# Abstract

Vision has a double life. On the one hand, vision represents the environment by describing it in a certain way. On the other hand, in order for visual descriptions to guide the execution of actions, vision must represent the environment by prescribing which motor acts are needed for an agent to complement with the relevant properties of the environment. Vision is, thus, Janus-faced: tied to both describing the environment as offering possibilities of action, as well as prescribing the motor acts that complement with that possibilities. While it may seem obvious that vision guides actions, it is no so obvious how the descriptions and the prescriptions it provides come together. In the recent years, though many have ventured into accounting for how visual processes represent the action-related properties of selected targets in the environment, few attention has been devoted to how vision actually prescribes the execution of the motor acts enabled by these properties. In this paper, we offer a framework that allows us to account for how visual representations can prescribe the way to execute an action.
Vision in Action is *Janus-faced*

Vision has a *double life*. On the one hand, vision represents the environment by *describing* it in a certain way. On the other hand, in order for visual descriptions to guide the execution of actions, vision must represent the environment by *prescribing* which motor acts are needed for an agent to complement with the relevant properties of the environment. Vision is, thus, *Janus-faced* tied to both *describing* the environment as offering possibilities of action, as well as *prescribing* the motor acts that complement with that possibilities. While it may seem obvious that vision guides actions, it is no so obvious how the *descriptions* and the *prescriptions* it provides come together. In the recent years, though many have ventured into accounting for how visual processes represent the action-related properties of selected targets in the environment, few attention has been devoted to how vision *actually prescribes* the execution of the motor acts enabled by these properties. In this paper, we offer a framework that allows us to account for how visual representations can *prescribe* the way to execute an action.

1. Introduction

Vision plays a crucial role in guiding the execution of our actions. Drinking a coffee, biting an apple, catching a ball on the fly, in all similar cases visual perception enables us to dynamically and smoothly execute the movements that we use in order to act in a satisfying manner. In doing this, however, vision has a *double life*.

On the one hand, vision represents the environment by *describing* it in a certain way, for example, as having relevant properties the representation of which leads to the execution of certain actions, on the basis of the agent’s bodily configuration, motor skills and with respect to the motor context (e.g., Clark 2001, 2007, 2009; Nanay 2011, 2021a, 2021b, 2013; Ferretti 2019, 2020b, 2021a, 2021b, 2021c; Ferretti and Zipoli Caiani 2019, 2021). Indeed, when visual perception fails to provide information about the relevant properties of the environmental targets, we usually fail in executing the related actions (Ibid.).

On the other hand, in order for visual descriptions to guide the execution of actions, vision must represent the environment by *prescribing* which motor acts are needed for an agent to complement with the relevant properties of the environment. In other words, to guide the execution of an action, it is not enough that vision detects the presence of the relevant action-related properties in the environment. Rather, it must also relate the visual information about such properties to the motor

---

1 It should be noted that the term “prescription” can be intended in many ways. Among the many meanings of this term, we intend a prescription as “the action of laying down rules or directions” (Merriam-Webster dictionary), namely, as something that admits being satisfied by a certain action or disregarded by a different one. Importantly, the term prescription is not intended here as a mandatory order that cannot be unattended.
skills, leading to the generation of the performances that are appropriate to act upon them. Thus, \textit{description} and \textit{prescription} must interlock in visual processing.

Importantly, were the function of vision limited to supplying descriptive information regarding the presence of certain environmental properties, vision would have no \textit{active} function in guiding actions. However, functional evidence suggests that vision plays more than a passive role.

It is well known, indeed, that a \textit{selective impairment} of the visual processing may cause the loss of the ability to correctly execute the motor acts that constitute an intended action, although the ability to \textit{visually represent} the properties of the environment remains unaltered. Notably, agents who are able to visually recognize an object and \textit{describe} it as offering action-related properties may lose the ability to \textit{prescribe} which motor acts complement with the relevant properties of the environment, without this implying damage to the musculoskeletal system. For example, patients suffering from \textit{optic ataxia}, an impairment concerning the dorsal visual system, or, better said, the visuomotor system responsible for vision-for-action, are no longer able to visually guide their motor acts according to their intentions, on the basis of their visual information on object properties, while, however, they are still able to visually recognize and describe that objects as endowed with action-related visual properties (Himmelbach & Karnath, 2005; Pisella et al., 2017; Schindler et al., 2004; Zipoli Caiani, 2017; Ferretti 2021a, see also Section 3 of this paper).

In sum, vision is \textit{Janus-faced}, tied to both \textit{describing} the environment as offering possibilities of action, as well as \textit{prescribing} the motor acts that complement with that possibilities. These two processes are two faces of the same \textit{visuomotor} coin. Now, though it may seem obvious that in many cases visual perception guides our actions, it is no so obvious how, in such cases, the \textit{descriptive} and the \textit{prescriptive} functions of vision come together.

Indeed, in recent years, the study of the interactions between vision and action has been the focus of a significant portion of the philosophical and cognitive science literature. Such an interest has allowed philosophers and scientists to point out the close connection between our perceptual and motor abilities, fostering a proliferation of empirical studies concerning the interaction between the visual and the motor systems, which have been deeply analyzed in the literature (this is just the tip of the iceberg of a very massive literature: Clark 2001, 2007, 2009; Brogaard, 2012; Chemero, 2011; Jacob & Jeannerod, 2003; Milner & Goodale, 1995/2006; Nanay, 2013; Noë, 2004; Zipoli Caiani and Ferretti 2016, 2019, 2020, 2021a, 2021b, 2021c Briscoe and Grush 2017).

However, despite assuming that \textit{descriptions} and \textit{prescriptions} are in play when it comes to \textit{vision-for-action}, this literature has not yet addressed how those two aspects fit together. Importantly, the initial hypothesis that the descriptive function and the prescriptive function of vision were instantiated by two (partially) segregated and sequential processes (e.g., Hoeren et al., 2013; Jacob & Jeannerod, 2003; Jacob & Vignemont, 2010; Milner & Goodale, 1995) has proved unsatisfactory in the light of the empirical
findings. Contrary to such an initial view, evidence show that the information flow subserving the visual detection of action possibilities in the environment and the information flow subserving the selection of a suitable motor plan for action integrate each other, rather than occurring sequentially (Zipoli Caiani and Ferretti 2017 for a review). So far, the most important lack is that, although many have ventured into accounting for how visual process represent the action-related properties of selected targets in the environment (for a discussion see: Ferretti, 2016, 2019, 2020, 2021a, 2021b, 2012c; Nanay, 2011, 2012a, 2012b, 2013; Siegel, 2014; Jacob & Jeannerod, 2003), fewer attention has been devoted to how vision actually prescribes the execution of the motor acts enabled by these properties. But, since not only does vision describe the external world, but it also prescribes which motor acts may be used on this world², the prescriptive function of vision-for-action has to find an explanation within those theories, along with its descriptive one. Then, we need a framework capable of explaining how vision describes the targets we act on and, at the same time, prescribes the actions to be performed on them.

In this respect, the reader should note that the prescriptive function of vision is of crucial importance to understand the difference between two kinds of visual representations: the visual representations that guide the execution of our actions and those that do not. Indeed, while the latter merely provide information about the visual properties of the environment, the former (especially its visuomotor component) also provides instructions on how to interact with these objects. This is the reason why – and this is also the crucial point – vision is not limited to describing the objects of the environment as action-able targets, but it is also involved in prescribing which motor acts have to be performed on the available targets in order to guide the execution of our actions.

Things being so, and given for granted that the descriptive aspect of vision-for-action has been extensively accounted for in the literature on visual processing, it becomes of particular importance to understand how visual representations can perform a prescription about the way to execute visually guided actions. So, here is the main question this paper is about, granted that visual representations describe the external environment:

Q: How does a visual representation prescribe the way to execute an action?

To answer this question entails facing two puzzles.

The first puzzle concerns the mental architecture that subserves our ability to perform actions that are guided by our capacity to visually perceive action-related targets in the environment, that is, the mental mechanisms that underlie the interaction between visual perception and motor execution in the

---

² Indeed, “we are right to think that perceptual experience informs us about the world we act in and that we act in the very world we perceive” (Clark 2007: 590).
visual control of an action. Importantly, solving this puzzle consists in establishing *what* mental states lead visual perception to determine the course of our actions in the way of action generation. Thus, the first puzzle related to Q is:

P1: What are the mental states that interface with vision in the guidance of actions?

Provided that, in order to face the first puzzle we have to posit the presence of both visual and motor states, this puzzle concerns the very connection between these states, namely, the way in which the information offered by the visual system interlocks with the information processed by the motor apparatus. At this point, it is worth noting that the challenge is to explain in which way the visual guidance of an action goes well with the intentional guidance of that action, so that vision may prescribe the execution of the motor acts according to the related action intention. Thus, the second puzzle starts as soon a solution to the first one is offered. Provided that there is a link between visual prescriptions and the actions intentions (as it will be shown in section 2.3):

P2: How do visual representations interlock with action intentions in the guidance of actions?

In this article, we propose an answer to our initial question Q by addressing both the related puzzles P1 and P2 (see also fig. 1 and 2). More precisely, we explain how visual representations can *prescribe* the execution of the visually available actions in two steps: first, we establish *what are* the mental representations involved in the execution of those actions (P1); second, we clarify *the way in which* these representations interlock with the last mental antecedents of actions, namely, motor representations (P2). In particular, the identification of the mental components that contribute to the visual guidance of an action (the solution to P1) and of the related connections to motor representations (the solution to P2) allow us to clarify how vision can guide an action by prescribing the appropriate motor acts in relation to the relevant motor intention. This is precisely what provides an answer to our initial question Q.

Note that this move is not trivial. While the literature has given for granted that there are visual states related to action, namely visuomotor states, nobody has provided a satisfactory analysis of the way their *descriptive* and *prescriptive* functions march in step in intentional action. The marriage between the two functions is *always assumed*, but *never explained*. This is clear by taking a look at the relevant literature on this point. Scholars have mainly focused on analyzing vision-for-action, mostly discussing the evidence about visuomotor states, in order to establish whether we *see action possibilities* and whether this vision is or is not conscious (Nanay 2013; Jacob and Jeannerod 2003; Ferretti 2016, 2021a, 2021b,
However, it remains to be explained the way the descriptive and the prescriptive aspects of vision for action interlock in intentional action.

1.2. The structure of the argument

The article proceeds as follows. In Section 2, we address P1 concerning what mental states, together with visual representations, underlie the execution of visually-guided actions. We start by dwelling on two general proposals which, more or less explicitly, are still influential in the contemporary debate. First (§2.1), we consider what we call the conscious vision hypothesis, according to which vision guides actions by means of the conscious representation of the action-related properties of the environment. Then (§2.2), we consider what we call the epistemic hypothesis, according to which vision guides action by fostering a propositional belief about how this action is to be performed. In both cases, however, we offer arguments showing that these hypotheses do not provide a suitable answer to P1. More precisely, we contend that, contrary to the conscious vision hypothesis, we do not consciously detect the action-properties in the visual environment, while, contrary to the epistemic hypothesis, we argue that propositionally structured beliefs are not the right kind of mental states that can determine the execution of an action.

Finally (§2.3), we consider a third alternative view, the intention hypothesis, according to which visual representations guide actions by interfacing with the relevant intentions to act. We defend this hypothesis by reviewing the available evidence proving that visual representations are strictly connected to our intentions since, not only do the former determine our intentions to act, but also the latter do shape the way we visually represent the environment.

In Section 3, on the basis of the intention hypothesis, we address P2, which is about how visual representations and action intentions interlock in visually-guided actions. In order to face this puzzle, we first provide arguments showing that action intentions have motor representations as their constituents. For this purpose, we draw on the Same Format Thesis (Ferretti & Zipoli Caiani, 2018, 2021b; Ferretti 2020), according to which, in order for an intention to determine the execution of an action, it must be constituted by an executable action concept which is instantiated in natural system as a motor representation (see also, Burnston, 2020; Shepherd, 2019 for similar views). Then, we offer empirical evidence proving that the same motor representation that constitutes our intention to act shapes the way we visually represent the environment. If our analysis is correct, we have a strong reason to assume that the visual representations that guide our actions and the underlying intentions share the same motor representations. In other words, vision prescribes actions by sharing the same motor representation that constitutes the related intention to act. We call this conclusion Visuo-Motor Thesis.

Our conclusion is crucial not only because it answers to our initial question Q, concerning how a visual representation prescribes the execution of an action, but also because it explains, for the first time
in the literature, the way in which the visual representations that guide our actions can have a double life as to be both descriptive and prescriptive. The view we cash out is that vision allows us to describe the environment by providing propositional information about the presence of action-related properties, while it allows us to prescribe the way to execute an action, inasmuch as such propositional information is constituted by the executable action concept that form our intention to act.

2. Explaining the prescriptive role of visual representations in visually-guided action

In the previous section (§1.1), we said that in order for vision to guide actions, the information offered by the visual input must interface with the information processed by the motor system. The question is: what are the mental states involved in such an interface between vision and action such that vision can perform a prescriptive task toward motor responses? Answering this question amounts to provide a solution to our puzzle P1 (§1, see fig.1).

![Fig.1](image)

A schematic representation of the puzzle P1

In this section, we consider three hypotheses about what mental states are involved in the visual guidance of an action. Such hypotheses are, respectively, the conscious vision hypothesis, the epistemic hypothesis, and the intention hypothesis. Then, we argue that while the conscious vision hypothesis, and the epistemic hypothesis suffer from flaws, the intention hypothesis is the best candidate to solve our puzzle P1.

2.1. The Conscious Vision Hypothesis

There is a lively debate on whether conscious visual experience guides action. On the one hand, the common sense seems to suggest that the properties of objects we consciously see are those we use to move in our environment (for different labels of this view and its different aspects, see Ferretti and

---

3 It might seem that the conclusion of this argument is the same of the nowadays classic account of perception developed by Jacob and Jeannerod (2003), according to which actions are guided by non-conceptual visuomotor representations. However, although according to Jacob & Jeannerod (2003), vision has a double life, such a view differs crucially from our thesis. In particular, according to Jacob and Jeannerod (2003), visuomotor representations serve as input for a different form of representation, namely, action intentions (Ibid., §§8.1, 8.2). Otherwise, we argue, mainly on the basis of empirical evidence (see Section 3 of this paper) that the visuomotor representations that guide our actions share the same cognitive resources of action intentions. Furthermore, our view is not committed with the assumption that the content of visuomotor representations has a non-conceptual content. Differently, we state that the content of our action-guiding visual representation has a prescriptive function, which can be suitably performed by an executive action concept (Pacherie, 2018). Finally, we do not endorse a segregation hypothesis as that assumed by Jacob and Jeannerod (2003), according to which the dorsal and the ventral streams do not functionally interact each other until the end of the processes (for a discussion, see Zipoli Caiani and Ferretti, 2017).
Zipoli Caiani 2018). This is what we call here the *conscious vision hypothesis*, according to which the type of mental states involved in the visual guidance of an action are conscious visual states.

**Conscious Vision Hypothesis**: Conscious visual experience triggers and shapes motor action.

According to this hypothesis, conscious visual experience would prescribe the action to be executed through its contents, so that this would be the answer to P1: the conscious visual experience of objects guides the motor reaction of the subject (see fig. 2).

Several problematic things have been addressed about this hypothesis – independently of the way it answers to P1, something we are not focusing on here. The main intuition behind this hypothesis has been challenged by the *Two Visual Systems Model* (Milner and Goodale 1995/2006). Long story made short, according to this model, the portions of our visual system that guide actions are not the same that offer us conscious visual experience. Two systems would run in parallel. A dorsal stream subserves unconscious vision, which then registers shape properties for motor purposes. A ventral stream subserves conscious visual experience of those properties. So, while the object source is the same, e.g. an apple, visual experience of the apple and motor programs to act toward it are shaped by different systems.

![Fig.2]

This idea also seems to challenge the view that we consciously see *affordances*, namely action possibilities offered by objects (Gibson 1979/1986). A view that has been defended, in different ways, in the philosophical literature (Siegel 2014; Nanay 2011, 2012a, 2021b, 2013).

That said, recently, it has been suggested that both streams, *per se*, can participate, even if partially, to both conscious recognition and action guidance. And, more important, that they interact in a way that allows to have a sort of conscious visual guidance of action (the literature is massive, for a review of these two points, see Ferretti 2018, 2021c; Briscoe and Schwenkler 2015).

However, in the light of this evidence, it has been suggested that even in case that vision-for-action can be conscious, so that conscious visual experience guides our motor programs, this falls short of the claim that we see action possibilities. Not only that we consciously see them, but that we see them either consciously or unconsciously. Indeed, it has been noted that, at best, what the new results
of the model suggest is that we can consciously see spatial properties, whose visual experience can, if this new view is right, directly guide our action. However, it should be noted that this does not amount to claim, sensu stricto, that there is something as conscious or unconscious vision of action possibilities. Seeing action possibilities entails seeing motoric information. But this claim is more demanding, and does not follow from the claim that we see spatial properties that our visual system flags to motor processing in order to shape an action program (for the complete review of this point, see Ferretti 2019, 2021a).

If that is true, then, even if we can consciously see spatial properties that are those used to shape motor processing, it does not follow that we consciously see action possibilities.

The reader should note that saying that we consciously see spatial properties that then guide action does not capture the idea that vision can prescribe the way a motor action can unfold. How the mere visual encoding of spatial properties can generate the prescription of a given motor act? Such a visual prescription would need more information about the action to be performed, the context, and the intention of the subject. Arguably, seeing that we see action possibilities would be a much more intriguing claim to spell out the prescriptive nature of vision. Suppose one sees an action possibility in the environment, say grasping an apple with a power grip that allows her to bring it to the mouth, conscious vision would, in this case, prescribe which sequence of motor acts she’d need to perform based on her intention of biting the apple. But there is no tenable reason in the literature to suggest that we consciously see action possibilities instead of spatial configuration of the environment (this has been show on the philosophical analysis of behavioral, phenomenal and neurobiological evidence on the perception of affordances, or action possibilities).

Notably, many pieces of evidence have shown that, in order for visual perception to guide an action, it is not necessary for visual representations to be conscious. For example, patients suffering from blindsight, an impairment due to a damage to the primary visual cortex (V1), show loss of visual consciousness in the affected visual field, but some of them can still exhibit the capacity of visually represent the environment by means of which they can suitably direct their hand or sight to an action target in the blind field (Danckert & Rossetti, 2005; Kinoshita et al., 2019; Pöppel et al., 1973; Tamietto et al., 2010). This supports the view that the visual perception of the environment shapes or determines our intentions to act, even independently of a peculiar associated phenomenology. Think also about the case of visual agnosia, in which conscious visual recogniton is not at work, but action performance is still in play (Milner and Goodale 1995/2006; Jacob and Jeannerdo 2003). Then, we need another candidate to explain the prescriptive nature of vision-for-action.
2.2. The Epistemic Hypothesis

According to the epistemic hypothesis, in order for vision to guide the execution of an action, two conditions have to be met. First, visual perception must represent an object as the suitable target for a certain action, then a (justified, true) belief must be formed about how the token action is to be performed. In other words, the epistemic hypothesis entails that vision guides an action through the suitable (practical) knowledge, that is, a propositional structure concerning the way to perform that action. Unlike the conscious vision hypothesis, the epistemic hypothesis assumes that, independently of any conscious connotation, a belief has to be the intermediate in between the visual and the motor state. This description paves the way for an analysis of the relationships between vision and action on a sub-personal level where consciousness does not play any role (Bratman, 1987; Davidson, 1963; Sinhababu, 2013; see also Nanay 2013 and Ferretti 2020 for a discussion of related examples).

More precisely, for the epistemic hypothesis, the visual perception of action possibilities and the related practical knowledge needed to execute them figure, respectively, as the premise and the conclusion of a propositional calculus (Fodor, 1983; Gregory, 1997; Marr, 1981 for classical loci). On this view, indeed, the visual representation that an object is action-able in a certain way involves the belief that such a way is suitable to perform the related action on this object, so that, in the appropriate circumstances, such a belief determines the execution of such an action, in precisely this way.

In sum, the epistemic hypothesis provides an answer to P1 by maintaining that vision guides the execution of an action, inasmuch as our visual representation of the relevant properties of the environment involves the beliefs which suitably describe the execution of that action (fig. 3).

Epistemic hypothesis: Visual representations and epistemic beliefs about action execution interlock in visually guided actions.

Importantly, for vision to involve the belief that determines an action, visual perceptions and beliefs must be conceived as mental states endowed with a propositional content, since propositions are the only suitable structures to figure in a logical calculus. Accordingly, if the epistemic hypothesis is true, our visually-guided actions are determined by the combination of the appropriate propositional representations.

![Fig.3](A schematic representation of the epistemic hypothesis.)
The problem with this view is that the literature is replete of arguments according to which propositional representations are not suitable to determine how to execute an action (e.g., Dickie, 2012; Ferretti & Zipoli Caiani, 2018, 2021b; Ferretti 2019, 2020; Fridland, 2013; Levy, 2015; Nanay, 2013; Noë, 2005). Indeed, there are at least two main reasons to be skeptical about the epistemic hypothesis.

The first reason is that, since the epistemic hypothesis entails only propositional structures to figure in the mental processing that triggers the action, it does not have enough resources to establish how coarsely or finely individuated is the way to execute a certain action (Fridland, 2013; Levy, 2015; Pacherie, 2008, 2011; Shepherd, 2019; Ferretti 2020; Ferretti and Zipoli Caiani, 2021b). Many actions, undeniably, require the execution of a large number of motor acts, so the problem is to establish how grained the propositional description of the action-related properties of the environment and the related movements should be. It is commonly agreed, indeed, that many actions display a sophistication that fails to be accounted for by a limited number of propositional tokens (Pacherie 2001; Jacob and Jeannerod 2003; Butterfill and Sinigaglia 2014; Ferretti 2020). To have a taste of this sophistication, consider that the hand has over 20 degrees of freedom and that hand movements require the coordinated interplay of 39 muscles acting on 18 joints, which allow to this structure a surprising number of action possibilities (Raos, et al. 2006). Things being so, the issue is to determine how many and which propositional contents are required for representing the relevant information related to the entire plethora of motor acts allowed by the hand. There are two options available here.

On the one hand, one may opt for a finely grained analysis, which would yet give rise to an explosion of propositional representations, with the consequence of making the computation underlying the visual guidance of our actions an extremely complex process, of which there is no evidence (Ferretti and Zipoli Caiani 2021b). On the other hand, one may opt for a coarser analysis and assume that also a less finely grained propositional description can account for how our visual representations guide our actions, with the consequence of considering only a portion of the motor opportunities offered by our bodies, which would be problematic when explaining the huge complexity of action performance (Ferretti 2020).

The second reason to be skeptical about the epistemic hypothesis is that, though the visual representation of a target for action involves the belief that a certain way is suitable to execute that action on that target, not all agents that form such beliefs are able to execute that action in the suitable way. The problem stems from the fact that, having the propositional representation of how to execute an action in a certain way does not always give rise to the ability to execute that action.

Consider the case in which one has the true belief that executing a certain chain of movements is the way to successfully perform a complex stunt, but that for some reason, she is not able to perform it herself (Dickie, 2012; Ferretti 2020; Levy, 2015; Noë, 2004, 2005). Cases like this challenge the view

10
In sum, if our propositional beliefs about how to execute an action are not enough to allow us to perform that action, then, contrary to the epistemic hypothesis, our visual representations cannot guide our actions by merely involving (justified, true) beliefs about the way to execute those actions (see Ferretti and Zipoli Caiani, 2021b). Accordingly, the epistemic hypothesis does not provide a solution to the puzzle P1.

2.3. The Intention Hypothesis

In the previous section (§1), we said that in order to answer our initial question Q, concerning how a visual representation can prescribe the execution of an action, we must first provide an answer to the puzzle P1, concerning what mental representations trigger our visually-guided actions. To solve this puzzle, we have considered two hypotheses that are among the most influential when considering the relationship between vision and action. According to the conscious hypothesis (§2.1), we represent the action-related properties of the environment by means of conscious mental states, while for the epistemic hypothesis vision guides the execution of an action through the function of a class of unconscious propositional representations concerning the way to perform that action (§2.2). However, both of them proved to be fraught with daunting problems.

In this section, we focus on a third proposal, namely, the intention hypothesis, which is the one we consider as being the best candidate to provide an answer to P1. According to the intention hypothesis, in order for a visual representation to guide an action, it must be integrated with the related intention to act.

Before addressing the intention hypothesis, let us briefly discuss what an intention to act is. An intention to act is usually conceived as a prescriptive mental antecedent of an action, that is, as a representational state whose condition of satisfaction is the execution of a certain action (e.g., Bratman, 1987; Mele, 2009; Pacherie, 2011; Searle, 1983). Unlike beliefs, action intentions do not refer to an already existing state of affair, but rather bring about changes in the world, so that the world matches the content of the intentions. More precisely, in order for an action to bring about changes in the world, it must interface with the relevant motor representation, namely, with the last mental antecedent that precedes the execution of an action (Brozzo, 2017; Burnston, 2017; Butterfill & Sinigaglia, 2014; Ferretti & Zipoli Caiani, 2018, 2021a, 2021b; Ferretti 2020; Pacherie, 2008a, 2018; Shepherd, 2017; Jeannerod 2006; Jacob and Jeannerod 2003). In other words, an intention to execute an action can be satisfied by having it interfacing with the appropriate motor representation capable of correctly generating the intended action.
While the literature has extensively focused on the interlock between motor representations and intentions (Butterfill and Sinigaglia 2014; Burnston 2017; Shepherd 2017; Mylopoulos and Pacherie 2017; Ferretti and Zipoli Caiani 2018, Fridland, 2021), as previously showed, nobody has meticulously investigated the way visual representations relate to the intentions to act. Despite of this, there are good reasons to believe that the execution of a visually-guided action depends on the interaction between the visual representation of the relevant properties of the environment and the particular intention that underlies the execution of that action, with the proviso that also the motor representation interlocks with a given intention.

Clarifying this point is crucial, as it helps us to understand how vision guides the execution of our actions in a prescriptive manner: it is commonly assumed that it is because motor representations and intentions interlock that we can act in a suitable manner. However, and this is something usually overlooked, but decisive, it is because vision interfaces with our intentions of acting that it can have an influence on the way we execute the related motor acts that form the intended action. Let’s call this the intention hypothesis (see fig. 4):

**Intention hypothesis**: visual representations interlock with intentions to act in visually guided actions.

In order to defend the intention hypothesis, one may start by recognizing that our intentions to act usually follow our visual representations of the environment. It seems obvious, indeed, that it is because one sees certain properties of the environment, say a certain orientation of a handle, that she forms the intention to grasp it in a certain way. Of course, one can also form the intention to perform a certain action prior to form a visual representation of the related target, but nevertheless she needs to rely on visual information to drive the action compatible with the intention. Think to the case of catching an apple. Though one may form an intention to act just before to visually perceive the apple, she needs to rely on visual information about the apple in order to perform the grasp. And, on the basis of the action she wants to perform, she needs to visually focus on the apple and intentionally shape an appropriate motor act (e.g., a power grip). In other words, it could be said that our intentions to act counterfactually depend on our visual representations.

![Fig. 4](image)  
A schematic representation of the intention hypothesis.
Importantly, evidence shows that not only does visual perception determine our intentions, but our intentions do change the way we visually perceive the properties of the environment too: visual representations seem to counterfactually depend on the actions we want to perform, i.e. on the intentions to act in a very specific manner (see Nanay 2018). Accordingly, the research in the field of visual attention has established that the visual selection of an action target is biased by the underlying intention to act. For example, Wykowska et al., (2009; see also Wykowska & Schubö, 2012) have showed that a preparatory grasping movement facilitates the detection of targets, resulting in faster reaction times compared to trials in which a pointing movement has to be prepared. Similarly, planning the execution of a grasping movement was reported to facilitate the visual detection of orientation targets compared to a pointing task (Bekkering & Neggers, 2002). Moreover, preparing the execution of a precision grip has been shown to facilitate the visual perception of a change in small objects, while preparing a power grip has been shown to foster the visual perception of a change in larger targets (Symes et al., 2008). Furthermore, in their experiment, Gutteling et al. (2015) showed that the action an agent intends to perform influences the early visual cortices, though agent’s retinal image remains unaltered (see also, Craighero et al., 1999; Dötsch et al., 2017; Fagioli et al., 2007).

Pieces of evidence such as these show that one’s intention to execute an action depends on the visual perception of the environment (Danckert & Rossetti, 2005; Kinoshita et al., 2019; Pöppel et al., 1973; Tamietto et al., 2010), but also that one’s visual perception depends on the action one intends to perform (in accordance with Nanay, 2018). Importantly, this view is consistent with the hypothesis that visual representations and action intentions directly interlock in visually guided actions, that is, without involving intermediate mental states, such as conscious states and beliefs (§§ 2.1, 2.2). Indeed, if conscious states and beliefs were mediating between visual representations and intentions to act, the bidirectional interaction between the activity of the visual system and the activity of the intentional system would not occur. Therefore, it is reasonable to assume that the intention hypothesis is not plagued by the problems afflicting the conscious vision hypothesis and the epistemic hypothesis.

Importantly, this provides a solution to P1, that is, to the puzzle concerning what mental states are involved, together with vision, in the visual guidance of actions. Our visually-guided actions are triggered by the interaction between our visual representations and our intentions to act, as stated by the intention hypothesis. Vision, thus, prescribes the way to execute an action by interlocking with the related intention to act.

This is however only half of the story about our initial question Q, concerning how does a visual representation prescribe the way to execute an action (§1). Indeed, to answer this question, it is not enough to answer the puzzle P1, showing that visual representations interface with action intentions in guiding actions, but it is also necessary to show how these different mental states actually interface each other. Accordingly, we still need to provide a solution to the puzzle P2, concerning how visual
representations interface with action intentions in the guidance of actions. This is the aim of the next section.

3. How intentions and visual representations interface in visually guided actions

In the previous section, we suggested that the intention hypothesis is a viable candidate to solve the puzzle P1 concerning what mental representations trigger the execution of visually-guided actions. The intention hypothesis assumes that, in order to guide an action, the visual representations of the action-related properties of the environment interlock with the intentions that determine the execution of those actions.

The reader should note that the relevance of this hypothesis is that, in addition to being empirically supported (§2.3), it does not suffer from the same issues that plague the conscious vision hypothesis and the epistemic hypothesis, inasmuch as it does not rely on conscious visual states and epistemic beliefs.

So far so good. P1 has been addressed. Now, in order to answer our main question Q, and establish how does a visual representation prescribe the way to execute an action, we must provide a solution to the second puzzle P2, about how visual and action intentions interface in visually-guided actions. More precisely, since the intention hypothesis assumes that our visual representations relate to our intentions to act, we have now to explain how these two mental states do actually interlock in the guidance of our actions (Fig. 5).

![Fig.5](image)

**Fig.5**

A schematic representation of the puzzle P2

To answer this question, we must first establish how action intentions interface with the motor representations that determine our actions (Burnston, 2017; Butterfill & Sinigaglia, 2014; Ferretti & Zipoli Caiani, 2018, 2021a, 2021b; Ferretti 2020; Pacherie, 2008a, 2018; Shepherd, 2017a; Jeannerod 2006; Jacob and Jeannerod 2003). To begin with, it is important to understand what is the representational format of the intentions underlying our actions, and then to establish the way in which our visual representations interlock with such intentions. Hence, we draw on a recent solution to this issue, also
known as the *Same Format Thesis* (SFT) (Ferretti & Zipoli Caiani, 2018, 2019, 2021a, 2021b, see also Burnston, 2020, Shepher, 2019 for similar views). According to SFT, an action intention is a propositionally structured representation constituted by an *executable action concept*. According to a common definition, executable action concepts are those concepts that activate the appropriate motor representation for a given type of action, so as to allow one to intend to perform this action (for a detailed introduction to this notion, see Israel et al., 1993; Gallese and Lakoff, 2005; Mylopoulos & Pacherie, 2016; Pacherie, 2011; Pavese, 2021)4 In other words, an executable action concept can be seen as reflecting an instruction to guide the activity of the body, through the appropriate motor representation, in order to execute the chain of motor movements that fits with the agent’s intention to act.

Before continuing, it should be noted that SFT is of crucial importance for understanding how vision can guide our actions. Indeed, in order to understand how our visual representations leads the execution of our motor acts, two questions must be addressed: (i) it must be established how our visual representations interface with our intentions of action, as recognized by the *intention hypothesis*; then, (ii) it must be established how our intentions of action interface with the motor representations that ultimately determine the course of the related actions. Importantly, SFT proposes a solution to the second question (ii), stating that for an intention to determine an action, it must be constituted by the relevant motor representation. Now, if SFT is true and our intention to execute an action is constituted by the motor representation of that action, as it appears according to the recent literature, then it is reasonable to assume that vision guides our actions inasmuch as visual representations and action intentions share the same relevant motor representations. Let’s call this the *Visuo-Motor Thesis* (VMT, see fig. 6):

**VMT**: Visual representations and action intentions share the same motor representations in the guidance of actions.

VMT has relevant features. First, it should be noted that VMT is consistent with the *intention hypothesis*, since both agree in assuming that visual representations and intentions to act interface in the visual guidance of our motor actions (§2.3). Second, VMT does not mention any role for conscious states and epistemic states, like beliefs, so that it does not suffer from the same issues plaguing other options (§ 2.1, 2.2). Moreover, VTM admits empirical consequences. Notably, if VMT is true, we can

---

4 Following Pacherie (2011), one can distinguish between two different types of action concept. On the one hand, there are those action concepts that do not hook up with the corresponding motor representations. These concepts, allow us to accurately categorize instances of action that fall under them, even without being able to perform the action ourselves. On the other hand, there are action concepts that allow us to activate appropriate motor representations, so that having the possibility to perform an action entails forming the appropriate executable action concept related to the action we are able to perform. The main difference between an executable and a *non*-executable action concepts is that the former involves the suitable representation of how to execute the related action, while the latter does not (Pacherie, 2011).
expect that the relevant areas of the brain are functionally activated by the processing of the visual stimulations linked to the action-related properties of the visual target in the environment (see §3.2; for a review, see also Ferretti 2021a). Accordingly, if VMT is correct, and the parietal and motor cortex are function involved in the processing of the visual representations of the action-properties related to a target, we gain an answer to the puzzle P2, concerning how vision and intentions interface in the guidance of an action.

According to VMT, vision is linked to action intention since they have in common, i.e. they computationally share, the motor representation of the relevant action. To support the view deriving from this thesis, we provide empirical evidence to show that, in the visual guidance of an action, also a visual representation can be conceived as a propositional structure constituted by the same executable action concept that forms the relevant intention to act.

Importantly, since VMT is about the way visual representations and action intentions interlock in the visual guidance of an action, it allows us to tackle our initial question Q, and thus to explain how visual representations can prescribe the execution of the available actions: it is because visual representations are constituted by a prescriptive motor representation that they are able to guide the suitable motor acts on the available targets.

### 3.1 How visual representations interface with action intentions, along with motor representations

We can now take a step forward and tackle our second puzzle P2, which concerns how visual and action representations interlock in visually-guided intentional action (cfr. Section 1). As anticipated, we maintain that P2 can be solved by means of VMT, that is, by assuming that visual representations and action intentions are constituted by the same motor representations. To support this view, we combine two previously established outcomes, namely, the intention hypothesis (§2.3) and SFT (Ferretti & Zipoli Caiani, 2018, 2019, 2021a, 2021b). More precisely, in Section 2.3, we have addressed the puzzle P1 about what mental states are involved in the visual guidance of an action and, following the intention
hypothesis, we established that the visual representation of the action-related properties of the environment interlock with the intention that cause the execution of the related action. Then, we draw on SFT (Ferretti & Zipoli Caiani, 2018, 2019, 2021a, 2021b), which establishes that in order for an intention to cause the execution of an action, it must be constituted by the relevant motor representation. More precisely, SFT provides a naturalistic framework for understanding how a motor representation can be related to the propositional content of an intention, by showing that the processing of motor representations and the processing of the executable action concepts that constitute the intentions share the same mental realizers. Things being so, an agent can propositionally intend to perform a certain action, because she possesses the executable action concept of that action, which is realized in the system as the related motor representation.

Now, since according to the intention hypothesis, in order for a visual representation to guide an action, it must interface with the intention that determines such an action, and since according to SFT in order for an intention to determine an action, it must be constituted by the motor representation that ultimately causes the execution of that action, then, it is reasonable to assume VMT as a viable solution to P2: in order to guide the execution of an action, a visual representation interlocks with an intention via the same motor representation that ultimately causes the execution of the related action. In other words, the visual information is related to the intention because visual information is part of a motor representation, and the motor representation shares the same realizer of the executable action concept in the intention (see fig.7). Importantly, VMT represents a significant advance over the literature considered so far, since it solves P2 by explaining how visual representations interface with action intentions and, ipso facto, it provides an answer to our main question Q by explaining how visual perception can guide the execution of an action.

---

5 A detailed defense of SFT and a clarification of the differences between SFT and other positions are provided in Ferretti & Zipoli Caiani, 2018, 2019, 2021a, 2021b), there is not necessity here to recap all the empirical evidence and arguments in support of this view that have been already provided.

6 The distinction between executable and non-executable action concepts is taken from Pacherie (2011). However, SFT differs from Pacherie’s approach to the interface problem (Mylopoulos & Pacherie, 2016), for the details see Ferretti and Zipoli Caiani, 2018.
Fig. 7
A schematic representation of how visual representations and action intentions interface by sharing the same motor representations.

Importantly, it must be noted that the assumption of VTM is also corroborated by empirical findings. Indeed, as previously stated (§3), if VMT is true, we can expect that the relevant areas of the motor cortex are functionally activated by the processing of the visual stimulations linked to the action-related properties of the target in the environment. Notably, VTM provides the best explanation regarding some important evidence about the processing of the visual stimulus. Over the recent years, indeed, a great deal of evidence has been provided to support the hypothesis that visually presented tools activate the same cortical areas that are functionally implicated in the planning and execution of those actions afforded by these tools. For example, in a nowadays classic experiment performed by Grafton et al. (1997) the activation of the dorsal premotor cortex has been recorded in agents observing frequently used tools. This result has been confirmed by another classic study conducted by Chao and Martin (2000) according to which the premotor cortex is sensitive to the visual detection of tools, and that the ability to detect motor-related information in a perceptual stimulus may depend on the function of the same cortical regions related to planning intentional interactions with the environment. Interestingly, Creem-Regehr and Lee (2005) have shown that the motor-related areas of the brain are active during the visual perception of action targets, and that such an activity is influenced by the practical meaning of the visual target for the agent. Indeed, although viewing an action target for which the practical function is unknown does not elicit the activation of the motor areas, the activity of the motor system is elicited if the agent knows how to execute a specific action with it.

Furthermore, experiments based on transcranial magnetic stimulation have been used to assess where and when the activation of the premotor system occurs during the observation of action-related objects. In an experiment conducted by Buccino et al. (2009), for example, the motor-evoked potentials (MEPs) from an area of the hand that is involved in grasping actions were recorded 200 ms after the presentation of a graspable object. Similarly, Makris et al. (2011) and Cardellicchio et al. (2011) have measured the presence of a strong motor-related activity in the parieto-frontal cortex induced by the presentation of a suitable visual stimulus around 300ms. More recently, Rowe et al. (2020) have found a significant activity in the parieto-frontal circuit between 280ms and 370ms after the visual stimulus onset (see also Rowe et al. 2017). Importantly, such results concerning the onset of an motor-related activity in the premotor areas show an involvement of these areas already in the first stage of the processing of the visual stimulus. On the basis of this, it is reasonable to hypothesize that the functional involvement of the premotor cortex in the processing of visual information is essential to guide the rapid execution of a visually guided action. Indeed, as already mentioned (§1), studies conducted on patients have shown that the selective damage of the parietal cortex linking the visual areas and the
premotor cortex results in a specific inability to correctly guide actions through vision, also known as optic ataxia (Cavina-Pratesi et al., 2013; Himmelbach & Karnath, 2005; Pisella et al., 2017; for a discussion Zipoli Caiani, 2017).

In addition, Pilacinski et al. (2018) have shown that the dorsal premotor cortex is sensitive to the visual consequences of an action, that is, to the intended visual outcome of a motor act. Notably, authors have shown that an activity related to the visual consequences of an action can be find in the premotor cortex during the early planning stage, suggesting that the premotor cortex codes for the action plan on the basis of the intended visual outcomes. Indeed, simulating the visual consequences of an action is a requisite to form the suitable intention to act given the available visual information.

Importantly, not only does the evidence show that the motor areas are sensitive to action-related visual stimuli, but also that the processing of the visual stimulus is actually influenced by agent’s motor representations. Evidence of the influence of our motor intentions on our visual representations has been provided by Gutteling et al. (2015), according to which preparing to execute an action, such as grasping an object or pointing to it, enhances the perception of the relevant features of the object. Similarly, Costantini et al., (2019) have shown that repeatedly performing an action on a target object influences subsequent visual categorization, affecting the time course of judgement about action-related visual properties of the manipulated target. A reasonable account of this effect requires involving the presence of rapid cortical feedbacks from the motor areas to the visual areas, so that it becomes reasonable to assume that motor processing and visual processing share, at least in part, the same neural substrate. More generally, the fact that the perception of action properties determines an activation of the motor areas and the fact that the preparation of a motor act determines an enhancement in the perception of the properties of action, can be explained by assuming that our visual representations interface with our intentions to perform a motor act by sharing with them the motor representation of the way to execute that act. Interestingly, this is precisely what is expected by assuming VMT.

To resume, in this section, we defended VTM by means of two different arguments. The first reason to assume VTM comes from combining the intention hypothesis (§2.3) with SFT (Ferretti & Zipoli Caiani, 2018). While according to the intention hypothesis our action guiding visual representations interface with our intentions to act, SFT states that our intentions determine actions since they are constituted by the motor representations of that actions. Accordingly, if the intention hypothesis and the SFT are true, it is reasonable to assume that the visual representations guiding our actions interface with our intentions to act by sharing with them the same motor representations, as stated by VTM. Moreover, the second reason to endorse VTM is that such a view provides an explanation for a great

---

7 A nowadays classic and widely shared theoretical framework points to the common coding of visual perception and action (Hommel et al., 2001; Prinz, 1997).
deal of evidence concerning the functional involvement of the premotor circuit in the processing of action-related visual information. Interestingly, VTM is a viable solution to P2, since it explains how do visual and action representations interlock in visually-guided actions.

Then, if our analysis is correct, we have all the ingredients to answer our original question Q about how can a visual representation prescribe the way to execute an action: the visual representation of an action target prescribes the way to execute the related action by sharing with the relevant intention to act the motor representation that causes that action.

4. Possible objections and replies about the relation between prescriptions and ascriptions in vision

There might be objections to the proposal that in the visual guidance of actions our visual representations of the environmental targets and our intentions to act on the same targets interface by sharing the same motor representations of the related actions. In what follows, we offer a reply to two possible objections, doing so provides the opportunity to further elucidate and clarify our proposal.

One objection may be that VMT does not explain how visual representations can have a double life. We started by saying that vision has a double life in guiding our actions (§1): on the one hand, vision describes the objects as having properties relevant to the execution of actions, on the other hand, it prescribes what actions are evoked by a target having such relevant properties. However, our argument aimed mainly to account for the prescriptive function of vision, while we left the descriptive one on the background. On the basis of this, someone might argue that, although our proposal is suitable to account for the prescriptive function of vision, it is not sufficient to explain how action can have a double life in guiding our actions. The problem is to understand how the prescriptive and the descriptive functions of vision come together.

To address this objection, we must first clarify how vision can have a descriptive function, then we must explain how such a descriptive function of vision comes together with the prescriptive one.

According to a widespread opinion, indeed, visual perception represents the environment by conveying informational contents that admit to be true or false (classically: Armstrong, 1988; Dennett, 1996; Dretske, 1969; Heil, 1983; McDowell, 1996; Peacocke, 1983; Searle, 1983, for a critical appraisal see Crane, 2009) According to this view, the structure of the content of our visual representations resembles that of our judgments, inasmuch as they have a common propositional format. Now, it is precisely by conceiving the content of vision as propositionally structured that we understand how it can perform a descriptive function. Indeed, it seems reasonable that when one observes an apple on the table, her visual system provides propositional information about the action-related properties of the apple that are relevant for the execution of compatible actions: grasping, biting, peeling (for a detailed analysis of this claim see Nanay, 2011, 2013) Accordingly, if our analysis is correct, we can maintain
that visual perception describes the objects of the environment as having action properties and that such description can be true or false. In other words, it makes sense to say that vision describes visual targets as having possibilities of action since it conveys contents about such properties in a propositional format.

So far, we have resumed what can be considered a straightforward way of thinking about the descriptive function of vision-in-action in terms of propositionally structured representations. Now, in order to answer to the first objection, we have to provide an account concerning how such a descriptive function comes together with the other function of vision in action, namely, the prescriptive one (§§ 1, 3). In brief, vision exerts a descriptive function inasmuch as it provides a propositional representation of the properties of the environment suitable for action, and it exerts a prescriptive function inasmuch as the action concept that constitutes such a propositional representation is realized by the same motor representation that constitutes our intentions to act, as stated by VMT (§ 3.2). But let us proceed slowly on this.

It should be noted that, a propositional content about an action-related property of a visual target is, *ipso facto*, about the action to which such a property is related. For example, the (visual) propositional representation of an apple as endowed with the property of being graspable in a certain way, is *ipso facto* about the way in which that apple can be grasped, namely, the very action of grasping that apple. Importantly, in order for a propositional structure to be about an action, it must contain the relevant *action concept* among its constituents. Hence, in order for a propositional structure to be about the graspability of an apple, it must be constituted by the suitable representation of the concept of “grasping”.

Importantly, as noted before (§ 3), action concepts can be of two types (Pacherie, 2011). An action concept is *non-executable* if its representation does not concern the motor acts that allow the execution of the relative action, while an action concept is *executable* if its representation coincides with the motor representation that determines the relative action (see originally Pacherie, 2011, and subsequently Ferretti and Zipoli Caiani, 2018, 2021; Ferretti 2020). Things being so, it is reasonable to assume that, at least in certain cases, the propositional representation of a visual target as endowed with action properties could be constituted by an executable action concept, i.e., by the same motor representation that determines the motor plan involved in the execution of the related action. It is precisely in such cases that vision has a double life. Notably, the visual representation that guides our actions exerts a descriptive function as it conveys a propositional content regarding action targets in the environment, while it exerts a prescriptive function as this content is constituted by an executable action concept, i.e., by the same motor representation which underlies our intentions to act, as stated by VMT. Things being so, VMT explains how the descriptive and the prescriptive functions of vision come together.
However, someone might still argue quite correctly that, although the empirical implications of VMT are confirmed by the experimental literature (§3.1), such evidence does not imply that VMT is true. Indeed, empirical evidence could be interpreted in a different way, that is, without entailing that vision and action intentions share the same motor representations in the guidance of actions, as stated by VMT.

Particularly, the abovementioned evidence of a recruitment of the parietal and premotor areas in the visual process is compatible with the hypothesis that visual information, before reaching the motor system, is influenced by already available semantic information. Indeed, it is nowadays accepted that the interaction between vision and action is not based on a segregated stream, but rather, it is an integrated process subserved by a network spanning the dorsal and ventral flow of visual processing (for a review, see Zipoli Caiani & Ferretti, 2017). Notably, the influence of the agent’s semantic competences on the visual perception of action targets and the related action guidance is not confined to higher planning levels, but rather extends to the low-level stages of sensorimotor computation. This means that the subset of visual processes that are not shaped by the recognition of semantic cues seems to be almost irrelevant for action guidance (Zipoli Caiani & Ferretti, 2017, Zipoli Caiani, 2018). Accordingly, being the interaction between the visual system and the motor system mediated by the semantic system, a further clarification is needed about how our visual representations can share the same motor representations of the action intentions that cause the execution of our actions, as stated by VMT.

In order to address this issue, first of all, it must be stated that our argument has an abductive form, so that although the evidence concerning the recruitment of the parietal and premotor system in the processing of the visual stimulus admits alternative, but less effective accounts, VMT is the best available explanation of those evidences.

Importantly, it is possible to argue that VMT fits adequately with the evidence relating to the integration of semantic and visual information (Zipoli Caiani & Ferretti, 2017). It should be noted, indeed, that it is nowadays widely accepted that the sensorimotor information plays a constitutive function in the formation of some relevant classes of semantic representations (e.g., Barsalou, 2008; Desai et al., 2010, 2015; Gianelli & Dalla Volta, 2015; Gijsels et al., 2018; Glenberg & Kaschak, 2002; Jirak et al., 2010). In particular, as already noted (§2.3), motor representations are relevant in the processing of executable action concepts (Brozzo, 2017; Gallese & Lakoff, 2005; Mylopoulos & Pacherie, 2016; Pacherie, 2011, for a discussion see Ferretti and Zipoli Caiani, 2018). In other words, on the basis of the recent literature, it can be said that, with regard to the conceptualization of executable actions, the information involved is of the same format as that of motor representations. Therefore, though the visual representations concerning an action target are semantically processed in terms of (executable) action concepts, it makes sense to hypothesize an integration between the
semantic information concerning executable action concepts and the visual information concerning action possibilities by means of motor representations. Indeed, given that the concepts concerning executable action are provided in a motor format, and given that visual information about the action properties of a target is semantically processed in terms of (executable) action concepts, then there are reasons to believe that visual representations are constituted by the same motor representations that guide our actions, as stated by VMT. Differently, if we assume that the visual representations that guide our actions are not constituted by motor representations, we have the problem of explaining how the visual representation of an action property can be relevant for the planning and execution of actions in which motor representations play a crucial role (Burnston, 2017; Butterfill & Sinigaglia, 2014; Ferretti & Zipoli Caiani, 2018; Mylopoulos & Pacherie, 2016; Shepherd, 2017b). The close integration of motor representations in visual information allows us to solve this problem. For this reason VMT is to be abductively preferred to alternative hypotheses.

Summing up, our framework explains how vision and intentions interlock in a way that also takes into account how visual states interlock with motor representations, which in turn interlock with intentions through executable action concepts.

5. Conclusion
In this essay we provided arguments to maintain that the content of vision is tied both to passively describe the environment as offering possibilities of action, as well as to actively prescribe what motor acts complement with that possibilities. In doing this, we have shown how vision-for-action has a double face.
References


