can perceptually read off the picture's subject from the picture's vehicle. Moreover, this phenomenological difference is neurologically implemented. For not only the cerebral areas that respectively implement such experiences are different, at least as far as the access to those experiences' respective high-level content is concerned. As is shown by the fact that one can selectively be impaired in the area respectively implementing the meaning vs. the seeing-in experience without losing one's pictorial vs. semantic competence respectively. But also, unlike meaning experiences, the area implementing the seeing-in experiential folds is perceptual as a whole. For not only a picture's subject can be accessed earlier than an expression's meaning, but also the neural underpinnings of such folds are located in the perceptual areas of the brain.

Keywords Meaning experiences, Seeing-in experiences · Neural implementation

Extended author information available on the last page of the article

Published online: 16 February 2023

1 Introduction

There are some complex experiences, such as the experiences that allow us to understand linguistic expressions and pictures respectively, which seem to be very similar. For they are stratified experiences in which, on top of grasping certain low-level properties, one also grasps some high-level semantic-like properties. Yet first of all, those similarities notwithstanding, we claim that a phenomenologically-based reflection shows that such experiences are different (§ 1). For a meaning experience has a high-level fold – in which one grasps the relevant expression’s meaning – that is not perceptual, but is only based on a low-level perceptual fold that merely grasps that expression in its visually or acoustically relevant properties (colors and shapes, or sounds, and possibly also its morpho-syntactic organization). While a pictorial experience, what Wollheim (1980, (1987, (1998, 2003a,b) takes to be a seeing-in experience, has two folds, the configurational and the recognitional fold – in which one respectively grasps the physical basis of a picture, its vehicle, and what the picture presents, its subject – that are both perceptual, insofar as they are intimately connected. For unlike a meaning experience, in a seeing-in experience one can perceptually read off the picture’s subject from the picture’s vehicle. Moreover and very interestingly, as we shall claim, this phenomenological difference is neurologically implemented. For not only the cerebral areas that respectively implement such experiences are different, at least as far as the access to those experiences’ respective high-level content is concerned. As is shown by the fact that one can selectively be impaired in the area respectively implementing the meaning vs. the seeing-in experience without losing one’s pictorial vs. semantic competence respectively (§ 2). But also, unlike meaning experiences, the area implementing the seeing-in experiential folds is perceptual as a whole. For not only a picture’s subject can be accessed earlier than an expression’s meaning, but also the neural underpinnings of such folds are located in the perceptual areas of the brain (§§ 2–3). As is *inter alia* shown by the particular case of one’s competence with ambiguous pictures on the one hand and with ambiguous expressions on the other hand (§ 3).

2 Unlike Seeing-in Experiences, Meaning Experiences are not Proper Fusion Experiences for They Are Not Perceptual, but Only Perceptually-Based

On the one hand, *seeing-in* experiences, the experiences that for Wollheim (1980, (1987, (1998, 2003a,b) determine what it is for a depiction (an intentionally-based picture such as a painting, a sketch and a drawing, as well as a causally-based picture such as a photo, a movie or TV shot, and perhaps also a mirror-image and a shadow) to be a *pictorial* representation, are *twofold* experiences. Their first fold, the *configurational fold*, consists in the perception of the pictorial *vehicle*, i.e., the picture in its organized physical basis. Its second fold, the *recognitional fold*, depending for its existence on the first fold, consists in the perception of the pictorial *subject*, i.e., the scene the picture presents.¹

¹ Some people (Husserl 2006; Nanay 2016, 2018; Voltolini 2015, 2018) further distinguish what the picture is about, the picture’s proper subject, from what the picture presents. For our purposes we do not need this distinction.

On the other hand, *meaning* experiences, the experiences Strawson (1994) labeled *as of understanding*, are also twofold experiences that are constituted by a first fold in which one perceives, either visually or auditorily, an expression in its morpho-syntactic structure and, on the top of the first, a second fold, the proper meaning fold, in which one experiences the meaning of that expression.²

Their similarity notwithstanding, these experiences are of a different kind. For, while there is room to consider seeing-in experiences as, though *sui generis*, perceptual experiences, meaning experiences can only be perceptually-based experiences. For, unlike the recognitional fold, the second fold of a meaning experience is experiential, but not perceptual in character (Voltolini 2020a).

In order to argue for this result, on the one hand, one may start with noticing that in the case of a seeing-in experience, one can read off what is grasped in the recognitional fold, the pictorial subject, from what is grasped in the configurational fold, the pictorial vehicle. To begin with, in order to understand how this reading-off works, one must remark that, as Wollheim himself (1987: 46) underlines, the two folds are not the same as the corresponding experiences of the vehicle and of the subject of a picture taken in isolation. In particular, the configurational fold is not the same as the perception taken in isolation of what stands in front of the picture's experiencer. One way of accounting for this difference is to claim that such a fold and that perception differ in their object, or better, in their object's properties (Voltolini 2015), since the fold has a content that is richer than that of that perception. That perception grasps the physical object facing the experiencer *qua* mere 2D object among other physical objects, let us call it the *mere* picture's vehicle. By contrast, the configurational fold grasps what we called the *pictorial* vehicle, or, as we can now say, the vehicle *qua* enriched by its *grouping properties*, i.e., the properties for its elements to be arranged in a certain way. In particular, these are the grouping properties organized in the third dimension;³ namely, the properties of the vehicle's elements to be arranged according to a certain direction along a certain dimension in a 3D space. This arrangement enables one to see in the configurational fold an item, the pictorial vehicle, which, unlike the vehicle taken in isolation, is not a mere 2D item, but a 3-D like item. Now as we said, such a grasp of the pictorial vehicle is still perceptual. For, although grouping properties are high-level properties, in their merely depending (generically)⁴ on the low-level perceptual properties of the vehicle, i.e., its colors and shapes, their apprehension is perceptual. For not only that apprehension is immediate, just as the apprehension of such low-level properties, but is also based on a perceptually relevant *selective* form of attention (Stokes 2018); notably, a *holistic* form of attention that enables one to perceive the vehicle as appropriately grouped. As it may be noticed by the fact that once this form of attention is activated, the scene one

² Typically (though not necessarily: cf. e.g. Wittgenstein 2009⁴: II,xi,§ 276), the layer of meaning involved in the meaning experiences is the lexical one. Incidentally, we are not talking of the meaning experiences, if any, occurring in the *production* and not in the *comprehension* phase of language. For insofar as they are *inner*, such experiences are hardly perceptual.

³ The perception of the vehicle taken in isolation may include its grouping properties as regards the first and the second dimension.

⁴ Generically, for the grouping may remain the same even though the low-level properties change. In the example we will provide below, the dalmatianwise grouping in the configurational fold may remain the same even if the spots to which it applies change either their color or their shape.

perceives radically changes. This is the form of attention that Nanay (2016, 2019) takes to be focused on an object and distributed across its properties. Indeed, as Calzavarini and Voltolini (2022) maintain, immediacy and holistic attention are not only necessary, but possibly jointly sufficient conditions, for the perception of high-level properties. There is no room here to properly deal with the issue of the distinction between perception and cognition (see Stokes 2018 for details), yet definitely such criteria may help in drawing a divide between the two kinds of mental states: perceptual states are states that only grasp either low-level properties or high-level properties singled out by means of those criteria.

Moreover, in the recognitional fold of the seeing-in experience one can read off the pictorial subject from the vehicle so arranged once one perceives that vehicle in that arrangement. For that arrangement enables one to perceptually recognize that subject in that vehicle. Given that enabling, indeed, that recognition has a perceptual status as well. In other words, perceiving the vehicle so arranged makes it the case that one perceives that subject as well. More precisely, the fact that the configurational fold has an enriched content mobilizing 3D grouping properties of the pictorial vehicle enables one to recognize, in the recognitional fold, a different 3D item in that vehicle – notably, a 3D scene (or, in a very similar proposal, a spatiotemporal region: Nanay 2022) – by virtue of the fact that the content of the latter fold matches the content of the former fold; in particular, elements to which is ascribed a certain 3D location in the former fold correspond to elements to which that location is ascribed in the latter fold (Voltolini 2015). As is proved by the fact that, as Wollheim himself intuited by distinguishing seeing-in experiences from the experiences of figures in the Rorschach tests (1980:138-9), there is no voluntary or anyway arbitrary element in the subject's apprehension, as it could be the case if that apprehension had an imaginative rather than a perceptual nature (see also Nanay 2022 for a non-imagistic but perceptual account of seeing-in experiences).

One can vividly realize that the two aforementioned folds work as stated in the seeing-in experience by appealing to a paradigmatic case, the case of experiencing 'aspect dawning' pictures. In this case, instead of perceiving a picture at once as one normally does, one can split an earlier perception of the mere picture's vehicle from a later seeing-in experience of the picture. The earlier perception is just a perception of the vehicle taken in isolation, a mere 2D item characterized by its low-level properties (its colors and shapes). By contrast, the seeing-in experience is constituted not only by a perception of the pictorial vehicle as the configurational fold of that experience, hence by something that has an enriched content due to the 3D grouping properties it mobilizes, but also by the recognitional fold of that experience in which the pictorial subject is also perceived, i.e., a 3D scene matching the 3D-like silhouettes that are perceived in the configurational fold. Consider the famous picture of a Dalmatian. At time t , one merely perceives an array of black and white spots. Yet at time t' , by means of holistically attending that array, one manages to group it according to a figure-ground 3D segmentation in which a 3D dalmatianwise item is protruded out of a background. So, by now facing a pictorial vehicle, at t' one grasps a content that is richer than the content one grasped at t , while facing a mere picture's vehicle. By virtue of that very segmentation, finally, one is able to perceive in the vehicle enriched by that segmentation the subject that one recognizes; namely, the 3D scene of a Dalmatian in front of a background.

On the other hand, in a meaning experience one certainly perceives the visual or acoustic properties of the relevant expression, including its morphosyntactic features. Yet one cannot read off the meaning of an expression from so perceiving that expression, even in its morphosyntactical complexity. For suppose even that perceiving that expression in its morphosyntactical complexity amounts to a high-level perception in which one perceives that expression by again holistically attending to its morpho-syntactic structure that depends (generally as well)⁵ on the low-level properties of that expression.⁶ Nevertheless, *pace* Wittgenstein (1991: § 869, 2009⁴: I§ 568), one cannot recognize in that expression so articulated its meaning in a perceptually relevant sense; by reading it off, so to say. For there definitely is no matching between the content of the perceptual fold of the meaning experience in which one sees or hears the expression in its morphosyntactic properties and the content of the other fold of that experience, the proper experience of the meaning of that expression. Indeed, as Schier (1986) originally noted, unlike pictures, linguistic expressions do not possess *natural generativity* (understanding one expression does not make one understand any other expression whatsoever, unless one knows its meaning), meaning is added *conventionally* to that expression, however one accounts for the nature of such conventionality. Hence, even if as regards a meaning experience one ends up having a twofold experience in whose first fold one perceives that expression in its morphosyntactical complexity, the second fold of that experience, the proper meaning fold, is merely juxtaposed to the first one, in the sense that, unlike a seeing-in experience, no real fusion experience arises from the simultaneous grasping of the two folds of the meaning experience. For, unlike a seeing-in experience, for the above reasons such two folds are not compenetrated.

Granted, in a meaning experience its second fold is experiential in character, and that character is irreducible to the character, admittedly perceptual, of the first fold. One's overall experience of the expression in question indeed changes, once one understands its meaning (Siewert 1998; Horgan and Tienson 2002; Pitt 2004, Strawson 1994, Chudnoff 2015). Yet *pace* Brogaard (2018), for the aforementioned reason of recognition failure it is too quick to say that such a character is perceptual as well.⁷ Thus, the overall meaning experience is not perceptual either; it only involves a *cognitive* form of phenomenology (Horgan and Tienson 2002; Pitt 2004, Strawson 1994, Chudnoff 2015). Simply, it is merely perceptually-based, since its first fold is admittedly perceptual.

In order to vividly grasp this point, consider first of all the phenomenon of *satiation*. Anyone has certainly experienced situations in which, by obsessively repeating a word (say, “fly”) one ends up uttering another word (say, “life”), or no word at all, but just

⁵ For the same morphosyntactic structure is compatible with different low-level visual or auditory properties of the expression.

⁶ For O’Callaghan (2011), a perceptual difference in phonological high-level properties already occurs in the case of a lexically ambiguous same-sounding expression. Brogaard (2018: 2978) admits that expression perception can be holistic, even though she would (erroneously, from our perspective) take this perception to affect expressions-*cum*-meaning.

⁷ Brogaard (2018: 2976,2979) would reply that the fact that, in their proper meaning fold, meaning experiences are evidence-resistant, since they persist notwithstanding the existence of defeaters, shows that such experiences are perceptual. But again, this is too quick. As Voltolini (2020a: 219) argues, an experience may be illusory even without being perceptual, as feelings clearly show. One may feel one’s limb as located in a certain position while knowing that it is not so, just as one may feel something as unreal while knowing that it is not such.

a mere noise. *Prima facie*, one may think that the phenomenal change at stake in such situations *perceptually* involves the semantic change experienced: first, one perceives the word in a certain meaning, second, one perceives the word in another meaning, or in no meaning at all. Yet this thought is wrong. For there definitely is a perceptual change in such situations, yet this change involves no semantic level, but only the morphosyntactic level qualifying the relevant word. As is proved by the fact that a similar perceptual change may occur also when *meaningless* words are involved: try e.g. with the meaningless “bly”, which will finally revert into the meaningless “libe” (or into a mere noise).

In the same vein, moreover, compare the difference between a structurally ambiguous yet lexically meaningless expression and a perceptually ambiguous picture. On the one hand, take the following well-known meaningless sentence from Lewis Carroll’s *Jabberwocky*:

(1) The slithy toves gyred the Jabberwock in the wabe.

From a morphosyntactical point of view, one can see or hear (1) in two different readings, depending on how one differently parses, viz. groups by differently holistically attending to it, the syntagms constituting it:

(1a) (The slithy toves in the wabe) (gyred (the Jabberwock)).

(1b) (The slithy toves) (gyred (the Jabberwock in the wabe)).

Yet, since no lexical meaning has been assigned to the nouns “toves” and “wabe”, the adjective “slithy” and the verb “to gyre”, neither (1a) nor (1b) has a lexically determined meaning. So a fortiori, no meaning can be read off from either (1a) or (1b). For perceiving those readings enables one to recognize no meaning in them. Granted, once meanings were conventionally assigned to the above words, one could experience different meanings in (1a) and (1b) respectively, just as one does with the sentence inaugurating Groucho Marx’s famous joke:

(2) Yesterday I saw an elephant in my pajamas.

One would then have two different twofold meaning experiences. Yet the meaning folds of such experiences would only be juxtaposed to the admittedly perceptual folds of such experiences in which one respectively grasps the different morpho-syntactic readings of (1), without any recognitional factor being involved. Hence, the meaning folds would not be perceptual. Thus, the resulting meaning experiences would not be perceptual, but merely perceptually-based.

The very same point can be made by appealing to lexically ambiguous sentences. Who claims that in:

(3) Dionysus is Greek.

one perceives the sentence’s name as meaning Dionysus the Elder, tyrant of Syracuse, will be troubled by discovering that, while perceiving exactly the same expressions (and possibly even having the very same mental images in mind), one can also experience that name in that sentence as meaning Dionysus the Younger, son of the preceding. Clearly, the two meaning experiences related to understanding this meaning difference are different as well (Siewert 1998; Horgan and Tienson 2002, O’Callaghan 2011). Yet no perceptual recognitional work, however mediated by attention, could allow one to experience this difference. One should only know by other means that the expression is ambiguous in order to experience its different meaning (cf. Martina and Voltolini 2017).

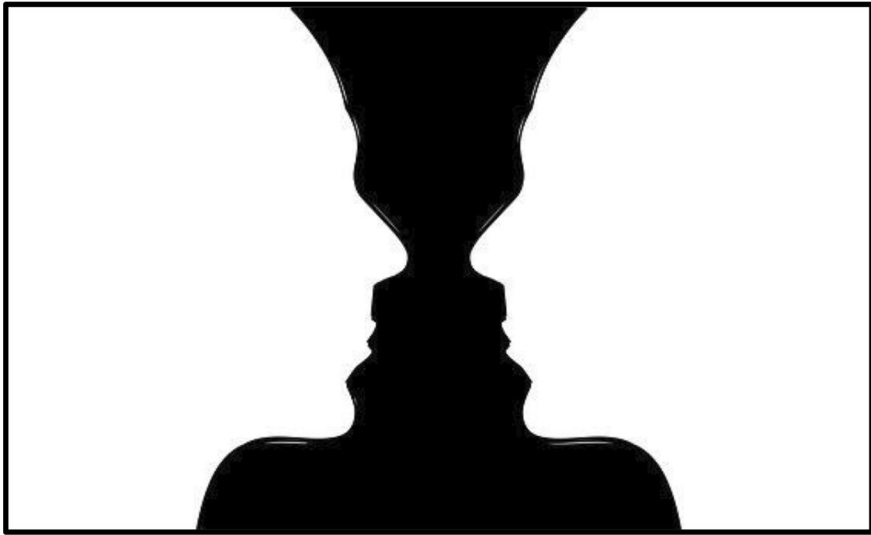


Fig. 1 The Rubin vase

On the other hand, take a perceptually ambiguous picture such as the Rubin vase (fig. 1) Depending on the different 3D figure-ground facewise and vase-wise segmentations of the very same mere picture's vehicle provided by differently attending that vehicle holistically, one can have different seeing-in experiences of that ambiguous picture such that one can read off the different subjects grasped in the respective recognitional folds of such experiences – namely, two white faces in profile on a black background vs. a black vase on a white background – from the respective configurational folds in which one respectively perceives those segmentations. Indeed, one can perceptually recognize such subjects respectively by virtue of such segmentations, so that those seeing-in experiences turn out to be perceptual as well.⁸

At this point, on behalf of meaning perceptualism one might remark that knowing the meaning of an expression induces a different perception of it. This remark would be correct, but only up to the extent that the new perception includes morphosyntactic features of that expression that were not included in one's original perception of it. If one knows, or even believes, that something has a certain meaning, one grasps that what one hears is not a mere noise, but (say) a morphosyntactically articulated sentence of a certain language. The import of that knowledge, or belief, would only be a form of *weak cognitive penetration*, in the sense defined by Macpherson (2012). Indeed, that knowledge, or belief, would basically induce a difference in the *phenom-*

⁸ Incidentally, one may remark that seeing-in experiences can be had utterly independently of meaning experiences. Non-human animals, for one, such as birds (Spetch and Weisman 2012) and primates (Fagot et al. 2010), appear to have the former while failing to have the latter experiences. However, some caution is needed on this point, for the ability of recognizing pictorial subjects in non-human animals greatly varies among species and appears to depend on critical factors such as presentation and training (for review, Bovet and Baclair 2000). Hence, further research is needed to conclusively establish the equivalence of seeing-in experiences in humans and non-human animals.

enal character of the perceptual experience involved, so that such an experience would definitely come to have a new content, yet utterly *non-conceptual*. One sees or hears an expression in certain morphosyntactic non-conceptualized features; namely, an item properly morphosyntactically grouped. In this respect, one may notice that also the configurational fold of a seeing-in experience is weakly cognitively penetrated in the very same sense. If one knows that what one is facing is a picture of a *Dalmatian*, one can entertain the non-conceptually relevant perceptual change in phenomenal character that transforms the perception of a mere 2D vehicle into the non-conceptual perception of a *pictorial* vehicle, featured by a 3D-like dalmatianwise item endowed with certain 3D grouping properties. However, this form of weak cognitive penetration does not amount to have a seeing-in experience as a whole yet. For this form does not yet mobilize the recognitional fold of that experience. Likewise, in letting one grasp only the morphosyntactic features of an expression, the form of weak cognitive penetration affecting the perception of that expression does not mobilize a meaning experience as a whole, but only its first, genuinely perceptual, fold. So, rightly observing that such a perception is weakly cognitively penetrated says nothing in favor of the perceptual character of the meaning experience as a whole.

Following McDowell (1998), however, someone might still reply that a meaning experience can be a form of non-sensory extended perception. Now definitely, the notion of a non-sensory form of perception is highly problematic. It is quite disputable, for example, whether intellectual intuition is perceptual in more than a metaphorical sense. How we just characterized perception of high-level properties does not allow intellectual intuition to be ranked as perceptual (see also Chudnoff 2015). For although intellectual intuition may be immediate, it is not imbued with holistic attention towards its object's properties. Granted, there is a form of perception that is admittedly non-sensory; namely, amodal perception. Yet we are unclear as to how a meaning experience could be a form of amodal perception. The paradigmatic cases of amodal perception are those in which parts of objects that are otherwise sensorily grasped are occluded from other such parts, so that they are grasped by no sensory modality; the dark side of the Moon, for example. Yet no such phenomenon occurs in the case of a meaning experience. The meaning of an expression is not something that the sensorily given features of that expression occlude, in any plausible sense of the term.

All in all, we can stress that the seeing-in experiences are *proper* fusion experiences, in which the overall experience is different from the sum of its parts (Stumpf 1890). For, as Wollheim intuited, its folds are compenetrated, by no longer being identical with the respective experiences of the picture's vehicle and of the picture's subject taken in isolation.⁹ By contrast, meaning experiences are not proper fusion experiences. For their second experiential fold is simply juxtaposed to its first, admittedly perceptual, fold, in its being not readable off from that fold by virtue of a content matching.

⁹ What makes the recognitional fold different from the face-to-face perception of the picture's subject taken in isolation may be that, unlike the latter, the former is a *strongly* cognitively penetrated perception, i.e., a cognitively penetrated perception whose penetration determines the conceptual character of its content. For, unlike the face-to-face perception of the picture's subject, in the seeing-in experience one needs concepts in order to single out in its recognitional fold that subject as a 3D scene differing from the 3D-like silhouettes grasped in its configurational fold. Cf. Wollheim (2003a), Voltolini (2015, 2020b).

3 There is a Common Semantic System for Seeing-In Experiences and Meaning Experiences, but only in Seeing-In Experiences the Semantic Access is Perceptual

In the previous Section, we have advanced, on a purely phenomenological basis, a series of philosophical considerations in support of the idea that meaning experiences and seeing-in experiences are typologically different, that is, are not experiences of the same kind. Unlike meaning experiences, seeing-in experiences are recognitional experiences of a sort that makes them perceptual experiences. In our opinion, we can arrive at the same conclusion if we consider empirical data in addition to phenomenological intuitions.¹⁰ In this respect, relevant questions are: Do seeing-in experiences and meaning experiences differ in timing and patterns of activation in the human brain? How do these differences (if any) relate to the nature of these experiences? As we will see, our philosophical considerations are strongly consistent with behavioral and neuroscience data.

In order to argue for this result, a first step is to show that the distinction between the two experiences is cognitively real, that is, that the two experiences are underpinned by distinct dimensions of the cognitive/neural architecture.

Against this hypothesis, however, a defender of the typological commonality might immediately rebut that, as regards their high-level aspects (that is, the recognitional fold and the proper meaning fold, respectively), there is a close relationship between those experiences in the human brain. In cognitive neuroscience, it is standardly believed that meaning experiences and seeing-in experiences ultimately converge within a shared central semantic store, a depository of conceptual representations that is equally accessible by linguistic expressions and picture forms (fig. 2). Evidence for a shared semantic system comes from observations that lesions in some cortical areas produce remarkably similar high-level deficits in both seeing-in and meaning experiences, as in the case of the patients affected by semantic dementia, who consistently show significant atrophy of the anterior temporal lobes of both hemispheres (Lambon Ralph et al. 2017a, b). Support for this hypothesis also comes from neuroimaging studies that have contrasted neural activity during semantic tasks performed either with linguistic expressions or with pictures (e.g., Vanderberghe et al. 1999, Moore and Price 1999, Bright et al. 2004 see also Binder et al. 2009). Using conjunctive analyses, these studies found robust semantic activation for both seeing-in experiences and meaning experiences in an extensive network of associative (i.e., modality-independent) areas in the left hemisphere, covering large sections of frontal and temporal regions. Yet, according to advocates of the Simulation Framework (e.g., Barsalou 1999, 2016), also called “neo-empiricism” (Prinz 2002), the common semantic system also extends to sensorimotor cortices. Within this framework, the access to the high-level proper meaning of concrete, high imageable words and sentences is supposed to re-activate regions of the brain that are involved in direct perception, such as the visual cortex (Kemmerer 2010).

¹⁰ According to a common methodological assumption (e.g. Drożdżowicz 2019), the nature of mental experiences cannot be decided merely on the basis of the phenomenology alone. Philosophical theories about mental experiences should be at least minimally not in contradiction with our empirical knowledge about them.

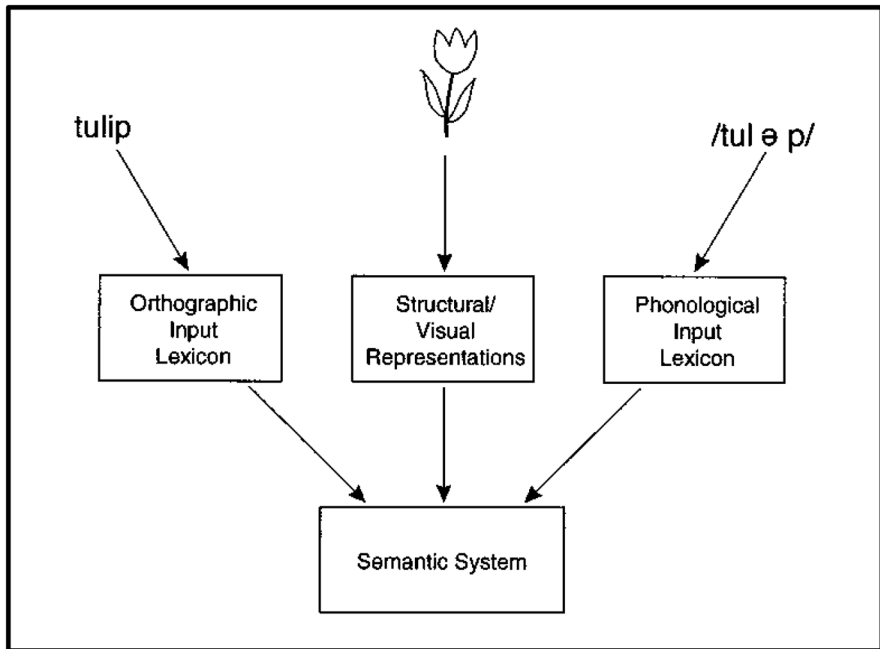


Fig. 2 A graphic representation of the semantic system (modified from Hillis & Caramazza 1991)

Although a common semantic system might be activated similarly during seeing-in experiences and meaning experiences, however, there is also evidence that these two semantic access routes are significantly different.

On the one hand, decades of empirical research have shown that the first fold of meaning experiences is supported by highly specialized neural structures in the visual and the auditory cortex, as well as in the associative cortex, with different underpinnings for orthographical (e.g., Dehaene and Cohen 2011), phonological (e.g., Lieberthal et al. 2005), and morphosyntactical (e.g., Matchin and Hickock 2020) processing. These neural structures appear to have no role in pictorial perception.

On the other hand, picture perception is known to be underpinned by a hierarchically organized perceptual stream that encodes progressively more complex information about the depicted objects and scenes, as well as information about the surface properties of the picture's vehicle (Nanay 2011; Ferretti 2018; Vishwanath 2014). It is unlikely that all this perceptual information is codified in linguistic expressions and mobilized during meaning experiences.¹¹ At present, *contra* the Simulation Framework, it is not even established that re-activation of a detailed perceptual representation of words' referents is necessary for language comprehension, at least not to the same degree as that activated during actual object recognition or seeing-in experiences (Calzavarini 2017; Mahon and

¹¹ For example, it is known that some specific structures in the ventral and dorsal visual streams are functionally dedicated to the extraction of 3D information from the visual scene (Welchman 2016). As we have seen, this process is a critical determinant of seeing-in experiences, allowing the grasping of a 2D object like a painting or a drawing in its pictorial dimension (the configurational fold). Certainly, this low-level aspect of seeing-in experiences have no direct correspondence in meaning experiences.

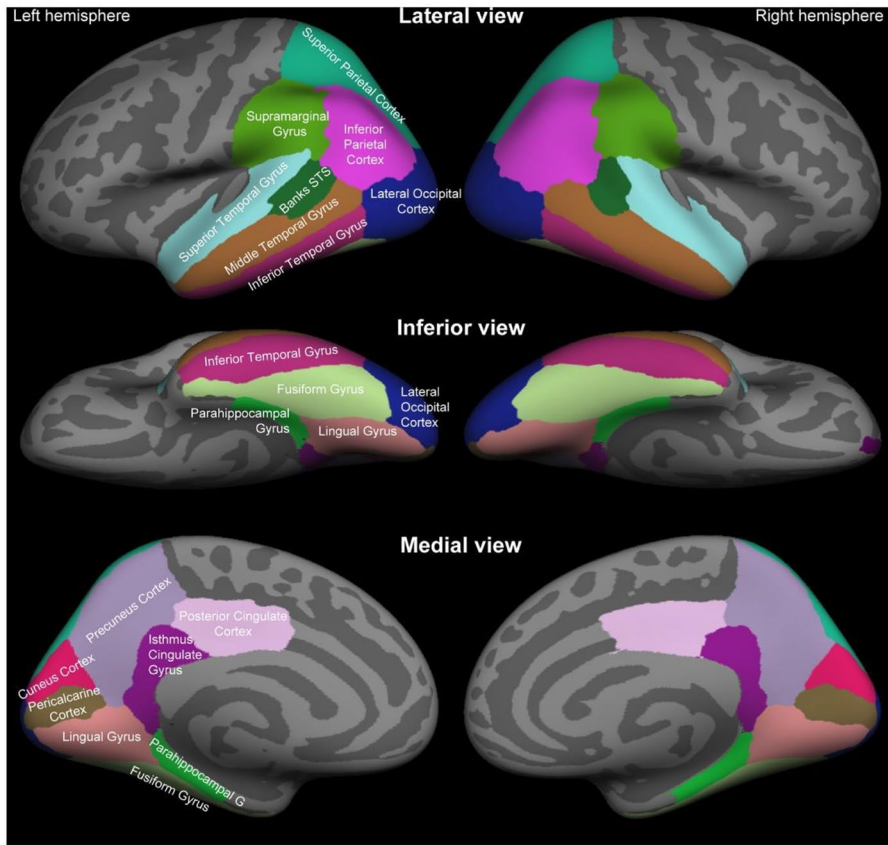


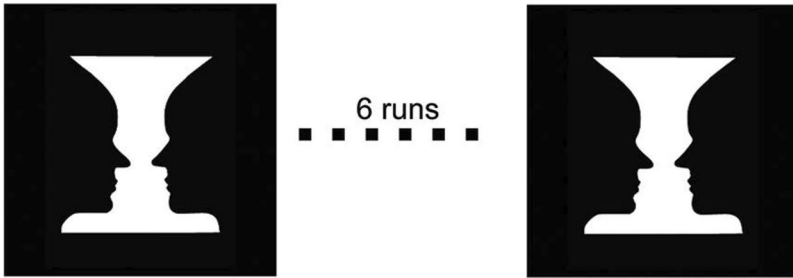
Fig. 3 Standard anatomical parcellation of the posterior section of the human brain. The fusiform gyrus is in light grey (from Ahveninen et al. 2012)

Caramazza 2008, 2009).¹² In addition, even if a portion of the visual cortex that underpins picture perception is re-activated during meaning experiences, this visual activation involves mainly top-down rather than bottom-up cognitive mechanisms, differently from pictorial perception.¹³

¹² The necessity or dispensability of the sensorimotor cortex for language understanding is a topic of intense debate in the cognitive neuroscience literature. On the one hand, in few cases lesions in (e.g.) the visual cortex have been associated with corresponding impairments in understanding concrete, visual words (Forde et al. 1997, Manning et al. 2000). These impairments, however, tend to be subtle and evident only in particular experimental conditions. On the other hand, many patients have been reported with lesions in the visual cortex and perfect semantic performances with both concrete and abstract words (e.g., Behrmann et al. 1994, Carlesimo et al. 1998; for review, Calzavarini 2017, 2020). These latter data support the hypothesis, *contra* the Simulation Framework, that the re-activation of visual neural structures is not a critical and necessary component of meaning experiences.

¹³ According to the Simulation Framework, access to the high-level visual cortex during meaning experiences is a top-down process mediated by “cross-modal convergence zones” in high-order, associative regions (see, e.g., Simmons and Barsalou 2003). For this reason, it is very different from access during seeing-in experiences (and visual experiences more generally), which is mainly bottom-up. Accordingly, advocates of the Simulation Framework consider language comprehension as a form of (conscious or unconscious) mental imagery, rather than a kind of perceptual achievement (Barsalou 1999).

A *Ambiguous condition*: subjects pressed 2 keys to report their subjective experiences of face or vase while scanning.



B *Unambiguous condition*: blocked design, 1-back task



Fig. 4 The experimental paradigm in the study of Wang et al. (2017). See the text for details

The existence of brain-damaged patients with profound deficits in seeing-in experiences but relatively intact meaning experiences, and vice versa, also argues against a complete overlap between the neural architectures underlying these two experiences. For example, patients with auditory verbal agnosia (Buchman et al. 1986) and deep dyslexia (Coltheart et al. 1980) are impaired on (auditory or visual) verbal understanding tasks but can normally perform on pictorial perception. Reciprocally, in several cases (Farah 2004), patients with visuoperceptual impairments showed severe pictorial impairments but achieved a normal level of verbal understanding on both spoken and written verbal tasks. Critically, several cases of semantic dementia patients have been observed whose temporal lobe atrophy was significantly more marked either on the left hemisphere or on the right hemisphere, and whose performance was disproportionately impaired in either seeing-in experiences or meaning experiences (for review, Gainotti 2012). In general, patients with left hemisphere atrophy tend to perform significantly worst on semantic tasks involving linguistic expressions as compared to pictures, while patients with right hemisphere atrophy tend to show the opposite pattern (e.g., Lambon Ralph and Howard 2000, Butler et al. 2009, Mion et al. 2010).

Considering these functional dissociations in accessing meaning from linguistic expressions as compared to pictures, several scholars have hypothesized that multiple semantic stores exist and that the pictorial and verbal access to the semantic system might be neuroanatomically segregated (e.g., Paivio 1986, Gainotti 2012, Hurley et al. 2018). In our opinion, neuroimaging studies are also consistent with this typological difference hypothesis. While several studies have been interpreted as supporting the common semantic system hypothesis, as we have argued above (e.g., Moore and Price 1999), all of these studies have reported some specific effects for seeing-in experiences and meaning experiences in addition to the regions of common activation, with a clear asymmetry between left and right hemispheres. On the one hand, meaning experiences have been associated with selective activation of the left superior and middle temporal lobes. On the other hand, seeing-in experiences increase activation in some ventral temporal regions of the right hemisphere, particularly the posterior and middle sections of the fusiform gyrus (e.g., Vandenberghe et al. 1996).¹⁴

Given the above dissimilarities, meaning experiences and seeing-in experiences are better construed as not the same kind of experience at the cognitive level. Specifically, empirical data are consistent with functional and anatomical differentiation along the way that pictures and linguistic expressions access their respective, too hastily hypothesized to be common, high-level experiential folds. But there is more than that. In our opinion, and this is our fundamental point here, there is evidence that, unlike verbal expressions, pictures access their high-level experiential folds via perceptual and recognitional cognitive resources.

In order to grasp this point, first consider that a long tradition of behavioral studies (e.g., Paivio 1986) and studies using the electrophysiological (ERP) technique (e.g., Leonardelli et al. 2019, Shaul and Rom 2019) has experimentally demonstrated that pictorial stimuli more readily contact the semantic system as compared to linguistic expressions. This “picture superiority” effect is generally believed to be an established finding in the literature about semantic memory activation. For instance, ERP studies that have directly contrasted meaning experiences and seeing-in experiences have reported that conceptual access for linguistic expressions is delayed by about 90 msec with respect to pictures (Leonardelli et al. 2019). As noted by Shaul and Rom, «the main processing of pictures happens during the first 300 msec, while the subject perceives the visual features of the figure. This processing may be enough to reach the meaning in pictures, but words need additional processing which happens later (between 400 and 500 msec) in order to reach the semantic presentation of the word» (2019: 249). This timing profile suggests that, although both seeing-in experiences and meaning experiences appear to be immediate at the phenomenological level, significant differences exist at the cognitive level: reading-off the pictorial subject from a pictorial’s vehicle is relatively faster, in cognitive terms, than grasping the meaning of a linguistic expression in the proper meaning experiential fold. Accordingly, several scholars have argued that pictures have a faster and more direct (“privileged”)

¹⁴ Other neuroimaging studies have reported a similar anatomical differentiation, with partly dissociated areas of activation for verbal and pictorial semantic tasks (e.g., Gates and Yon 2005, Reinholz and Pollman 2005).

access to their high-level semantic fold, while words and sentences are interpreted to require additional translation at the representational level before accessing the semantic system (e.g., Hillis and Caramazza 1990).¹⁵

Admittedly, a defender of the typological community might insist that these findings by themselves do not conclusively establish that seeing-in experiences are perceptual in nature. To be sure, immediacy is merely a necessary but not sufficient condition for an experience to be perceptual (Martina and Voltolini 2017; Nes 2016). Yet, critically, unlike meaning experiences, in seeing-in experiences semantic access is mediated by neural structures that have been independently associated with perceptual recognition.

Let us analyze this point more in detail. As outlined above, there is a clear left-right hemisphere asymmetry in the neural underpinnings of the two kinds of experiences. As is known, a dominant view that emerges from decades of experimental research in neuropsychology and neuroscience is that the left hemisphere is specialized for amodal and language processes, whilst the right hemisphere is specialized for visual object recognition (e.g., Gazzaniga 2000). This general trend reinforces the conjecture that, unlike meaning experiences, the overall seeing-in experience is perceptual in character.

More specific evidence comes from neuroimaging studies. On the one hand, as we have seen, semantic processing of pictorial stimuli selectively activates the right fusiform gyrus (e.g., Bright et al. 2004), a region in the secondary visual cortex which is known to be involved in the processing of high-level visual information (Palejwala et al. 2020). Since a focal lesion in this area appears to be sufficient for generating visual recognition disorders (Konen et al. 2011), it has been suggested that the right fusiform gyrus is the main cortical substrate of the structural description system (Zannino et al. 2011) (fig. 3). According to most models of visual processing (Marr 1982, Humphreys and Forde 2001), the «structural description system represents the highest level in the visual processing stream, where incoming percepts match structural representations before accessing the semantic system» (Zannino et al. 2011: 2878). On the other hand, the semantic processing of linguistic expressions selectively engages some traditional language areas of the left temporal lobes, such as the posterior middle temporal lobe (e.g., Vandenberghe et al. 1996). According to an influent neurocognitive model of language comprehension (Hickock and Poeppel 2007), this cerebral region serves as an associative, non-perceptual interface that «maps between phonological-level representations of words or morphological roots and distributed conceptual representations» (Hickock and Small 2015: 304).¹⁶

To sum up: empirical knowledge from cognitive neuroscience appears to vindicate the phenomenologically-based philosophical considerations we have provided in Sect. 1. From the cognitive point of view, a seeing-in experience is typologically different from a meaning experience because of its specific perceptual way of being a recognitional experience.

¹⁵ According to some scholars, the privileged access hypothesis is also supported by the observation that semantic performances with words, but not pictures, is affected by the frequency with which the corresponding item (i.e., the referent or the depicted object) is encountered (Taikh et al. 2015).

¹⁶ Actually, the vast majority of psycholinguistic models assume that the phonological and the semantic systems are functionally independent, and that there is an associative, non-perceptual link between the two systems (e.g., Levelt 2001). Even most of the neural models developed in the Simulation Framework postulate associative, Hebbian-like connections between phonological representations in the perisylvian cortex and semantic representations in the sensorimotor cortices (e.g., Pulvermuller 1999).

4 Unlike Meaning Experiences, in Seeing-in Experiences it is Possible to Read off the High-Level Content Because of Their Perceptual (Neural) Basis

In light of its putative perceptual nature, one might expect that the link between the first and the second fold in seeing-in experiences is more robust and less susceptible to brain damage as compared to what happens in meaning experiences. Only seeing-in experiences, we have argued, are proper fusion experiences. Interestingly enough, neuropsychological evidence appears to provide some support for this conjecture. On the one hand, in meaning experiences, the access to the proper meaning fold can sometimes be impaired after brain damage without this affecting the perception not only of phonological but also of morpho-syntactic properties (*morpho-syntactically enriched expression fold without proper meaning fold*). A notable example of this condition are patients affected by transcortical sensory aphasia, a neuropsychological syndrome that is supposed to «result from result from a one-way disruption between left hemisphere phonology and lexical–semantic processing» (Boatman et al. 2000: 1634). On the other hand, it is very rare that a brain-damaged patient knows that a certain object is a picture (in its 3D organization) without being able to access what the picture represents, i.e., the picture’s subject (*configurational fold without recognition fold*).¹⁷ This observation supports the philosophical intuition that, unlike meaning experiences, in seeing-in experiences the high-level aspect is not juxtaposed to the low-level aspect, but is intimately connected to it.

In our opinion, the typological difference between seeing-in experiences and meaning experiences is further supported by an analysis of the neural underpinnings of perceptual ambiguity and lexical ambiguity. As we will see, such analysis clearly suggests that the two processes are differentiated in the human brain.

Against this hypothesis, a defender of the typological commonality might immediately rebut that that the perception of ambiguous figures such as the Necker cube or the Rubin vase, on the one hand, and the perception of lexically ambiguous words such as “bank” or “pole”, on the other, tend to activate a similar network of high-order neural structures in frontal, temporal, and parietal lobes (for reviews, Brascamp et al. 2018 and Vitello and Rodd 2015, respectively). The inferior frontal gyrus, a neural structure that has been implicated in attentive and executive functions in many studies, has consistently shown increased activation for both ambiguous pictures (e.g., Knapen et al. 2011) and words (e.g., Rodd et al. 2005). This region has been indicated as one of the most likely candidates for playing a critical role in both perceptual transitions and lexical ambiguity resolution, suggesting a close relationship between the top-down cognitive resources necessary for shifting between different semantic readings of words and pictures. This communality is also suggested, one might argue, by the advantages of bilingual children in understand-

¹⁷ A defender of the typological commonality might reply that patients affected by associative visual agnosia (Farah 2004) are a clear counterexample to this claim. For in such syndrome visual processes are preserved, but patients are still unable to recognize what a picture represents. However, careful analyses suggest that most of these patients have significant impairments in extracting 3D information from pictorial representations such as photographs, drawings, and even stereoscopic computerized images (for review, Chainay and Humphreys 2001, Hellar 2019). Thus, the pictorial impairment in this syndrome seems to be located at the level of the implementation of the configurational fold, put in Wollheim’s terms, rather than at the level of the implementation of the recognitional fold.

ing perceptual figures (e.g., Bialystok and Shapero 2005, Chung-Fat-Yim et al. 2017). This advantage might indicate the existence of common selection/inhibition attentional processes involved in both picture perception and language understanding.

Yet, these similarities notwithstanding, the neural underpinnings of ambiguous pictures and ambiguous words are clearly dissociated, with a significant hemispheric asymmetry characterizing the fronto-temporo-parietal network involved in the two processes. On the one hand, seeing-in experiences involving ambiguous pictures tend to activate right hemisphere regions (Brascamp et al. 2018). On the other hand, meaning experiences with ambiguous words are clearly left-lateralized (e.g., Hoffman and Tamm 2020).

More importantly, there is evidence that pictorial ambiguity belongs to the broader class of perceptual phenomena, while lexical ambiguity is better considered as a full high-order, cognitive process. This evidence reinforces the intuition that, unlike meaning experiences, seeing-in experiences have a perceptual nature.

To illustrate this claim, we may first rely on evidence about the time course of the cognitive shifting between different readings of ambiguous pictures and ambiguous words. On the one hand, studies using the ERP technique have revealed an early neural signal correlated with endogenous reversals of ambiguous pictures, called “reversal positivity”, which appears 130 msec after stimulus onset at occipital positions, where the early visual cortex is located (review in Kornmeier and Back 2012). The existence of this early neural signal, which has been observed for a range of ambiguous pictures such as the Necker cube, the Necker lattice, and the Old/Young woman, strongly suggests that perceptual reversals can be initiated during the first visual processing step – although high-order cognitive processes can modulate it at later stages (Abdallah and Brooks 2020).

This timing profile is certainly compatible with the involvement of perceptual, bottom-up mechanisms in seeing-in experiences with ambiguous pictures. As is well known, the existence of passive, sensory-like cognitive processes in ambiguous picture perception is confirmed by a number of traditional findings, such as the observed patterns of reversals over time (which suggest the automaticity and fatigue-like nature of this process), or the presence of adaptational effects in perceptual ambiguity (for a review, Long and Toppino 2004). Under certain accounts, alternations in ambiguous figures result from mutual inhibition/suppression processes between separate pools of neurons located in the visual cortex, each representing the information pertaining to one of the two (or more) perceptual interpretations of those figures (e.g., Toppino and Long 1987).¹⁸ This might explain why, as noted by Block (2014: 567), alternate experiences in ambiguous figure perception are characterized by *exclusivity* (they are not given simultaneously), *inevitability* (one way of seeing the faced object will eventually replace another), and *randomness* (the duration of one alternative experience is not a function of previous duration).

On the other hand, empirical data shows that the shifting between different meaning folds of the same word is a significantly slower process. The standard conception in the lexical ambiguity literature is that, when listening to an ambiguous word, the different meanings are simultaneously accessed, and a first semantic selection is made starting from 200 msec from stimulus onset (Vitello and Rodd 2015). If new information

¹⁸ Granted, this does not mean that top-down cognitive components such as attention, beliefs, or expectation have no role in the processing of ambiguous pictures. Today, the vast majority of neurocognitive models of perceptual ambiguity have a dual or hybrid architecture, in that they recognize the interactions between top-down and bottom-up mechanisms (e.g., Long and Toppino 2004, Knapen et al. 2011).

is acquired that is inconsistent with this interpretation, the word must be reinterpreted. Experimental research suggests that semantic reinterpretation is a cognitively demanding process, as demonstrated by several behavioral processing costs (e.g., Rodd et al. 2010). It is commonly believed that these costs are associated with longer times for inhibiting the contextually inappropriate meaning of the ambiguous word (for example, the dog-meaning of “bark”) and (re)activate the contextually appropriate meaning (e.g., the tree-meaning). ERP studies have demonstrated that this shifting in meaning experiences starts at least 800 msec after the onset of the disambiguating cue (MacGregor et al. 2020). Indeed, this timing profile is not compatible with the involvement of sensory, bottom-up mechanisms in the semantic processing of ambiguous words. Accordingly, dominant models of lexical ambiguity resolution (e.g., Duffy et al. 2001) postulate a combination of higher-order, top-down factors involved in meaning selection and semantic reinterpretation, such as contextual knowledge or knowledge about meaning frequency. This timing profile is also at odds with the idea that meaning experiences are characterized by a specifically perceptual form of immediacy (Nes 2016; Brogaard 2018).

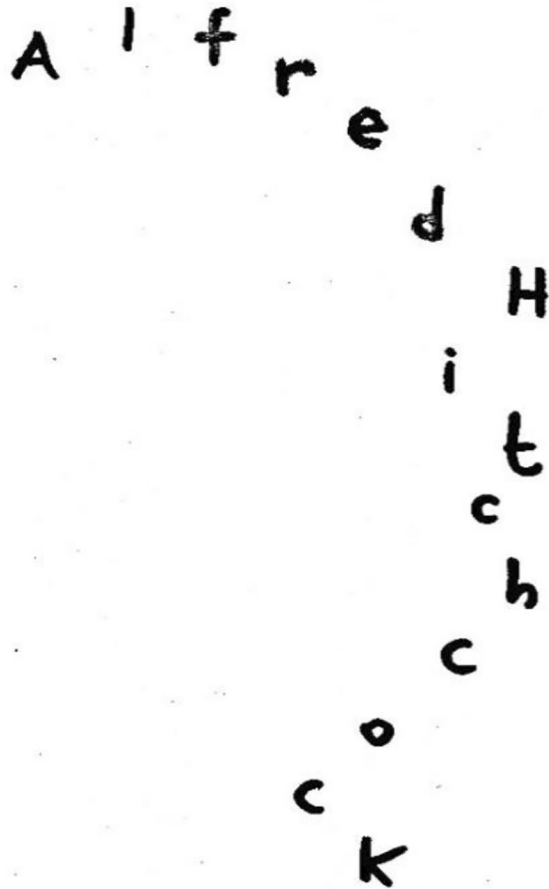
Again, a defender of the typological community might insist that these findings by themselves do not conclusively establish that seeing-in experiences with ambiguous figures or pictures are a perceptual phenomenon. As we have said, immediacy is merely a necessary but not sufficient condition for an experience to be perceptual.

Yet, numerous neuroimaging studies using standard, univariate fMRI have demonstrated that neural activity in both primary and secondary visual cortex correlates with the content of alternative interpretations of ambiguous pictures in seeing-in experiences (review in Sterzer et al. 2009). To take a paradigmatic case, when subjects are presented with the Rubin vase in the fMRI scanner, visual regions in the fusiform gyrus that are known to be selective for faces (e.g., the “face fusiform area”) show increased activation during face-interpretations as compared to vase-interpretations (Andrews et al. 2002). Similarly, studies using magnetoencephalography (MEG) have demonstrated that behavioral reports of alternative face and vase interpretations of the Rubin vase correlate with activity in the early visual cortex (Parkkonen et al. 2008).

In principle, these correlations may be caused by other experimental manipulations rather than the picture’s content (for example, the greater visual effort requested by processing faces as compared to objects). This is because neuroimaging techniques such as univariate fMRI or MEG are only sensitive to quantitative variations in the hemodynamic or electrical activity of the brain, and not to neural information *per se*.¹⁹ In a recent study, however, Wang et al. (2017) used multivoxel pattern analysis (MVPA; see Norman 2006) to further explore the hypothesis that visual regions, in their activity patterns, carry information about fluctuating content during perception of ambiguous pictures. In the experimental condition (*ambiguous condition*), the subjects were presented with the

¹⁹ According to the received view in the neuroimaging literature, standard univariate fMRI methods, although extremely useful, have limitations in decoding the information that is represented in a subject’s brain (e.g., Norman et al. 2006). For these methods are only able to analyze voxels (i.e., the minimum unit of neuroimaging analysis) in isolation. In contrast to univariate methods, multivoxel pattern analysis is a technique that analyzes neuroimaging data by considering the pattern of BOLD responses across many voxels simultaneously. MVPA is supposed to decode the category of a stimulus from these patterns revealing the representations that a brain region contains (Norman et al. 2006). For example, «a brain region in which faces of men and women elicit distinct multivoxel patterns but faces of the same sex yield similar patterns may represent sex» (Contreas et al. 2013: 1).

Fig. 5 Hitchcock's nominal silhouette (Voltolini 2015)



Rubin vase and were asked to report, by pressing one of two buttons, any alternations between face and vase interpretations as soon as it was perceived. In the control condition (*unambiguous condition*), the subjects were presented with unambiguous black and white photographs of faces and vases (fig. 4). Results of this study confirm that activity patterns in the early visual cortex and the face-selective regions in the fusiform gyrus are sufficient to discriminate between facewise and vasewise segmentation of the Rubin vase. In other words, it is possible to use activity patterns in these visual regions to predict which of the two alternative perceptual contents (face or vase) is activated.

The considerations above suggest that, in seeing-in experiences, the recognitional fold can emerge from neural activity in the perceptual regions of the brain. As regards meaning experiences, this does not seem to be the case.

Hoffman and Tamm (2020), for instance, have recently used a combination of univariate and multivariate (MVPA) fMRI analyses to investigate the brain regions involved in the processing of balanced ambiguous words (that is, ambiguous words in which neither meaning was highly dominant over the other). Results of the multivariate analysis study showed that different frontal and temporal regions in the left hemisphere could

discriminate between the presentation of the same words in different semantic contexts (e.g., “bark” following “tree” vs “bark” following “dog”). Thus, neural activity in these areas could be used to reliably predict which of the ambiguous word’s proper meanings was grasped by the subjects. Critically, however, all of these regions are supposed to be high-level, associative nodes distant from primary sensory cortices. For instance, one of the regions highlighted in the study was the left anterior temporal lobe. This region is modulated by conceptual processing independently to the input modality (Lambon Ralph et al. 2017a, b). Due to its multimodal neurofunctional profile, it has been suggested that the anterior temporal lobe constitutes a supramodal or amodal “hub” where conceptual information is distilled and represented in non-modal form (Patterson et al. 2007).

Again, empirical knowledge from cognitive neuroscience appears to vindicate the phenomenologically-based philosophical considerations we have provided in Sect. 1. Only in seeing-in experiences, we have seen, it is possible for the high-level fold to emerge from its perceptual (neural) basis. This reinforces the intuition that, unlike the recognitional fold, the second fold of a meaning experience is definitely experiential, but not perceptual in character.

5 Conclusion

To sum up. Both phenomenological considerations and available data in cognitive neuroscience supports the claim that, although they seem very similar, seeing-in experiences and meaning experiences are typologically different. Only when understood in a pictorial way, representations elicit a specific perceptual phenomenology and recruit specific perceptual resources of the brain. Indeed, as Goodman (1968) originally suggested, there is nothing in the representation itself that makes it pictorial or non-pictorial (verbal); everything depends on the representational system it is understood to belong to. In order to illustrate this point, let us consider the following nominal silhouette (cf. Voltolini 2015) (fig. 5):

In this arrangement, the mark “Alfred Hitchcock” can be naturally understood either as a word or as a picture of the famous British director. Thus, in the case of nominal silhouettes, the same representation can elicit both a seeing-in experience and a meaning experience. Given the results of this article, one should expect not only those cerebral areas that respectively implement such experiences are (at least partly) different, but also that only during the pictorial reading of nominal silhouettes semantic access is supported by perceptual regions of the brain. Interestingly, no experiment has directly contrasted neural activity during pictorial vs. verbal readings of nominal silhouettes. Further experimental research in this area might shed light on this issue, providing more support for the typological difference between seeing-in experiences and meaning experiences.²⁰

Acknowledgements The paper has been presented at a MUMBLE seminar at the University of Turin, June 30, 2021. We thank all the participants for their inspiring remarks.

²⁰ Although the paper has been conceived and elaborated together, Author 1 is specifically responsible of Sects. 2–3, while Author 2 is specifically responsible of Sect. 1.

Funding Open access funding provided by Università degli Studi di Torino within the CRUI-CARE Agreement.

Declarations

Conflict of Interest None.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abdallah, D., and J. L. Brooks. 2020. "Response dependence of reversal-related ERP components in perception of ambiguous figures". *Psychophysiology* 57: e13685.
- Ahveninen, J., Jääskeläinen, I. P., Belliveau, J. W., Hämäläinen, M., Lin, F. H., & Raij, T. (2012). Dissociable influences of auditory object vs. spatial attention on visual system oscillatory activity. *PLoS one*, 7(6), e38511.
- Andrews, T., D. Schluppeck, D. Homfray, P. Matthews, and C. Blakemore. 2002. "Activity in the fusiform gyrus predicts conscious perception of Rubin's vase-face illusion". *Neuroimage* 17: 890–901.
- Barsalou, L. 1999. "Perceptual symbol systems". *The behavioral and brain sciences* 22: 577–609, 660.
- Barsalou, L. 2016. "On staying grounded and avoiding quixotic dead ends". *Psychon bull rev* 23: 1122–1142.
- Behrmann, M., M. Moscovitch, and G. Winocur. 1994. "Intact visual imagery and impaired visual perception in a patient with visual agnosia". *Journal of experimental psychology* 20: 1068–1087.
- Bialystok, E., and D. Shapero. 2005. "Ambiguous benefits: the effect of bilingualism on reversing ambiguous figures". *Developmental science* 8: 595–604.
- Binder, J., R. Desai, W. Graves, and L. Conant. 2009. "Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies". *Cerebral cortex* 19: 2767–2796.
- Boatman, D., B. Gordon, J. Hart, O. Selnes, D. Miglioretti, and F. Lenz. 2000. "Transcortical sensory aphasia: revisited and revised". *Brain* 123: 1634–1642.
- Bovet, D., and J. Vauclair. 2000. "Picture recognition in animals and humans". *Behavioural brain research* 109: 143–165.
- Brascamp, J., P. Sterzer, R. Blake, and T. Knapen. 2018. "Multistable perception and the role of the frontoparietal cortex in perceptual inference". *Annual review of psychology* 69: 77–103.
- Bright, P., H. Moss, and L. Tyler. 2004. "Unitary vs multiple semantics: PET studies of word and picture processing". *Brain and language* 89: 417–432.
- Brogaard, B. 2018. "In defense of hearing meanings". *Synthese* 195: 2967–2983.
- Buchman, A., D. Garron, J. Trost-Cardamone, M. Wichter, and M. Schwartz. 1986. "Word deafness: one hundred years later". *Journal of neurology neurosurgery and psychiatry* 49: 489–499.
- Butler, C., S. Bramati, B. Miller, and M. Gorno-Tempini. 2009. "The neural correlates of verbal and non-verbal semantic processing deficits in neurodegenerative disease". *Cognitive and behavioral neurology* 22: 73–80.
- Calzavarini, F. 2017. "Inferential and referential lexical semantic competence: a critical review of the supporting evidence". *Journal of neurolinguistics* 44: 163–189.
- Calzavarini, F. (2020). *Brain and the lexicon*, Springer, Berlin.
- Calzavarini, F., and A. Voltolini. 2022. "Perception of faces and other progressively higher-order properties". *Topoi*. <https://doi.org/10.1007/s11245-022-09802-4>.

- Caramazza, A., and A. Hillis. 1990. "Where do semantic errors come from?". *Cortex; A Journal Devoted To The Study Of The Nervous System And Behavior* 26: 95–122.
- Carlesimo, G., P. Casadio, M. Sabbadini, and C. Caltagirone. 1998. "Associative visual agnosia resulting from a disconnection between intact visual memory and semantic systems". *Cortex; A Journal Devoted To The Study Of The Nervous System And Behavior* 34: 563–576.
- Chainay, H., and G. Humphreys. 2001. "The real-object advantage in agnosia: evidence for a role of surface and depth information in object recognition". *Cognitive neuropsychology* 18: 175–191.
- Chudnoff, E. 2015. *Cognitive phenomenology*. London: Routledge.
- Chung-Fat-Yim, A., G. B. Sorge, and E. Bialystok. 2017. "The relationship between bilingualism and selective attention in young adults: evidence from an ambiguous figures task". *Quarterly journal of experimental psychology* 70: 366–372.
- Coltheart, M., K. Patterson, and J. Marshall. 1980. *Deep dyslexia*. London: Routledge and Kegan Paul.
- Contreras, J., M. Banaji, and J. Mitchell. 2013. "Multivoxel patterns in fusiform face area differentiate faces by sex and race". *PLoS one* 8: e69684.
- Dehaene, S., and L. Cohen. 2011. "The unique role of the visual word form area in reading". *Trends in cognitive sciences* 15: 254–262.
- Drozdowicz, A. 2019. "Do we hear meanings? – between perception and cognition". *Inquiry* 1–33.
- Duffy, S. A., G. Kambé, and K. Rayner. 2001. "The effect of prior disambiguating context on the comprehension of ambiguous words: evidence from eye movements". In *On the consequences of meaning selection: perspectives on resolving lexical ambiguity*, ed. S. Gorfein, 27–43. Washington: American Psychological Association.
- Farah, M. 2004. *Visual agnosia*. Cambridge MA: MIT Press.
- Fagot, J., R. K. Thompson, and C. Parron. 2010. "How to read a picture: lessons from nonhuman primates". *Proceeding of the national academy of science U.S.A* 107: 519–520.
- Ferretti, G. 2018. "The neural dynamics of seeing-in". *Erkenntnis* 84: 1285–1324.
- Forde, E.M.E., Francis, D., Riddoch, M.J., Rumiati, R.I. & Humphreys, G.W. (1997) On the links between visual knowledge and naming: A single case study of a patient with a category-specific impairment for living things, *Cognitive Neuropsychology*, 14 (3), pp. 403–458.
- Gainotti, G. 2012. "The format of conceptual representations disrupted in semantic dementia: a position paper". *Cortex; A Journal Devoted To The Study Of The Nervous System And Behavior* 48: 521–529.
- Gates, L., and M. Yoon. 2005. "Distinct and shared cortical regions of the human brain activated by pictorial depictions versus verbal descriptions: an fMRI study". *Neuroimage* 24: 473–486.
- Gazzaniga, M. 2000. "Cerebral specialization and interhemispheric communication: does the corpus callosum enable the human condition?". *Brain* 123: 1293–1326.
- Goodman, N. 1968. *The languages of art*. Indianapolis: Boobs-Merrill.
- Hickok, G., and D. Poeppel. 2007. "The cortical organization of speech processing". *Nature review of neuroscience* 8: 393–402.
- Hickok, G., and S. Small. 2015. *Neurobiology of language*. New York: Academic Press.
- Hillis, A. E., & Caramazza, A. (1991). Mechanisms for accessing lexical representations for output: evidence from a category-specific semantic deficit. *Brain and language*, 40(1), 106–144.
- Hoffman, P., and A. Tamm. 2020. "Barking up the right tree: Univariate and multivariate fMRI analyses of homonym comprehension". *Neuroimage* 219: 117050.
- Holler, D., M. Behrmann, and J. Snow. 2019. "Real-world size coding of solid objects, but not 2-D or 3-D images, in visual agnosia patients with bilateral ventral lesions". *Cortex; A Journal Devoted To The Study Of The Nervous System And Behavior* 119: 555–568.
- Horgan, T., and J. Tienson. 2002. "The intentionality of phenomenology and the phenomenology of intentionality". In *Philosophy of mind*, ed. D. Chalmers, 520–533. Oxford: Oxford University Press.
- Humphreys, G. W., & Forde, E. M. (2001). Hierarchies, similarity, and interactivity in object recognition: "category-specific" neuropsychological deficits. *The Behavioral and brain sciences*, 24(3), 453–509.
- Hurley, R., M. Mesulam, J. Sridhar, E. Rogalski, and C. Thompson. 2018. "A nonverbal route to conceptual knowledge involving the right anterior temporal lobe". *Neuropsychologia* 117: 92–101.
- Husserl, E. 2006. *Phantasy, image consciousness, memory*. Dordrecht: Springer.
- Kemmerer, D. 2010. "How words capture visual experience: the perspective from cognitive neuroscience". In *Words and the mind: how words capture human experience*, eds. B. Malt, and P. Wolff, 287–327. Oxford: Oxford University Press.
- Knapen, T., J. Brascamp, J. Pearson, R. Ee, and R. Blake. 2011. "The role of frontal and parietal brain areas in bistable perception". *The journal of neuroscience*: 31: 10293–10301.

- Konen, C., M. Behrmann, M. Nishimura, and S. Kastner. 2011. "The functional neuroanatomy of object agnosia: a case study". *Neuron* 71: 49–60.
- Kormmeier, J., and M. Bach. 2012. "Ambiguous figures - what happens in the brain when perception changes but not the stimulus". *Frontiers in human neuroscience* 6: 51.
- Lambon Ralph, M., E. Jefferies, and K. Patterson, et al. 2017a. "The neural and computational bases of semantic cognition". *Nature reviews neuroscience* 18: 42–55.
- Lambon Ralph, M., E. Jefferies, K. Patterson, and T. Rogers. 2017b. "The neural and computational bases of semantic cognition". *Nature reviews of Neuroscience* 18: 42–55.
- Leonardelli, E., E. Fait, and S. Fairhall. 2019. "Temporal dynamics of access to amodal representations of category-level conceptual information". *Scientific reports* 9: 239.
- Levelt, W. 2001. "Spoken word production: a theory of lexical access". *Proceedings of the National Academy of Sciences of the United States of America* 98: 13464–13471.
- Liebenthal, E., J. Binder, S. Spitzer, E. Posing, and D. Medler. 2005. "Neural substrates of phonemic perception". *Cerebral cortex* 15: 1621–1631.
- Long, G. M., and T. C. Toppino. 2004. "Enduring interest in perceptual ambiguity: alternating views of reversible figures". *Psychological bulletin* 130: 748–768.
- Macpherson, F. 2012. "Cognitive penetration of color experience: rethinking the issue in light of an indirect mechanism". *Philosophy and phenomenological research* 84: 24–62.
- MacGregor, L., J. Rodd, R. Gilbert, O. Hauk, E. Sohoglu, and M. Davis. 2020. "The neural time course of semantic ambiguity resolution in speech comprehension". *Journal of cognitive neuroscience* 32: 403–425.
- Mahon, B. Z., and A. Caramazza. 2008. "A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content". *Journal of physiology* 102: 59–70.
- Mahon, B., and A. Caramazza. 2009. "Concepts and categories: a cognitive neuropsychological perspective". *Annual review of psychology* 60: 27–51.
- Manning, L. (2000) Loss of visual imagery and defective recognition of parts of wholes in optic aphasia, *Neurocase*, 6 (2), pp. 111–128.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Cambridge, MA: MIT Press.
- Martina, G., and A. Voltolini. 2017. "Perceiving groupings, experiencing meanings". *Rivista di estetica* 66: 22–46.
- Matchin, W., and G. Hickok. 2020. "The cortical organization of syntax". *Cerebral Cortex* 14: 1481–1498.
- McDowell, J. 1998. "In defense of modesty". In *Meaning, knowledge, and reality*, 87–107. Cambridge MA: Harvard University Press.
- Mion, M., K. Patterson, J. Acosta-Cabronero, G. Pengas, D. Izquierdo-Garcia, and Y. T. Hong, et al. 2010. "What the left and right anteriorfusiform gyri tell us about semantic memory". *Brain* 133: 3256–3268.
- Moore, C., and C. Price. 1999. "A functional neuroimaging study of the variables that generate category-specific object processing differences". *Brain* 122: 943–962.
- Nanay, B. (2011). Perceiving pictures. *Phenomenology and the Cognitive Sciences*, 10, 461–480.
- Nanay, B. 2016. *Aesthetics as philosophy of perception*. Oxford: Oxford University Press.
- Nanay, B. 2018. "Threefoldness". *Philosophical studies* 175: 163–182.
- Nanay, B. 2019. *Aesthetics. A very short introduction*. Oxford: Oxford University Press.
- Nanay, B. 2022. "What do we see in pictures? The sensory individuals of picture perception". *Philosophical Studies*. <https://doi.org/10.1007/s11098-022-01864-9>.
- Nes, A. 2016. "On what we experience when we hear people speak". *Phenomenology and mind* 10: 58–85.
- Norman, K., S. Polyn, G. Detre, and J. Haxby. 2006. "Beyond mind-reading: multi-voxel pattern analysis of fMRI data". *Trends in cognitive sciences* 10: 424–430.
- O' Callaghan, C. 2011. "Against hearing meanings". *The philosophical quarterly* 61: 783–807.
- Paivio, A. 1986. *Mental representations: a dual coding approach*. Oxford: Oxford University Press.
- Palejwala, A., K. O'Connor, and C. Milton, et al. 2020. "Anatomy and white matter connections of the fusiform gyrus". *Scientific reports* 10: 13489.
- Parkkonen, L., L. Andersson, J. Hämäläinen, M., & Hari, R. 2008. "Early visual brain areas reflect the percept of an ambiguous scene". *Pnas* 105: 20500–20504.
- Patterson, K., P. J. Nestor, and T. Rogers. 2007. "Where do you know what you know? The representation of semantic knowledge in the human brain". *Nature reviews of Neuroscience* 8: 976–987.
- Pitt, D. 2004. "The phenomenology of cognition, or, what is it like to think that P?". *Philosophy and phenomenological research* 69: 1–36.
- Pulvermüller, F. 1999. "Words in the brain's language". *Behavioral and brain sciences* 22: 253–336.
- Prinz, J. 2002. *Furnishing the mind*. Oxford: Oxford University Press.

- Ralph, M., and D. Howard. 2000. "Gogi aphasia or semantic dementia? Simulating and assessing poor verbal comprehension in a case of progressive fluent aphasia". *Cognitive neuropsychology* 17: 437–465.
- Reinholz, J., and S. Pollmann. 2005. "Differential activation of object-selective visual areas by passive viewing of pictures and words". *Cogn brain res* 24: 702–714.
- Rodd, J. M., M. H. Davis, and I. S. Johnsrude. 2005. "The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity". *Cerebral cortex* 15: 1261–1269.
- Rodd, J., I. Johnsrude, and M. Davis. 2010. "The role of domain-general frontal systems in language comprehension: evidence from dual-task interference and semantic ambiguity". *Brain and language* 115: 182–188.
- Schier, F. 1986. *Deeper into pictures*. Cambridge: Cambridge University Press.
- Shaul, S., and Z. Rom. 2019. "The differences in semantic processing of pictures and words between dyslexic and typical-reading university students". *Psychology* 10: 235–255.
- Siewert, C. 1998. *The significance of consciousness*. Princeton: Princeton University Press.
- Simmons, W. K., & Barsalou, L. W. (2003). The similarity-intopography principle: reconciling theories of conceptual deficits. *Cognitive neuropsychology*, 20(3), 451–486.
- Spetch, M. L., and R. L. Weisman. 2012. "Birds' perception of depth and objects in Pictures". In *How animals see the World: comparative behavior, Biology, and evolution of Vision*, eds. O. F. Lazareva, T. Shimizu, and E. A. Wasserman, 216–231. Oxford: Oxford University Press.
- Sterzer, P., A. Kleinschmidt, and G. Rees. 2009. "The neural bases of multistable perception". *Trends in cognitive sciences* 13: 310–318.
- Stokes, D. 2018. "Attention and the cognitive penetrability of perception". *Australasian journal of philosophy* 96: 303–318.
- Strawson, G. 1994. *Mental reality*. Cambridge MA: The MIT Press.
- Stumpf, C. 1890. *Tonpsychologie*. vol. II. Leipzig: Hirzel.
- Taikh, A., I. Hargreaves, M. Yap, and P. Pexman. 2015. "Semantic classification of pictures and words". *Quarterly journal of experimental psychology* 68: 1502–1518.
- Toppino, T. C., and G. M. Long. 1987. Selective adaptation with reversible figures: don't change that channel. *Perception & psychophysics* 42: 37–48.
- Vandenbergh, R., C. Price, R. Wise, O. Josephs, and R. S. Frackowiak. 1996. "Functional anatomy of a common semantic system for words and pictures". *Nature* 383: 254–256.
- Vishwanath, D. 2014. "Toward a new theory of stereopsis". *Psychological review* 121: 151–178.
- Vitello, S., and J. M. Rodd. 2015. "Resolving semantic ambiguities in sentences: cognitive processes and brain mechanisms". *Language and linguistics Compass* 9: 391–405.
- Voltolini, A. 2015. *A syncretistic theory of depiction*. Basingstoke: Palgrave.
- Voltolini, A. 2018. "Twofoldness and three-layeredness in pictorial representation". *Estetika* 55: 89–111.
- Voltolini, A. 2020a. "Different kinds of fusion experiences". *Review of philosophy and psychology* 11: 203–222.
- Voltolini, A. 2020b. "Qua seeing-in, pictorial experience is a superstrongly cognitively penetrated perception". *Studies on art and architecture* 29: 13–30.
- Wang, X., N. Sang, L. Hao, Y. Zhang, T. Bi, and J. Qiu. 2017. "Category selectivity of human visual cortex in perception of rubin face-vase illusion". *Frontiers in psychology* 8: 1543.
- Welchman, A. 2016. "The human brain in depth: how we see in 3d". *Annual review in visual sciences* 14: 345–376.
- Wittgenstein, L. 1991. *Remarks on the philosophy of psychology vol. 1*. Oxford: Blackwell.
- Wittgenstein, L. 2009⁴. *Philosophical investigations*. Oxford: Blackwell.
- Wollheim, R. 1980. *Art and its objects*. Cambridge: Cambridge University Press.
- Wollheim, R. 1987. *Painting as an art*. Princeton: Princeton University Press.
- Wollheim, R. 1998. "On pictorial representation". *The journal of aesthetics and art criticism* 56: 217–226.
- Wollheim, R. 2003a. "In defense of seeing-in". In *Looking into pictures*, eds. H. Hecht, R. Schwartz, and M. Atherton, 3–15. Cambridge MA: The MIT Press.
- Wollheim, R. 2003b. "What makes representational painting truly visual?". *Proceedings of the aristotelian society suppl. vol. 77*: 131–147.
- Zannino, G., F. Barban, E. Macaluso, C. Caltagirone, and G. Carlesimo. 2011. "The neural correlates of object familiarity and domain specificity in the human visual cortex: an fMRI study". *Journal of cognitive neuroscience* 23: 2878–2891.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Fabrizio Calzavarini¹ · Alberto Voltolini¹

✉ Alberto Voltolini
alberto.voltolini@unito.it

Fabrizio Calzavarini
fabrizio.calzavarini@unito.it

¹ University of Turin, Turin, Italy