

Cognitive-Semantic Concept Structuring in Visual Perception and Visual Imagery: Identifying a New Basic Top-Down Processing System for Visual Experience

Kurt Stocker

Department of Humanities, Social and Political Sciences, Chair of
Cognitive Science, ETH Swiss Federal Institute of Technology Zurich,
Zurich, Switzerland

Psychopharmacology Research, Division of Clinical Pharmacology &
Toxicology, Department of Clinical Research, University Hospital Basel
and University of Basel, Basel, Switzerland

Department of Psychology, University of Zurich, Switzerland

Faculty of Psychology, Swiss Distance University Institute, Brig, Switzerland

kurt.stocker@gess.ethz.ch

Abstract

Visual experience encompasses both visual perception and visual imagery. Using cognitive-semantic analysis, it is outlined how visual experience can be characterized as involving (among other top-down processes) a basic top-down-processing system that is organized by fundamental aspects of the four main Talmyan concept structuring systems (fundamental aspects of configurational structure, perspective, attention, and—not in principle, but quite often—force dynamics). It is proposed that this visual concept structuring represents a thus far unidentified basic top-down-processing system of visual experience which structures the conceptual content of the visual/imagistic stimulus. While Talmyan concept structuring has in the literature thus far mainly been applied to language, the main conclusion of this paper is that Talmyan concept structuring is a basic component of visual top-down processing even when language is not involved. Implications for sensory modalities other than the visual one are discussed, and possible future research is formulated.

Keywords

seeing – visual perception – visual imagery – Talmyan concept structuring – configurational structure – perspective point (PP) – attention – force dynamics

1 Introduction

Visual experience in this article encompasses both visual perception and visual imagery. *Visual perception* is the process through which we with our eyes take raw visual sensations from the environment and give them meaning using knowledge, experience, and understanding of the world (Oliva and Torralba, 2007; Bernstein, 2016: 88). The processing of raw visual sensations from the environment starts with light entering the eye, which is then transported via retina and thalamus to the visual cortex and from there to still other cortical regions—a process referred to as *bottom-up visual processing*. The meaningful interpretation (meaning assignment) of the raw visual sensations takes place by applying mental faculties such as knowledge, experience, and understanding to these sensations—a process referred to as *top-down visual processing* (e.g., Li, 2019; Goldstein and Van Hooff, 2018: 57–8). Thus, all visual perception is considered to be the result of both bottom-up processing and top-down processing. One example of visual top-down processing using knowledge, experience, and understanding of the world is the *light-from-above assumption*: the everyday-experience of light usually coming from above can lead us to interpret the very same visual object either as convex or as concave, depending on whether the object is perceived to be lit from above or from below (Kleffner and Ramachandran, 1992).

Visual imagery occurs when we imagine something—when we “see” with our “mental eyes” something in our mind that is not physically present (Kosslyn et al., 2006). Since in visual imagery the bottom-up input from the physical eye and retina is missing, visual imagery is considered to physiologically solely involve visual top-down processing in the brain—processing which often involves similar brain areas as in visual perception (see e.g., Goldstein and Van Hooff 2018: 297–305 for an overview on neurophysiological similarities and differences between visual perception and visual imagery). While visual perception and visual imagery obviously differ from one another—one involving seeing and the other imagining—they also share basic similarities (Goldstein and Van Hooff 2018: 293–305). For example: looking along an object from an initial point of focus to a point farther removed from this initial point, normally takes longer, the farther the point is away. While we all know this to be true for

objects in visual perception, this has also been shown to be true for imagined objects, for objects in visual imagery (Kosslyn, 1973; Goldstein and Van Hooff, 2018: 320–321). Findings like this show that visual imagery and visual perception share some common spatial and temporal characteristics—such as that looking along an object takes longer when the looked-along distance is longer. It is due to such similarities between these two mental systems that they can often be referred to collectively as “visual experience” in this article because this article often looks at spatial, temporal, and still other characteristics that are common to both visual perception and visual imagery. More specifically, this article proposes that both visual perception and visual imagery—besides being top-down processed by mental faculties such as experience, knowledge, and understanding—are always also structured in a top-down manner by specific cognitive-semantic mechanisms. Such a cognitive-semantic view of perception and imagery is in line with the view of Leonard Talmy, the founding father of the discipline cognitive semantics:

... the word “semantic” simply refers to the specifically linguistic form of the more generic notion “conceptual.” Thus, general conception—that is, thought—includes linguistic meaning within its greater compass ... research on cognitive semantics is research on conceptual content and its organization in language and, hence, on the nature of conceptual content in general. In this formulation, *conceptual content is understood to encompass not just ideational content but any experiential content, including affect and perception.* (Talmy, 2000: 4; my emphasis)

More specifically, what Talmy proposes here in relation to perception is that perception (like language) is organized by what he calls “concept structuring systems” (Talmy 2000: 40). The main concept structuring systems which Talmy implies to be at work in virtually all perception, imagery, and thought are the following: configurational structure, perspective, attention, and—if causal or causality-related thought is involved—force dynamics (Talmy, 2000; Stocker, 2015). Yet, while Talmy’s writings can be interpreted to *imply* that concept structuring systems also structure perception and imagery, there is no account yet thereof that shows that this may be so. Therefore, this article sets out to propose and demonstrate that the process through which we structure the content of a visual percept or a visual mental image is organized by basic aspects of the four main Talmyan concept structuring systems (by basic aspects of configurational structure, perspective, attention, and—not in principle, but quite often—force dynamics).

In order to exemplify how looking at a visual object (or seeing it in one's mind) can be proposed to be subjected to Talmyan concept structuring—henceforth this is often referred to as *concept-structure top-down processing* in this article—we use the famous painting of a chair by Van Gogh for exemplification (Figure 1). Throughout the article, the visual percept (or mental image thereof) of Figure 1 is referred to as CHAIR ON FLOOR. Let us now proceed to examine how seeing (by the example CHAIR ON FLOOR) can be characterized to involve concept-structure top-down processing—involving visual perception in combination with possibly added visual imagery. For this we characterize how perceiving CHAIR ON FLOOR (Figure 1) is subjected to basic aspects of the concept structuring systems configurational structure (Section 2), perspective (Section 3), attention (Section 4), and force dynamics (Section 5). Section 2



FIGURE 1 Vincent Van Gogh, *Van Gogh's Chair*. 1888. Currently held by the National Gallery, London. SOURCE: [HTTPS://COMMONS.WIKIMEDIA.ORG/WIKI/FILE:VINCENT_WILLEM_VAN_GOGH_138.JPG](https://commons.wikimedia.org/wiki/File:Vincent_Willem_Van_Gogh_138.jpg). THIS WORK IS IN THE PUBLIC DOMAIN, I.E., NO COPYRIGHT APPLIES TO THIS WORK OF ART.

is mainly concerned with visual perception, and from section 3 onwards visual imagery is incorporated into the overall descriptive framework. For the analysis in this article, we mainly draw on the decade-long work by Talmy (2000), who has worked out these concept-structuring systems using cognitive-semantic analysis, and on the work by Stocker (2015), who additionally showed how these four basic concept-structuring systems can feature simultaneously within a single mental representation—in our case, this will be shown for the mental representation of the percept and mental image CHAIR ON FLOOR.

2 Spatiotemporal Configurational Structure

Talmy uses the terms *concept structuring systems* and *schematic systems* interchangeably. *Configurational structure* is one of the four fundamental concept structuring systems (Talmy, 2000).

The first schematic system we treat is that of **configurational structure**. This system comprises the schematic structuring or *geometric delineations in space or time* or other qualitative domain that closed-class forms can specify. *Closed-class forms can ascribe such structure to the whole of a referent scene, thus partitioning that scene into entities in particular relationships*, or to any of those entities themselves, or to the paths described by such entities when their interrelationships change through time. (Talmy, 2000: 47; his emphasis [bold], my emphasis [italics])

In linguistics, one can make a distinction between open-class and closed-class forms. A class of morphemes is considered open if it is quite large and open to new entries (such as roots of nouns, verbs and adjectives), whereas a class of morphemes is considered closed if it is relatively small and fixed in membership (such as determiners, prepositions, conjunctions, particles, inflections, derivations) (Talmy, 2000: 22–24). The main closed-class form at work in CHAIR ON FLOOR is the spatial-prepositional relation ON, as we will see as the discussion in this section proceeds.

Let us now examine how our specimen percept CHAIR ON FLOOR is top-down-processed in configurational-structural terms. Based on extensive cognitive-semantic analysis and experimental review, Stocker (2014b) proposed that what is commonly referred to as “space” in relation to the mind always also contains “time,” and that what is commonly referred to as “time” in relation to the mind always also contains “space”—and thus he introduced the

terms object spacetime and event spacetime for mental “space” and mental “time,” respectively. Taking over this framework, we can characterize CHAIR ON FLOOR as a mental representation in object spacetime. Object spacetime consists of three-dimensional cognitive space xyz with time t unfolding along the fourth dimension as an abstract nonspatial correlate, resulting overall in an $xyzt$ spacetime conglomerate (Stocker, 2014b: 81, Fig. 1a). In *spatial* terms, visually perceiving CHAIR ON FLOOR involves a mental representation where the image can mentally unfold along the three mental xyz -axes, giving the image not only height and width, but also mentally constructed depth, so that we for instance perceive the *door* in CHAIR ON FLOOR to be behind the *chair* (cf. Figure 1). Psychological demonstration that we mentally construct depth comes for instance from the classical Ponzo size illusion (see e.g., Bernstein, 2016: 123–124). When it comes to the *temporal* part of CHAIR ON FLOOR, it is natural to assume that in order to visually perceive CHAIR ON FLOOR, it must also mentally extend in time (t), as any perception must have a certain duration in order for it to be perceived. For a principled cognitive-semantic account why even a static visual scene (such as CHAIR ON FLOOR) always contains time see Stocker (2014b: 72–76). In Section 5, we will see that the potential force-dynamic part of the percept CHAIR ON FLOOR breaks up the temporal extension of this percept into two distinct time points (t_1, t_2), given further evidence of the underlying temporal dimension in this percept. For now, let it suffice to say that in terms of the schematic system configurational structure, visually perceiving CHAIR ON FLOOR involves an underlying geometric $xyzt$ scaffolding.

As can be seen in the last quote from Talmy above, the concept-structuring system *configurational structure* not only encompasses dimensional extensions in mental spacetime, but also the partitioning of such a spatiotemporal scene into particular relationships, relationships that, when verbalized, are most likely expressed with closed-class forms. Thus, when we look at CHAIR ON FLOOR, we most likely partition the overall percept into spatial-prepositional (closed-class) relationships like ON and NEXT TO, such as CHAIR ON FLOOR, PIPE ON CHAIR, TOBACCO POUCH NEXT TO PIPE, and so on. For the sake of this paper, it suffices to use one of these possible mentally perceived spatial relationships for exemplification: CHAIR ON FLOOR.

While this paper is a theoretical account of how concept structuring might be a thus far unidentified basic top-down processing system for visual experience – i.e., an account which lacks empirical evidence that has directly investigated this theoretical account – it can nevertheless be noticed that there is also evidence from cognitive-neuroscience studies and computational simulations that we perceive a scene like CHAIR ON FLOOR in spatial-relational terms not

only in language, but also in visual perception and spatial cognition (a summary of such studies can be found in Laeng et al., 2011: 298).

Having established top-down-processed configurational structure of the percept in Figure 1—simplifying it to CHAIR ON FLOOR and ignoring further configurational structure such as PIPE ON CHAIR—we may now note that a very crucial part of the percept is so far still missing: the self. Because arguably, if there were no self functioning as a perceiver, there would be no percept. In Talmyan concept structuring, the self/perceiver is captured with the concept-structuring system *perspective* (Talmy 2000; Stocker, 2012), the system to which we turn to next.

3 Object-Exterior, Object-Interior, and Object-Co-Located Perspective Point

Perspective is a second of the altogether four fundamental Talmyan concept structuring systems and can be described as:

The present schematic system consists of the **perspective** that one can have on ... [a referent] entity ... This system thus *establishes a conceptual perspective point [PP] from which the entity is cognitively regarded*. While this schematic system is presumably neutral to particular sensory modalities, it is most readily characterized in visual terms as, in effect, pertaining to where one places one's "mental eyes" to "look out" upon a referent structure. (Talmy, 2000: 68; his emphasis [bold], my emphasis [italics])

The most basic distinction in perspective in relation to a referent entity is if the self/perceiver takes on a perspective point that is outside the reference object or inside or at least co-located with the reference object. (The term "object" is used here in the generic sense of Stocker (2014b), in which an "object" can refer to an animate or inanimate entity.) Talmy exemplifies it thus:

Consider, for example, the following bit of narrative: "She sat in the rocker near her bed and looked out the window. How lovely the sky was!" In the first sentence, the use of a third-person pronoun together with the objective scene description invites the listener to place his perspective point [PP] somewhere in the depicted room looking at the sitting woman. But in the second sentence, the exclamatory *how*-construction, together with the expression of subjective experience, induces the listener to relocate

his perspective point [PP] to the location of the sitting woman, in effect, looking out through her eyes. (Talmy, 2000: 68–69; his emphasis)

The first PP (“somewhere in the depicted room”) represents what we will refer to as an *object-exterior perspective point*: the self/perceiver construes a mental viewpoint outside the looked-upon reference object (outside the woman). In contrast, the second PP (“looking through her eyes”) represents what we will refer to as an *object-interior perspective point*: the self/perceiver construes a mental vantage point from inside the reference object (inside the woman), or at least a vantage PP that is co-located with the referent object. While in the example by Talmy above, it is the third-person pronoun and the objective scene description that invites the listener to take on an object-exterior PP, the visual main referent-entity in Figure 1 (the chair) also invites the viewer to take on an external PP. Instead of a linguistic device evoking a third-person perspective (using “she”), we have a painted device that evokes a third-person perspective (painting an “it” such as chair), and instead of a linguistic objective scene description, we here have a painted objective scene description (but as is typical for Van Gogh the “objective” world is overlaid with much subjectivity).

While the *woman-looking-at-sky* example of Talmy above demonstrates possible relocation of an object-exterior PP to an object-interior PP in relation to an *animate* referent entity, it is worthy of note that this—location of one’s mental eyes to a “viewpoint” within or co-located with an entity other than oneself—can also be proposed to be linguistically evoked in relation to an *inanimate* referent entity:

... the prospect path ascribed to an inanimate configuration, such as a cliff wall or a window, is often associated with an actual viewer located at that configuration and directing her or his visual path along the same path as the prospect line. Thus, in (i), one readily imagines a viewer standing at the cliff edge or in the bedroom looking out along the same path as is associated with the cliff wall or the window.

- (i) a. The cliff wall faces/looks out toward the butte.
- b. The bedroom window faces/looks out/opens out toward the butte/onto the patio. (Talmy, 2000: 173)

Thus, according to Talmy, the self/perceiver would take on object-co-located PP in (i), co-locating the PP with the cliff or the bedroom window. The further theoretical possibility of taking on an object-interior PP in such linguistically evoked scenes, as Talmy proposed in relation to the *woman-looking-out-the-window*

example above, is left unexplored by Talmy in relation to (i). Nevertheless, we can basically distinguish object-exterior PPs, object-co-located PPs, and object-interior PPs as three basic types of PPs. If looking at CHAIR ON FLOOR (Figure 1) were accompanied by a sentence like “The backrest of the chair faces Van Gogh, who is walking toward the chair,” one might also be invited to relocate one’s mental eyes to looking out along the same path that is associated with the backrest toward an imaginary Van Gogh approaching the chair (which could in principle represent either an object-co-located or object-interior PP). However, without such additional linguistic evoking, the most likely mental viewpoint when looking at Figure 1 is of course an object-external PP, as a chair is more naturally cognized as being viewed from a perspective that is external from it.

As mentioned above, this paper is a theoretical account of how concept structuring might be a thus far unidentified basic top-down processing system for visual experience that lacks direct empirical evidence for such an account. Keeping the lack of direct empirical evidence in mind, it is still worthwhile to point out that there is also cognitive-psychological evidence that we—triggered by linguistic or other contextual stimuli—adjust the location of our mental PP in a top-down manner. Franklin and colleagues (1992), for instance, showed that while reading a story one can—depending on the linguistic devices employed—take on the perspective of the described character or a disembodied perspective between the two described characters. Similarly, cognitive-psychological studies have also shown that while communicating (Schober, 1993) or while (similar to looking at a painting) looking at a photograph (Tversky and Martin Hard, 2009), we might either take our own perspective point or co-locate our perspective point with the person in sight.

Finally, although Talmy has only defined the PP in relation to “mental eyes” (Talmy, 2000: 68; cf. above), it can be noted that it is clear that in his theory of cognitive semantics he uses a generic notion of a perspective point which encompasses *both* how *mental eyes* (visual imagery) and how *physical eyes* (visual perception) look out upon a scene. Accordingly, Talmy for instance characterizes the PP with descriptions like “an observer (or conceiver) has in sight (or mind)” (Talmy, 2000: 313). In this article I take over this generic notion of the PP which encompasses both—the looking out upon a scene with mental or physical eyes—and refer to this notion as the Talmyan PP (or only as the PP).

In relation to top-down conceptual structuring of the percept CHAIR ON FLOOR (Figure 1), we have thus far established its top-down four-dimensional unfolding in mental object spacetime, including the top-down assignment of spatial relationships of the objects to one another (in Section 2), as well as the

typical object-external perspective point of the perceiver in relation to such a percept (in this section).

It is worthy of note that an object-external PP in visual perception (an object-external PP involving physical eyes) is likely to also involve *bottom-up*-processing parts. After all, when for example looking at the object CHAIR ON FLOOR (Figure 1), the perceiver/self is located externally to this object when light is reflected from the surface of this object and falls into her/his eye to initiate the visual-perception process of this object. Clearly though, even an object-exterior PP in visual perception will always also consist of top-down-processing parts, such as constructing a certain distance/depth between the PP and the externally located object. In contrast, an object-exterior PP in visual imagery (an object-external PP involving mental eyes) can be proposed to always involve top-down processing only, as visual imagination can be proposed to be principally of a top-down manner (cf. Introduction).

In the next section, we now turn to the question why in the percept CHAIR ON FLOOR for most people the chair might be more in the foreground than the floor.

4 Figure-Ground Attentional Assignment

The third of the four fundamental Talmyan concept structuring systems we look at is *attention*. Attention—as conceptualized in Talmyan concept structuring—is a complex construct that involves windowing/gapping mechanisms of attention (Talmy 2000; for empirical evidence see Laeng et al., 2011), as well as the foundational attentional notions of figure and ground:

[There is a] pervasive system by which language establishes one concept as a reference point or anchor for another concept. It posits the existence in language of two fundamental cognitive functions, that of the **Figure**, performed by the concept that needs anchoring, and that of the **Ground**, performed by the concept that does the anchoring. The pair of concepts can be of two objects relating to each other in space...[o]r the pair of concepts can be of two events relating to each other in a temporal, causal, or other type of situation ... [such] [*c*]ognitive anchoring mainly involves ... *attention* and its differential distribution. (Talmy, 2000: 311; his emphasis [bold], my emphasis ([italics]))

This linguistically-based bipartite partitioning of a mental scene into figure and ground corresponds to some degree to figure and ground partitioning in

Gestalt perceptual psychology (Talmy, 2000: 313; Herskovits, 1986: 197). Going beyond the Gestalt figure/ground dichotomy, Talmy has also worked out a tripartite partitioning of a mental scene into figure, ground, and reference frame (Talmy, 2000: 311–313)—but this tripartite partitioning in the case of the percept CHAIR ON FLOOR only becomes relevant when potential imagined motion in relation to this percept additionally comes into play. This possibility—and along with it attentional imagistic tripartite partitioning—will be examined in the next section, while in this section it suffices to consider the bipartite partitioning of the percept CHAIR ON FLOOR solely into figure and ground. If we were to verbalize CHAIR ON FLOOR, we might express this with a sentence like “The chair is on the floor,” but not with a sentence like “*The floor is on the chair.”¹ It is a general linguistic rule that in a simple sentence the subject nominal (“chair” in our case) functions as figure and the object nominal (“floor” in our case) functions as ground (Talmy, 2000, p. 347). In the present case, this linguistic bipartite partitioning of the expression “The chair is on the floor” into the figure “chair” and the ground “floor” is identical to the visual-perceptual bipartite partitioning (Gestalt psychology) of CHAIR ON FLOOR into the perceptual figure CHAIR and the perceptual ground FLOOR (cf. Rubin 1915/1921; Bernstein, 2016: 117–118).

To sum up, the perceptual entities within the overall percept CHAIR ON FLOOR are not all given the same amount of attention; rather a top-down process involving figure-ground discrimination—“[t]he ability to organize a visual scene so that it contains meaningful figures set against a less relevant ground” (Bernstein, 2016: 118)—imposes onto this percept a foregrounding of one object (chair) at the expense of backgrounding other objects (floor and still other typically backgrounded objects such as the box).

1 Adopting a widely used convention in linguistics, I mark semantically or grammatically unacceptable or highly unusual sentences with an asterisk. Note that “*The floor is on the chair” is not a perceptual impossibility. If you compare it to a percept like LAMP ON CEILING, which implies that the lamp is below the ceiling but attached to it, one could—with a stretch of imagination—also mentally picture that the chair in CHAIR ON FLOOR is actually hovering in space and that the floor is below and attached to the bottom parts of the chair’s legs. If one for instance had a play chair and glued a play floor (say a piece of cardboard) onto the bottom parts of the chair, then one could quite felicitously comment on this with a statement like “Look, now the floor is on the chair, and it sticks quite well.” But this is of course not how most of us would typically perceive the relational pattern of the chair and the floor when looking at CHAIR ON FLOOR.

5 Force-Dynamic Affordance

As we have seen, the first top-down concept-structuring system described in the Talmyan framework in this section was *configurational structure* (unfolding of spatiotemporal dimensions and unfolding of spatiotemporal relationships between or among entities, Section 2); the second *perspective* (the positioning of the perceiver's perspective point, Section 3); and the third *attention* (foregrounding and backgrounding of certain entities, Section 4). In relation to this thus far tripartite (configurational/perspectival/attentional) concept-structuring complex, "*force dynamics* is a fourth schematic system: to the preceding basically pictorial complex [the tripartite configurational/perspectival/attentional complex], one now adds the forces that the elements of the structural framework exert on each other" (Talmy, 2000: 467; my emphasis).

Both classic force dynamics (Talmy, 2000) and a more modern version of forced dynamics called elementary force dynamics (Stocker, 2014a) involve the assumption that mental representation of causal and causality-related relations involves the conceptualization of two opposed entities that are engaged in a force interaction. Force-dynamics can formally describe a number of different thinking/reasoning types such as causation, modality, epistemic reasoning, counterfactual/probabilistic reasoning, and stock-flow reasoning (Talmy, 2000; Stocker 2014a; Stocker and Funke, 2019; Wolff, 2003, 2007).² We now examine the possibility that the visual top-down processing of CHAIR ON FLOOR can also involve force-dynamic aspects, but for this we need to introduce the concept of *affordance* first.

A basic finding in psychology is that viewing visual objects in many cases involve what has been termed *affordance* (Gibson, 1966, 1979). Affordance refers to the possibility that perceiving a visual object can involve the cognition of one or more of its action options. According to this, visually perceiving a chair for instance, is likely to not only involve the perception of the chair per se, but can also involve the additional incorporation of an action option such as *sit* (Norman, 1988). For modern experimental evidence that perceiving a visual object often also involves the (not necessarily conscious) conceptualizing of a typically associated action option with it—for instance, seeing a cup

2 For a principled illustration that distinguishes "force-dynamically neutral expressions with ones that do exhibit force-dynamic patterns" (Talmy, 2000, p. 412) see Talmy (2000, pp. 412–413). There it is for example theoretically laid out why a mental representation underlying the cognition of the sentence "The ball was rolling along the green" can be considered force-dynamic neutral, whereas the mental representation underlying the cognition of the sentence "The ball kept (on) rolling along the green" can be considered to involve force-dynamic parts.

with a handle evokes the action option of grabbing the cup—see for instance Fischer and Dahl (2007).

Although this has to my knowledge never been undertaken thus far in the literature, affordance being a possible part of visual-object perception is readily described in terms of Talmy's force dynamics (Talmy, 2000), especially in the revised force-dynamic version by Stocker (2014a). In order to be able to demonstrate how affordance (using the example *sit*) might occur while looking at the percept CHAIR ON FLOOR, let it be assumed that an affordance thought like the following occurs while looking at this percept:

- (1) The chair is unoccupied; therefore, someone could sit on it.

In a force-dynamic framework, the *chair* functions as the *agonist* (*ago*). In the revised version of force dynamics (Stocker 2014a), *ago* is a cognitive entity which can take on any given state or action value. In (1), *ago* is the *chair*, which initially takes on the value *being unoccupied* ("unoccupied" here only means no-one is sitting on it, as there are still objects lying on the chair). The entity opposed to the *ago* entity is referred to as the *antagonist* (*ant*) in force-dynamic theory. In (1) the imagined *someone* who could sit on the chair takes on this *ant* function. In Stocker's force dynamics, *ant* is always conceptualized as potentially imposing a state or action onto *ago* that is *different* from the out-set action or state of *ago*. Thus, when *ago* is initially associated with the state *being unoccupied* (as in (1)), *ant* will by definition potentially impose a value or state other than *being unoccupied*. Hence in (1) the visual perceiver of the chair (*ant*)—if she or he conceptualizes sitting affordance in relation to the chair—imposes the value of *being occupied* onto the chair, as she/he (unconsciously or consciously) imagines someone (or even her- or himself) sitting on it. In elementary force dynamics (Stocker 2014a), this could be formally expressed as:

- (2) C: $\text{ago}_{\text{chair}-x_{\text{unoccupied}}}$, $\text{ant}_{\text{visual-perceiver}-x_{\text{diffoccupied}}(++)} \rightarrow \text{E: } \text{ago}_{\text{chair}-x_{\text{diffoccupied}}}$

(2) reads: the cause C involves conceptualizing the chair in the function of *ago* with the initial state value (*x*) of *being unoccupied*. *ant* (the perceiver) is conceptualized as imposing the different state value (x_{diff}) of *being occupied* onto *ago* (the chair). As *ant*'s imposing action is imagined as being carried out ("++" associated to the *ant* entry signifies that the *ant* tendency can be acted out), the cognized effect E is indeed that the perceiver (*ant*) causes the chair (*ago*) to be imagined as occupied. If in contrast to (1), the visual perceiver of CHAIR ON FLOOR for instance were to conceptualize a thought like

- (3) Although the chair is unoccupied, no-one should sit on it because the pipe and the tobacco pouch on it might signify that it is occupied.

then this would result in a force-dynamic conceptualization where ant is not able to act out its tendency of occupying the chair, and this would accordingly result in a different force-dynamic representation (see Talmy, 2000 and Stocker, 2014a for details).

Note that (2) also brings about two distinct time points, as there is a force-dynamic cause C (visually perceiving the chair as being unoccupied) followed by the force-dynamic effect E (fusion of visual perception and visual imagery resulting in imagining that someone is now occupying the chair). In principle, any cognized cause-and-effect relationship can be argued to have an underlying temporal relationship with the cause cognized as occurring first and the effect cognized as occurring second; this is even the case when we see a cause and effect being at work simultaneously, with simultaneity being overlaid over an underlying mental cause-first/effect-second pattern—see Kant (1781/1998: 312) and Stocker (2014a: 127–128) for detailed argumentation. Thus, (2) implies a time point 1 (t_1) at force-dynamic C, and time point 2 (t_2) at force-dynamic E. Such a possible fusion of visual perception (seeing the chair at t_1) and visual imagery (additionally imagining a sitting action in relation to the chair at t_2) means that in the *xyz*t spacetime extension underlying the perception of CHAIR ON FLOOR (cf. Section 2), *t* is partitioned to two distinct time points (t_1 , t_2). Furthermore, imagining sitting on the chair could—as one possibility—also evoke some imagined motion (for instance, imagining someone going over to the chair and/or sitting down on it). As Talmy has noticed, as soon as motion is involved in a mental scene that involves figure and ground (cf. Section 4), then also a *reference frame* in addition to figure and ground must be assumed. He exemplifies this with the sentence “The pen rolled off the table” in relation to which he writes:

For ... if an observer (or conceiver) has in sight (or mind) only the Figure object [the pen], she can know only that the object exists, but nothing of change of position. Even when ... the observer sees both Figure and Ground objects [pen and table]—still without any reference frame, however—she can additionally know only that there is a change from the two object's being together to their being apart, but could not know which object moved (or if both moved), nor whether there is any further motion once the two objects are apart, as there is no way to determine (change of) distance. Only when the observer sees both objects within a framework ... can she know which object is stationary, which object moves, by how much, and along what path. (Talmy, 2000: 313)

Similarly, in our case of a possible visual perceiver imagining someone or herself/himself with an approaching or sitting-down motion in relation to Van

Gogh's chair (Figure 1), she or he could then only know which object is stationary (the chair), which object moves (the imagined animated entity performing the approaching or sitting), and along what path (toward the chair and/or toward the seat surface), if this were imagined in an overall reference frame that encompasses figure and ground. Note that this possible adding of an approaching or sitting-down motion in relation to the chair would also involve a new figure-ground arrangement. Let this be illustrated with the imagined sitting-down motion. In the mental image SOMEONE SITTING DOWN ON CHAIR, the figure is now this someone (subject nominal if verbalized), the ground is the chair (object nominal if verbalized), and the floor is now part of the reference frame, part of the overall background in relation to which the new figure-ground arrangement is mentally construed in a top-down manner.

Let it also be noted that imagined *sitting* in relation to CHAIR ON FLOOR is only one possible top-down affordance action option in relation to this percept/mental image. There could be other affordance action options which might spontaneously occur in the perceiver in relation to the CHAIR, as for instance removing the PIPE and TOBACCO POUCH from it. In principle, any imagined top-down affordance action option in relation to a referent entity within a percept or mental image could be accounted for in force-dynamic terms.

6 Discussion

With the example of CHAIR ON FLOOR, it was demonstrated in this article how visual experience (exemplified with a combined cognition of visual perception and visual imagery) can be proposed to be conceptually structured in a quadripartite manner—namely top-down structured in relation to its spatiotemporal unfolding and configuration (Section 2), in relation to perspective (Section 3), in relation to attentional foregrounding/backgrounding (Section 4), and in relation to force-dynamic affordance (Section 5). Figure 2 sums up the proposed top-down quadripartite concept-structuring of visual experience in a single image. Thus, in order to be able to conceptually structure a visual scene, it must be mentally constructed in *xyzt* spacetime; it must be viewed by the self from a certain (in our example: object-exterior) perspective; it is most likely viewed in a specific figure-ground arrangement (for instance, the chair being mentally construed to be on the floor rather than vice versa)—and if affordance takes place, this then further results: in a force-dynamic structure with the perceiver imagining an ant (such as *someone*) imposing an action value (such as *sit*) onto ago (such as onto the *chair*); in *t* being partitioned into two distinct time points (beginning and end point of the mentally

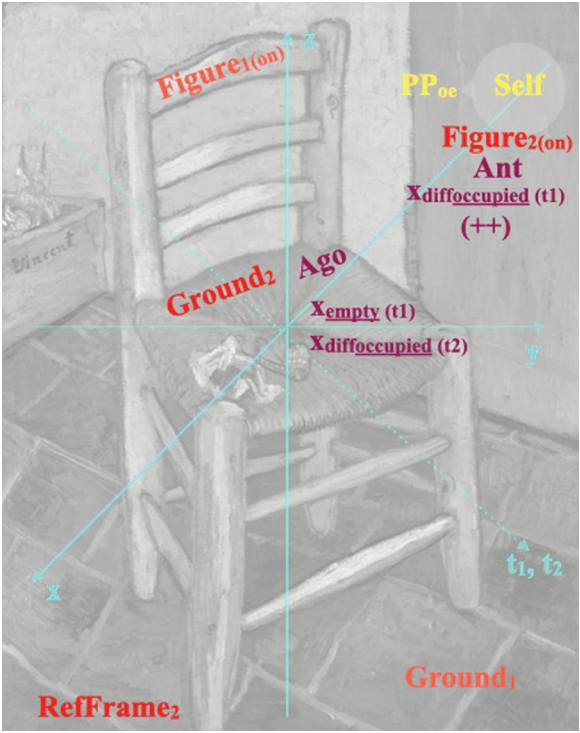


FIGURE 2 Visual experience of Van Gogh's Chair and top-down conceptual structuring thereof in terms of spatiotemporal configuration (green), perspective (yellow), attention (bright red, red), and force dynamics (purple). Green = object spacetime; xyz = four dimensions of object spacetime; t_i = time point 1, t_2 = time point 2. Yellow = perspective of self; PP_{oe} = object-exterior perspective point (round circle symbolizes self/perceiver; pointed "nose" symbolizes rough viewing direction of perceiver). Bright red: $Figure_i/Ground_i$ = Figure/Ground at t_i ; (on) = spatial-prepositional relation of figure to ground. Red: $Figure_2/Ground_2/Reference\ Frame_2$ = Figure/Ground/Reference Frame at t_2 ; (on) = spatial-prepositional relation of figure to ground within reference frame. Purple = force dynamics; Ago = agonist, Ant = antagonist, x = state or action value of agonist; x_{diff} = state or action value of antagonist (by definition different from state or action value of agonist); (++) associated with ant = ant tendency is acted out.

construed affordance cause-and-effect action in relation to the perceptual referent entity); and in a new figure/ground arrangement (e.g., imagined someone becoming the new figure, the chair changing from figure to ground, and the floor changing from ground to being part of the reference frame).

Figure 2 sums up (with the example of the percept CHAIR ON FLOOR) how visual experience (visual perception and/or visual imagery) can be proposed to be overlaid with complex top-down Talmyan mental concept-structuring. While this structuring (like all top-down processing) cannot be seen by the eye, it nevertheless seems to function as an invisible scaffolding that helps us to conceptually structure and organize what we see—either with our physical eyes and/or our mental eyes.

In this article, a theoretical framework for concept-structure top-down processing of visual experience (i.e., for visual perception and for visual imagery) has been developed. Using cognitive-semantic analysis, it has been proposed that all visual experience (all visual perception and all visual imagery) involves—besides top-down processing by mental faculties such as knowledge, experience, and understanding—top-down processing that consists of Talmyan concept structuring. Thus the overall picture that emerges for *visual perception* is to view it as a combination of bottom-up processing (visual information being processed by the eye/retina/thalamus/cortex) and of *two* basic types of top-down processing: i) top-down processing (guided by mental faculties such as knowledge, experience, and understanding of the world) that identifies the conceptual *content* of the visual stimulus (henceforth called *concept-content top-down processing*), and ii) top-down processing which *structures* this conceptual content in terms of basic aspects of Talmyan spatiotemporal-configurational, perspectival, attentional, and (quite often, but not in principle) force-dynamic aspects (*concept-structure top-down processing*). In contrast, *visual imagery*—given its lack of bottom-up (eye to brain) processing—can be viewed to solely consist of the two basic types of top-down processing: of concept-content top-down processing and of concept-structure top-down processing. The new overall proposal for the basic nature of visual experience is schematically outlined in Figure 3.

As one question for future work, we may ask what the implications of this new extended theoretical account of visual experience (as summarized in Figure 3) are for sensory modalities other than the visual one, such as the auditory or the tactile modality. To consider this question briefly, let the researcher who has been perhaps the main psychological researcher over the past decades in studying visual imagery, Stephen Kosslyn, be quoted here:

First, let us be clear on what we mean by the term *mental image*: in our usage, a mental image occurs when a representation of the type created

during the initial phases of perception is present but the stimulus is not actually being perceived; such representations preserve the perceptible properties of the stimulus and ultimately give rise to the subjective experience of perception. As this characterizing makes clear, we are not limiting mental imagery to the visual modality. Although visual imagery is accompanied by the experience of “seeing with the mind’s eye,” auditory mental imagery is accompanied by the experience of “hearing with the mind’s ear,” and tactile imagery is accompanied by the experience of “feeling with the mind’s skin” and so forth. (Kosslyn et al., 2006: 3–4; their emphasis)

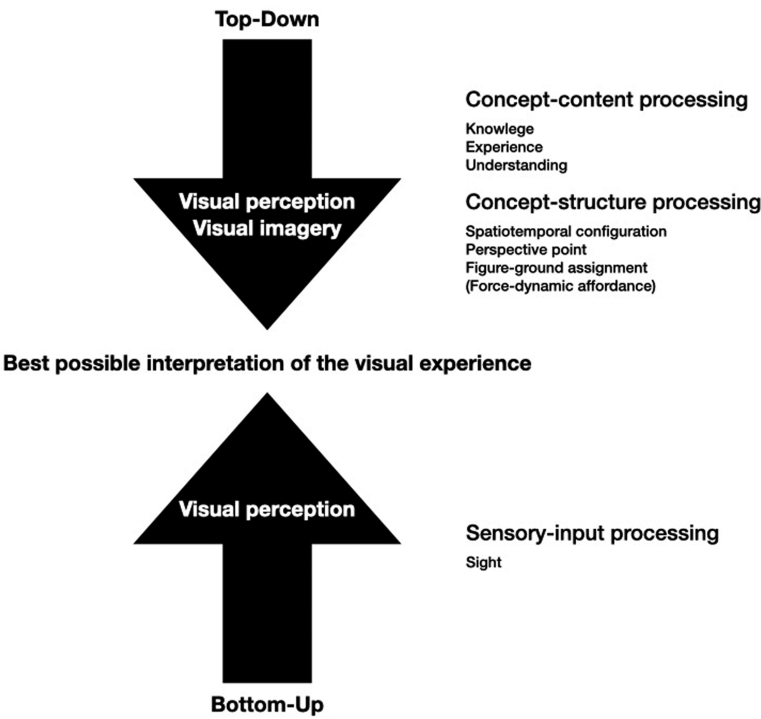


FIGURE 3 Visual experience. Visual perception is proposed to always involve a combination of bottom-up processing (eye to retina to thalamus to cortex) with two basic types of top-down processing: with concept-content top-down processing and with concept-structure top-down processing (both only processed in the brain). Visual imagery is proposed to solely consist of a combination of concept-content top-down processing and concept-structure top-down processing (again both only processed in the brain). (The basic idea for this Figure is derived from a Figure by Goldstein and Van Hooff (2018: 58), and has then been expanded upon.)

Taking over this generalization of imagery across sensory modalities from Kosslyn and his colleagues, one can—in analogy to this article’s visual experience standing for visual perception *and* for visual imagery—speak of auditory experience (standing for auditory perception and auditory imagery), tactile experience (standing for tactile perception and tactile imagery), olfactory experience (standing for olfactory perception and olfactory imagery), and so forth. The intriguing question arising for future research is, *whether sensory modalities other than visual experience—such as auditory experience or tactile experience—can also be described as (besides concept-content top-down processing) being top-down processed by the four main Talmyan cognitive concept structuring systems (or close analogues thereof)—namely by basic aspects of configurational structure, perspective, attention, and possibly added force dynamics*. It is for instance well known that attention plays a major role not only in the visual modality, but also in other modalities – such as the auditory modality (as for example shown by the famous cocktail party effect and the related dichotic listening task) or the olfactory modality (e.g., Zelano et al., 2005; Li, 2016). However, how well auditory or olfactory attention can be characterized with Talmyan attentional concept structuring—and *how well all perception and imagery can be characterized with Talmyan concept structuring*—is a question that is unstudied yet, and is a question that would deserve future investigation.

Finally, the concept-structuring theoretical framework of visual experience as developed in this paper is the only one I know of that can account for spatiotemporal structure, perspective, attentional foregrounding/backgrounding, and possible affordance within a single coherent theory. Nevertheless, it is of course still more than needed for future investigation to see if, and if so how, the present account could be fruitfully combined/integrated with other accounts of vision, such as with aspects of the vision theories of Marr (1982/2010) or Li (2019). One major emphasis of Marr’s theory is, for instance, how we derive at a “3-D model representation” (Marr, 1982/2010: 418) in visual perception, and one major emphasis of Li’s theory is how the “attentional bottleneck” (Li, 2019, p. 1) is at work in visual perception. The three-dimensional representations of visual percepts have in the current account been developed to a four-dimensional spatiotemporal account (spatiotemporal-configurational concept structuring, Section 2), and the attentional aspects in visual perception have been developed into tripartite partitioning of the visual scene into figure, ground, and reference frame (figure-ground-attentional concept structuring, Section 4). However, only very careful and extensive future theoretical and empirical work on visual experience will be able to determine whether aspects of the theories of Marr, Li, the current theory, and still other theories could be

integrated fruitfully with one another to arrive at a larger overall understanding of human visual experience.

In relation to “unraveling the mysteries of human visual perception” (Marr, 1982/2010: 419), David Marr once wrote:

Indeed, part of the fun is that we never really know where the next key is going to come from—a piece of daily experience, the report of a neurological deficit, a theorem about three-dimensional geometry, a psychophysical finding in hyperacuity[,] a neurophysiological observation, or the careful analysis of a representational problem. (Marr, 1982/2010: 419)

It is hoped that the present article might persuade some readers that perhaps one of the pieces for a next step forward in unraveling the mysteries of human visual experience is integrating (Talmyan) top-down concept structuring into the scientific description. On the one hand, tackling this question in the future would give cognitive semanticists a cognitive realm to study (sensory experience) in addition to language, and on the other hand, it would give perception/imagery researchers a theoretical framework—Talmyan concept structuring—whose conceptual rigor, precision, and detail in terms of mental phenomena might assist the field considerably. Relating to the latter, specific cognitive-psychological and neurophysiological future studies could be designed to see whether Talmyan concept-structure top-down processing of sensory experience could also be established on experimental grounds.

Acknowledgements

I thank the reviewers who have been involved in the reviewing process of this manuscript because their comments have added valuable new aspects to it.

References

- Bernstein, Douglas A. 2016. *Psychology: Foundations and Frontiers* [10th ed.]. Boston: Cengage Learning.
- Fischer, Martin H. and Christoph D. Dahl. 2007. The time course of visuo-motor affordances. *Experimental Brain Research*, 76 (3): 519–524.
- Franklin, Nancy, Barbara Tversky and Vicky Coon. 1992. Switching points of view in spatial mental models. *Memory & Cognition*, 20: 507–518.

- Gibson, James J. 1966. *The Senses Considered as Perceptual Systems*. London: Allen and Unwin.
- Gibson, James J. 1979. *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin Harcourt.
- Goldstein, Bruce E. and Johanna C. Van Hooff. 2018. *Cognitive Psychology*. Andover, Hampshire: Cengage Learning.
- Herskovits, Annette. 1986. *Language and Spatial Cognition*. New York: Cambridge University Press.
- Kant, Immanuel. 1781/1998. *The Critique of Pure Reason*. Translated by Paul Guyer and Allen. W. Wood. New York: Cambridge University Press.
- Kleffner, Dorothy A. and Vilayanur S. Ramachandran. 1992. On the perception of shape from shading. *Perception & Psychophysics*, 52(1): 18–36.
- Kosslyn, Stephen M. 1973. Scanning visual images: Some structural implications. *Perception & Psychophysics*, 14(1): 90–94.
- Kosslyn, Stephen M., William L. Thompson and Giorgio Ganis. 2006. *The Case for Mental Imagery*. New York: Oxford University Press.
- Laeng, Bruno, Matia Okubo, Ayako Saneyoshi and Chikashi Michimata. 2011. Processing spatial relations with different apertures of attention. *Cognitive Science*, 35(1): 297–329.
- Li, Zhaoping. 2016. Olfactory object recognition, segmentation, adaptation, target seeking, and discrimination by the network of the olfactory bulb and cortex: Computational model and experimental data. *Current Opinion in Behavioral Sciences*, 11: 30–39.
- Li, Zhaoping. 2019. A new framework for understanding vision from the perspective of the primary visual cortex. *Current Opinion in Neurobiology*, 58: 1–10.
- Marr, David. 1982/2010. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Cambridge, MA: The MIT Press.
- Norman, Donald A. 1988. *The Psychology of Everyday Things*. New York: Basic Books.
- Oliva, Aude and Antonio Torralba. 2007. The role of context in object recognition. *Trends in Cognitive Sciences*, 11(12): 520–527.
- Rubin, Edgar. 1915/1921. *Visuell wahrgenommene Figuren*. Copenhagen: Gyldendalske Boghandel.
- Schober, Michael F. 1993. Spatial perspective-taking in conversation. *Cognition*, 47: 1–24.
- Stocker, Kurt. 2012. The time machine in our mind. *Cognitive Science*, 36(3), 385–420.
- Stocker, Kurt. 2014a. The elements of cause and effect. *International Journal of Cognitive Linguistics*, 5(2): 121–145.
- Stocker, Kurt. 2014b. The theory of cognitive spacetime. *Metaphor and Symbol*, 29(2): 71–93.

- Stocker, Kurt. 2015. Toward an embodied cognitive semantics. *Cognitive Semantics* 1(2): 178–212.
- Stocker, Kurt and Joachim Funke. 2019. How we conceptualize climate change: Revealing the force-dynamic structure underlying stock-flow reasoning. *Journal of Dynamic Decision Making*, 5: Article 1.
- Talmy, Leonard. 2000. *Toward a Cognitive Semantics. Vol. 1: Concept Structuring Systems*. Cambridge, MA: The MIT Press.
- Tversky, Barbara and Bridgette Martin Hard. 2009. Embodied and disembodied cognition: Spatial perspective-taking. *Cognition*, 110(1): 124–129.
- Wolff, Phillip. 2003. Direct causation in the linguistic coding and individuation of causal events. *Cognition*, 88(1): 1–48.
- Wolff, Phillip. 2007. Representing causation. *Journal of Experimental Psychology: General*, 136(1): 82–111.
- Zelano, Cristina, Moustafa Bensafi, Jess Porter, Joel Mainland, Brad Johnson, Elizabeth Bremner, Christina Telles, Rehan Khan and Noam Sobel. 2005. Attentional modulation in human primary olfactory cortex. *Nature Neuroscience*, 8(1): 114–120.