Nadiya El-Sourani

Neural Correlates of Exploiting Contextual Information During Action Observation: Evidence from Contextual Objects

- 2019 -

Neural Correlates of Exploiting Contextual Information During Action Observation: Evidence from Contextual Objects

Dissertation
zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)
im Fach Psychologie

eingereicht im Fachbereich Psychologie und Sportwissenschaft
der Westfälischen Wilhelms-Universität Münster

von
Nadiya El-Sourani
geboren am 10.08.1988 in Leer (Ostfriesland)

Dekan: Prof. Dr. Guido Hertel

Erste Gutachterin: Prof. Dr. Ricarda I. Schubotz

Zweite Gutachterin: Prof. Dr. Pienie Zwitserlood

Datum der mündlichen Prüfung: ________________________

Datum der Promotion: ________________________

Acknowledgements

First and foremost, I would like to thank Professor Ricarda I. Schubotz for giving me the opportunity to work in her lab and hence to pursue my Ph.D. I am grateful for her inspirational passion for science, her knowledge and her insights, as well as for her support and the trust she put in me and my work throughout the years. She always encouraged me to follow new ideas and to scientifically grow. I am also very thankful to my second supervisor Professor Pienie Zwitzerlood for her interest in my work, her encouragement and her cheerful manner. Many thanks go to Moritz Wurm, for inspiring me scientifically on all levels. I am profoundly grateful for his co-supervision, as he always encouraged me especially when things got tricky. I am truly inspired by his passion for science, his rigorous scientific integrity, and his always supportive and kind manners. I would also like to thank Professor Gereon R. Fink for his reliable and profound supervision and for always taking the time to discuss results and new ideas.

Furthermore, I want to express my gratitude to all former and current members of the Biological Psychology group (Bipsys) at the University of Münster, and especially to Ima Trempler. Ima has not only been a very dear colleague, supporting me in all work-related issues, but she has also become a true friend. Over the past couple of years, we stuck together through thick and thin: scientifically and emotionally. Ima, you were supportive whenever it was needed, you shared in my ups and downs and always lent an ear — even to my silliest thoughts. With you, even the (seemingly) worst things ended up as great experiences. I could not have asked for a better friend throughout this challenging time of ours. I also owe my gratitude to Christiane Ahlheim, who took Ima and me under her wing when we started working together. Thank you so much for your great sense of humour, your affectionate teasing, and your constant support and encouragement. I truly miss working with you. I would also like to thank Mari Hrkać, Klara Hagelweide, Daniel Kluger, Laura Quante, Jennifer Pomp, Nina Heins, Marlen Roehe, Lena Schliephake and Amelie Huebner for their company, offering fruitful discussions and help whenever needed; thanks to Monika Mertens and Jutta Linke, who helped me carry out my studies and organize my work environment, and provided encouraging chats whenever I needed them.

Finally, I want to thank my family. I thank my two older brothers, Nail and Nader, who always loved and supported me, and who, at the same time, knew how to tease me. In this way, you prepared me perfectly for working in science ☺. Moreover, I am beyond grateful for the love, trust and support of my parents. Papa, your true passion for science has deeply impacted and inspired me and my thinking. Thank you for constantly pushing me and showing me that one can achieve anything through consistent and hard work. Mama, thank you for being the best role model I could have wished for. There is truly not one other person in the world who has had my back as you have. Words cannot express the gratitude and admiration I feel for you! I also want to thank my grandparents for their support and love through this whole period. *Oma und Opa, Danke für eure Liebe, euer stetig offenes Ohr und den Mut den ihr mir zugesprochen habt, wann immer ich es gebraucht habe.* Last but not least, I want to thank my fiancé Haris Orucevic, who put up with me the last couple of years. No matter what, he always supported me and my work as fully as possible. He was understanding when I needed him to be and encouraging when I felt like things might not work out. Thank you for your imperturbable trust in me and my abilities, and of course for constantly reminding me of that. I am beyond grateful for your unconditional love, even at times when I was only able to display the worst version of myself. I could not wish for a better lifelong companion and best friend.

Without every one of you, I would not be at this place. Your support encouraged me to turn this challenging period into a great and ultimately successful experience. Thank you.

Table of Contents

Summary

Proper social interaction requires that we understand what people around us are doing and can infer their intentions and goals. Perceiving the actions of others is thus a prerequisite for action understanding. However, it remains unclear what information is de facto exploited by action observers to make accurate predictions concerning other people's actions when confronted by the large number of possible action options.

During action perception, information about manipulated objects and the movements of corresponding body parts (i.e., action kinematics) is exploited. However, this primary action information is often not perceived under perfect visual conditions. In addition, it is not sufficient for long-term goal inferences. Actions are naturally embedded in a contextual situation, i.e., they take place within particular settings and are performed by intentional actors. As such, contextual cues can help to fill in the gaps (when needed) and allow higher-level meaning to be inferred. Thus, to enable a full understanding of an action, it has been suggested that this *secondary* or *contextual information* is integrated as part of the perception of actions.

Building on these findings, the present work aimed at investigating whether and how *contextual objects* (COs), which are part of an action scene yet not integrated into the action itself, have an impact on action observation on a neural level. I aimed to answer questions concerning which (CO) information is read by the observer of an action and which factors influence the selection of the information that the observer notices. This was tackled by modulating two factors of the CO: the semantic relation to the observed action (*goal affinity* (GA)) and the spatial relation to the observed action (*location ergonomics* (LE)). Behavioural studies and two functional magnetic resonance imaging (fMRI) experiments were conducted to test 1) whether CO information is read by action observers and influenced by its semantic and spatial relation to the observed action; 2) whether, depending on the CO's location, participants establish their own action space by perceiving COs within an observed action scene, not only

with reference to the action but also with reference to themselves; and 3) whether the processing of CO information is driven by a CO's informative value to the observed action and its predicted (higher-level) outcome.

Study 1 investigated the neural mechanisms underlying the integration of COs in semantic and spatial relation to observed actions. Participants observed action videos in which COs and observed actions either possessed or lacked common overarching goals (*goal affinity*), and in which their location relative to the actor varied (*location ergonomics*). High goal affinity increased bilateral activity in the occipitotemporal cortex (OTC) and the intraparietal sulcus (IPS), but also in the precuneus and middle frontal gyri, implying that the semantic relation between COs and actions is considered during action observation and triggers processes beyond those usually involved in action observation. Moreover, COs with high goal affinity located close to the actor's dominant hand additionally engaged the bilateral inferior frontal gyrus (IFG), corroborating the view that the IFG is critically involved in the integration of action steps in pursuit of a common overarching goal.

Study 2 investigated whether, depending on their location, COs can trigger an observer-referenced processing. I reasoned that depending on the location a CO, action observers might also establish their own (virtual) action space by perceiving the CO not only regarding the observed action but also with reference to themselves. Participants were presented with the action videos from Study 1 that contained COs at varying locations on a table. A behavioural study assessing reachability scores of the COs from an observing perspective was conducted. Results suggest that, in addition to an actor space, participants created their own action space within the presented scene. These scores were subsequently used to model the BOLD response in the fMRI data set from Study 1. The fMRI analysis revealed significant activations in the premotor, cingulate, inferior parietal, and insular cortices in addition to the left IFG, parametrically increasing with reachability of the presented CO for the observer. These findings suggest that proximal located COs furnish action observers with their own action possibilities, and hence, even just observing an action from 3pp can trigger the involvement of an observer-referenced perspective.

Based on the findings of Studies 1 and 2, I reasoned that the processing of COs might be driven by a particular CO's informative impact on the observed action, rather than by congruency per se. In contrast to Study 1, Study 3 thus investigated whether *incongruent* COs, which match neither the current action nor the corresponding room category, exert a strong impact on action observation. I hypothesized that both highly congruent COs and incongruent COs provide rich information regarding potential upcoming action steps. Specifically, I tested the hypothesis that the IFG subserves this selection and integration process. As hypothesized, the processing of both highly congruent and incongruent COs engaged the OTC and the IPS. Moreover, the activation of all IFG compartments significantly increased for both CO levels, whereas the engagement of Brodmann Area (BA) 47 was significantly stronger for incongruent COs.

Together, the experimental data support the view that CO information is selectively used by action observers in order to construe the higher-level meaning of an action and thus predict an action's unfolding. This selection seems to be based on the informative impact of the CO on the observed action.

List of Original Publications

This thesis is based on the following original research articles:

Study 1.	El-Sourani, N., Wurm, M. F., Trempler, I., Fink, G.R., & Schubotz, R.I. (2018). Making sense of objects lying around: How contextual objects shape brain activity during action observation. *Neuroimage*, 167, 429–437.
Study 2.	El-Sourani, N., Trempler, I., Wurm, M. F., Fink, G.R., & Schubotz, R.I. (2019). Is it mine or is it yours? Proximal unused objects invite third-person perspective action observers to create their own action space. *Manuscript under review (Neuroimage).*
Study 3.	El-Sourani, N., Trempler, I., Wurm, M. F., Fink, G.R., & Schubotz, R.I. (2019). Predictive impact of contextual objects during action observation: evidence from fMRI. *Manuscript submitted to Journal (Journal of Cognitive Neuroscience).*

1. Theoretical and Empirical Background

From the beginning of our lives, we are constantly observing what other people are doing and thus trying to understand what their aims are. This ability is vital in order to navigate through our social world and to properly react and interact with one another. Thus, the capability to derive intentions, to capture action goals and to predict the future steps of an unfolding action is highly important for decent social interaction. Consequently, action observation and understanding has become an intensively investigated and discussed research field in human cognitive (neuro)science (cf. Overwalle & Beatens, 2009; Caspers, Zilles, Laird, & Eickhoff, 2010).

An everyday action is an extremely complex stimulus carrying many different types and amounts of information. It remains unsettled how we understand, simply by observation, the actions of others. When we observe an action, we are seldom presented solely with a single action *step*, consisting of a manipulation movement and a manipulated object (*core information*), performed in isolation. Instead, we are exposed to countless cues that may impact the perception of an action (*contextual information*) and especially its outcome, i.e., the overarching action goal. To give some examples, an action is usually embedded in a contextual scene that may include room information, actor information and additional object information, as well as spatial relations between objects and agents. Due to this contextual information, seeing somebody cracking an egg can have different implications. For example, if the manipulation is performed within a bakery setting (or at least in the presence of a baking bowl), one could infer that the person is trying to bake a cake. However, when performed in a typical kitchen and in the presence of a frying pan, one might instead expect the person to be preparing scrambled eggs. If the observed action scene additionally contained another actor who was buttering a bread, one might further expect that breakfast was going to be prepared.

In this way, action recognition involves not only the reading of body movements and of the objects involved, but also the integration of contextual information, i.e., the environment in which an action takes place (cf. Malcolm, Groen, & Baker, 2016). The process of inferring superordinate goals and generating predictions about forthcoming action steps should benefit especially from screening the action's immediate environment, and in particular, additional objects that are located within reach of an agent, and could thus be used in action steps subsequent to the observed action. In fact, objects themselves provide action observers with their functional use — that is, not only in terms of *how* they must be used (i.e., kinematics) but in terms of their function (i.e., intention) (e.g., a credit card is *for* paying) — thereby providing additional meaningful information (Bach, Nicholson, & Hudson, 2014; Nicholson, Roser, & Bach, 2017). As such, objects located close to oneself (i.e., the observer) can also provide meaningful information during action observation. Our social world comprises situations which involve not only the understanding of another person's intention to act, but at the same time, one's own ability and opportunity to act and interact with others and with the environment. This ability is necessary to adequately select actions when facing cooperation or competition.

In this thesis, I will focus on the impact of contextual information on action understanding in terms of such additional objects, referred to as *contextual objects.* Contextual objects are part of an observed action scene, i.e., part of its contextual setting, but never involved in the observed action itself. I investigate whether and under which circumstances this type of contextual information is exploited to constrain an observer's expectations concerning an observed action and its outcome, thereby potentially subserving goal anticipation. More specifically, I investigate the neural mechanisms underlying these processes.

In the following sections, I will give a detailed description of action as a stimulus, including different hierarchies of an action and its outcome, as well as the differentiation of core (i.e., primary) and contextual (i.e., secondary) action information types. I will then briefly give an overview of the existing theories of action understanding, followed by a summary of current findings on the impact of contextual information on object recognition and action observation. Finally, the relevant neural correlates underlying these issues will be discussed.

1.1 Action as a stimulus

In order to investigate how we understand what others aim to do by simple observation, action observation paradigms, which treat action as a stimulus, are invaluable. Since actions can be understood on different levels of abstraction, it is crucial to review the literature with regard to the concepts and definitions the term *action* can entail.

An *action* is by definition a change in the environment induced by an (intentional) actor (Zacks & Tversky, 2001). Accordingly, actions are goal-directed: that is, they are subject to a specific purpose (Csibra & Gergely, 2007). Noppeney (2008) differentiates actions into i) hand actions and ii) whole-body movements: while whole-body movements are i*ntransitive* in nature (e.g., running), hand actions can be either intransitive (e.g., waving somebody goodbye or pointing to something) or *transitive*, i.e., involving objects. The majority of hand actions, however, are transitive and thus depict object manipulations.[1] Depending on the definition of the beginning and end of an action, actions can be temporally structured on various scales: for example, “making coffee” can be regarded as a single action. When considered on a more fine-grained scale, however, it also can be considered as a sequence of action steps: that is, taking the coffee filter, putting it into a plastic filter, filling the filter with coffee powder, and so on (cf. Hrkać, Wurm, & Schubotz, 2014). On the other hand, when considered on a larger temporal scale, an action may be part of another overarching action, such as “preparing breakfast” (Kurby & Zacks, 2008). Like using a camera zoom on an object consisting of various objects we are

---

[1] In this thesis, transitive actions are the main focus. Thus, when using the term *action*, I am referring to an object manipulation.

posed with the problem of a wide action gamut in terms of (non-atomic) definition and inherent understanding. It has been suggested that this convoluted organization of an action is also reflected in its neural representation. In order to properly investigate action observation and understanding, it is important to differentiate between the types of information that can relate to an action (as a stimulus), the hierarchical subdivisions of an action and the different levels of its outcome, i.e., different levels of action goals. I will account for these issues in the following sub-sections.

1.1.1 Core vs contextual action information

When we observe an action (e.g., squeezing a lemon), we are typically confronted with a *manipulation movement* (e.g., squeezing) and the corresponding *manipulated objects* (e.g., the lemon and the squeezer). While the visual analysis of the manipulation movement (or action kinematics) incorporates its movement trajectories and the posture of the hands, the visual analysis of an object includes its pragmatic information (shape and size), as well as its semantic information, i.e., its functional use (or affordance, see below) (e.g., Bach et al., 2014; Nicholson et al., 2017; see Gibson, 1977, 1979). This information has to be processed in order for the observed action to be unambiguously recognized. It is termed *core information* (or *primary action information*). Thus, when participants are able to name the action (squeezing a lemon) after having observed the corresponding object manipulation, they have processed both object and manipulation information. Here, fMRI has proven an effective method for disentangling the signatures involved in the processing of these two core information types, which are simultaneously available (Grafton & Hamilton, 2007; Schubotz & von Cramon, 2009; Wurm, von Cramon, & Schubotz, 2012).

For example, studies have shown that the posterior superior temporal sulcus (pSTS) is sensitive to biological motion (Grossman et al., 2000), and that the lateral occipital complex

(LOC) — including the fusiform gyrus — is selective for processing visual object information (Grill-Spector, Kourtzi, & Kanwisher, 2001). However, action observation engages a larger fronto-parietal network and is not limited to areas processing these information sources. The underlying neural correlates will be thoroughly examined in Section 1.4.

Studies have shown that an action observer processes both sources of information even when the object information proves useless or even hinders the recognition of the performed manipulation, as is the case in pantomime with incompatible objects (Schubotz & von Cramon, 2009; Wurm et al., 2012).[2] This may be owing to the fact that objects are recognized earlier (< 200 ms) than manipulations, as manipulations only develop over time. In addition, object information itself implies potential actions (Buxbaum, Kyle, Tang, & Detre, 2006; Johnson-Frey, 2004). This is referred to as the concept of *affordance* (see Section 1.3), according to which the environment (including objects) implies information about possible actions (Gibson, 1977, 1979; Jeannerod, 1994; Fagg & Arbib, 1998). As such, objects provide an observer with a visuomotor priming effect when their perception facilitates an action that is congruent with their visual properties (Craighero, Fadiga, Rizzolatti, & Umilta, 1998; Craighero, Fadiga, Umilta, & Rizzolatti, 1996). Interestingly, this effect is dependent on the object's reachability (Costantini, Ambrosini, Sinigaglia, & Committeri, 2010; Costantini, Committeri, & Sinigaglia, 2011): that is, it only occurs when the object is within reach (cf. Maranesi, Bonini, & Fogassi, 2014) of either the actor (Costantini et al., 2010) or the observer (Costantini et al., 2011; Brozzoli, Gentile, Bergouignan, & Ehrsson, 2013; see also Creem-Regehr, Gagnon, Geuss, & Stefanucci, 2013). As a matter of fact, recent monkey (Ishida, Nakajima, Inase, & Murata, 2010) and human (Brozzoli et al., 2013; Brozzoli, Ehrsson, & Farnè, 2014) studies have shown that the same neuronal populations in the parietal and ventral premotor cortices respond to both

---

[2] Pantomime with incompatible objects depicts specific action kinematics (e.g., cracking an egg) that are performed on objects that are incompatible with regard to that action (e.g., a lemon and a squeezer).

objects close to one's own hand and objects close to another person's hand. Object affordances are thus *space* dependent.[3]

On the other hand, it has been found that observing action kinematics can help us understand and predict actions of others (Hamilton, Joyce, Flanagan, Frith, & Wolpert, 2007; Aglioti, Cesari, Romani, & Urgesi, 2008; Gowen, 2012). This is especially the case for complex social stimuli, i.e., real-life situations. Actions might be performed within cooperative or competitive situations (Manera, Becchio, Cavallo, Sartori, & Castiello, 2011); they might include multiple objects and/or imply different actions which can be performed with them (Schubotz, Wurm, Wittmann, & von Cramon, 2014). For example, when observing someone sitting in front of an apple and a knife, two actions could be expected: peeling or cutting the apple. It is not until the manipulation movement begins (cutting) that one can precisely name the observed action (cutting an apple).

Hence, if core information is available an observer can identify the immediate consequence of an object manipulation, thus labelling the action (e.g., squeezing a lemon). Moreover, one might even be able to infer possible goals of that action based on prior experience (e.g., preparing lemonade). However, it is not until the contextual setting of that observed action is taken into account (e.g., a kitchen, a salad bowl next to the action site) that a prediction can be made about the ongoing action, which is more likely than others (e.g., preparing a salad). Hence, information that may not actually be relevant for the recognition of an ongoing action step is still often crucial for higher-level processes like the inference of goals and intentions, and thus the prediction of forthcoming action steps. This information type is called *contextual information* (or *secondary action information*). *Context* per se has been

---

[3] The space around an individual can be divided into three areas: *personal space, peripersonal space* (PPS) and *extrapersonal space.* Personal space covers the surface of our body, whereas PPS or "action space" is the space surrounding our body that is reachable with our arms and hands. Finally, extrapersonal space is defined as the space that is unreachable with simple arm movements (di Pellegrino & Làdavas, 2015; Rizzolatti, Fadiga, Fogassi, & Gallese, 1997).

defined as everything within a given situation that shows or impacts the meaning of a target stimulus (Kokinov, 1997), or “the sensory/behavioural/cognitive milieu that influences the way each sensory feature is perceived” (Albright & Stoner, 2002, p. 340). Contextual information may be broadly subdivided into two classes: i) *person-related* contextual information, and ii) *scene-/situation-related* contextual information. While the former is allocated to the performing agent of an action and includes information on mental or social states, the latter is, as suggested by the name, derived from the scene in which the observed action takes place. This might include prototypical social or cultural events (e.g., a wedding) but also more fine-grained scene aspects, such as the room in which an action takes place (Wurm & Schubotz, 2012). It is worth noting that scenes not only contain specific objects or even object-categories, but also several non-moveable elements with fixed spatial locations, such as floors, ceilings, trees, etc., which contribute to the gist of a scene (e.g., a kitchen vs a forest) (cf. Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Malcolm et al., 2016). As humans tend to process information with regard to a global understanding (Fiske & Taylor, 1991) — that is, there is a natural human tendency to draw together several pieces of information to construct higher-level meaning in context (Frith & Happe, 1994)[4] — the scene in which an action takes place should possess information relevant for action understanding.

Contextual information can impact action perception on different levels. Hence, before reviewing important findings regarding the impact of contextual information on action observation, it is necessary to consider the different hierarchies of an action and the different levels of its outcome, i.e., different levels of action goals.

---

[4] This tendency is referred to as the concept of *Central Coherence* (Frith & Happe, 1994). Central Coherence includes two aspects: *global perception* and *contextual sensitivity*. The former is used in order “to see the forest for the trees”, whereas the latter refers to the ability “to use the forest in order to quickly recognize the trees as trees” (Vermeulen, 2015). Whereas the former might be associated with using context to infer overarching action goals, the latter might be associated with contextual facilitation during object or action identification.

1.1.2 Action hierarchies

In this section, I will consider the different hierarchies of an action. Regarding the hierarchy of an action, different models have been proposed (Byrne & Russon, 1998; Csibra, 2007; Grafton & Hamilton, 2007; Hamilton & Grafton, 2006, 2007; Jeannerod, 1994; Kilner, Friston, & Frith, 2007; Kilner, 2011; Wolpert, 1997; Wolpert, Doya, & Kawato, 2003). Grafton and Hamilton (2007) have suggested three levels in the hierarchy of an action: the *muscle level*, the *kinematic level* and the *goal level*. At the bottom of the hierarchy is the muscle level, representing the activity of the muscles needed to perform the action. On that level, motor neurons in the brainstem and the spinal cord are innervated, eventually inducing movement (Porter, 1985). The intermediate level, the kinematic level, represents the movement in space of an effector. Here, upper motor neurons in the premotor and motor cortices induce the movement kinematics of the effector (Dum & Strick, 2002; Graziano, 2006). Finally, at the top of the hierarchy is the goal level, representing the intention of the actor and the desired outcome of that action. Hamilton and Grafton (2007) have proposed the goal level to be coded by the parietal cortex, and more specifically, the anterior intraparietal sulcus.

Kilner et al.’s model (2007) resembles Hamilton and Grafton’s model (2007) in many aspects, but it makes a further distinction between goals and intentions, thereby proposing four levels in the hierarchy of an action. According to Kilner and colleagues (Kilner et al., 2007; Kilner, 2011), the goal level describes *short-term* goals (e.g., reaching to grasp a glass of water) that are necessary to achieve the long-term intention, while the intention level defines the *long-term* goal of an action (e.g., drinking *to quench thirst*). This further distinction is reasonable given that intentions are mental states of an agent, not directly visually accessible to an observer (Premack & Woodruff, 1987), and are further thought to precede an action (Zhu & Buckareff, 2006). On the other hand, a goal is the (anticipated) outcome of an action, and will therefore eventually be achieved by an observed action; goals are thus likely to be visually accessible to an observer, at least in the future. Kilner (2011) makes this distinction between goals and intentions on a conceptual level but not on a neural level. In contrast to Hamilton and Grafton (2007), Kilner (2011) suggests middle temporal areas and the anterior IFG (BA 47) to be responsible for coding goal inferences.

Taken together, the neural correlates of the lower levels of an action hierarchy have been well studied, whereas the neural correlates of higher levels (i.e., goals and intentions) seem not to have been conclusively ascertained yet. This might in part be owing to the fact that psychologists and neuroscientists have used the terms *goal* and *intention* interchangeably. Moreover, the term goal has been used ambiguously in the research literature, as it can refer to different sub-states of an action (or its outcome) (cf. Section 1.1). Therefore, I will consider goals and intentions more closely in the following section.

1.1.3 Goals and intentions

The term *goal* itself has already been found to be used in at least three different ways throughout the literature (Uithol, van Rooij, Bekkering, & Haselager, 2011). On the most basic level, a *goal* has been described as a target in space, thus depicting (e.g.,) pointing to (Wohlschlager & Bekkering, 2002), reaching for (Bekkering, Wohlschlager, & Gattis, 2000) or grasping an object (Hamilton & Grafton, 2006). On the intermediate level, the term *goal* has been used interchangeably with the term *action*, commonly brought down by early macaque literature on mirror neurons (e.g., Gallese & Goldman, 1998; Rizzolatti & Craighero, 2004). Finally, goals have been described as a desired outcome or state of an action, such as “a prepared breakfast” or “a full cup of tea” (cf. Csibra & Gegely, 2007). Apart from the wide range of definitions used to describe an action goal itself, there have also been ambiguities regarding the usage of the terms *intention* and *goal* in the action literature (Hamilton, 2009; Iacoboni et al., 2005). As briefly described in Section 1.1.2, the main difference between intentions and goals is that an intention is a mental state, not directly visually accessible (Premack & Woodruff, 1987), that precedes the execution of an action, while a goal refers to the forthcoming states in the external world and thus should be at least perceivable as a future state of an observed action. In fact, Iacoboni et al. (2005) have defined an intention as the 'why' of an action. However, when asking participants in studies about the 'why' of an observed action, they were found either to infer goals or to envision the mental state of the agent, which relate to two different cognitive processes (cf. Frith & Frith, 2005, 2006; Saxe, 2005). Thus, using the term *intention* may be disadvantageous due to its potential to confuse mental state and goal inferences. To give an example, on a basic level, the 'why' and thus the intention of the action *grasping* could have (e.g.,) two intentions: "grasping to drink" or "grasping to write". *Writing* or *drinking,* in turn, can themselves be described as actions, or at the very least, as the more distal action goals of the sub-goal *grasping* that may or may not become true as its future state. Similarly, there could be different intentions behind writing, e.g., "writing a summary" to "study for a test" or "writing a letter" to "stay in contact with someone/socially interact". To avoid ambiguity, in the present thesis when referring to the term *goal*, I mean a desired end state or outcome of an action. In the next section, I will give an overview of the three main neurocognitive theories on action understanding.

1.2 Theories on action understanding

There are three major neurocognitive accounts regarding the understanding of action and the involvement of the motor system: *direct matching*, *action reconstruction* and *predictive coding*. I will briefly describe each theory within this section, before turning to more specific neural correlates in Section 1.4.

1.2.1 Direct matching theory

After the discovery of mirror neurons in the macaque premotor area F5 (di Pellegrino, Fadiga, & Fogassi, 1992), and later in the parietal areas termed PFG and AIP (Fogassi et al., 2005; for a review, see Rizzolatti & Sinigaglia, 2010), the idea of shared representations of action and perception attracted much attention among neuroscientists. Mirror neurons in the macaque monkey were found to modulate their activity during the execution of an action as well as during the observation of the same (or similar) action performed by another individual. This *mirror neuron network* in the macaque monkey corresponds to a fronto-parietal network in the human brain (for a review, see Rizzolatti & Sinigaglia, 2010). Based on the findings relating to mirror neurons, a theory called *direct matching* was put forward that suggests that an observed action is directly mapped to a corresponding motor representation of that action in the observer's motor repertoire, i.e., the motor system of the action observer resonates with the observed action (Rizolatti, Fogassi, & Gallese, 2001). Through this simulation process (or "mirroring"), goals and intentions of the observed action can be inferred (Gallese & Goldman, 1998) and subsequently understood. In this sense, action understanding is *retrodictive* (Zentgraf et al., 2011), i.e., a visually perceived action is propagated bottom-up from the OTC and directly mapped to the motor system. Thereby, access to knowledge about action goals is enabled. In this sense, action mirroring provides a simulation device for goal understanding. Actions are understood only after the visuomotor transformation has been performed. Hence, the meaning of the contextual setting in which an action takes place is ignored.

In spite of its popularity, there have been theoretical concerns with the direct matching theory (e.g., Csibra, 2007; Hickok & Hauser, 2010; Jacob, 2008; Kilner et al., 2007; Kilner, 2011; Friston, Mattout, & Kilner, 2011). One important concern relates to the fact that an action goal can be achieved by different means (i.e., actions), while a specific action can have different goals, i.e., there is *no one-to-one mapping* (Kilner et al., 2007). For example, a cup can be grasped to drink or to clean (Iacoboni et al., 2005). Just by directly simulating the observed motor act in one's own motor system, one will not necessarily know what the goal of (e.g.,) this "grasping action" will be. Therefore, alternative theories have been proposed: namely, *action reconstruction* (Csibra, 2007) and *predictive coding* (Kilner et al., 2007; Kilner, 2011; Friston et al., 2011). These theories suggest goal inferences through emulative processes and take the contextual constraints of a given observed action into account.

Before turning to action reconstruction and predictive coding in detail, it is worth mentioning the *affordance matching hypothesis* (Bach et al., 2014). This is an approach to action understanding that incorporates the available data on direct matching but places it in a model of action understanding that emphasizes the role of object knowledge, which is thought to facilitate the prediction and interpretation of any observed action. Thereby, this model aims to account for several of the inconsistencies found with regard to the direct matching account (such as the one-to-one mapping problem). Objects and their functional use — as opposed to action kinematics — are viewed as the main driver of action understanding (Bach et al., 2014). Accordingly, an understanding of object-directed actions does not initially result from a matching of motor parameters, but is based on the identity of objects, as well as one's prior experience of their function (what an object is for) and use (how it is used). This proposition is especially interesting given that during action observation, objects are recognized earlier than action kinematics. Support for this theory comes from a recent neuroimaging study revealing overlapping brain areas for object and higher goal processing, but not movement-related activations (Nicholson et al., 2017). For a detailed description of this theory, see Bach et al. (2014).

1.2.2 Action reconstruction theory

According to the action reconstruction theory, also referred to as the *inferential model,* it is suggested that the involvement of the motor system during action perception does not subserve action goal inference (Csibra, 2007). Instead, action goals originate outside the motor system: for example, by propagating relevant information directly from the OTC to the prefrontal cortex. Hence, it is suggested that the observed action is reproduced through its interpretation, which is based on an initial visual analysis of that action. During this visual analysis, further information such as the context can be included to help select the most appropriate goal (cf. Iacoboni et al., 2005). Information is then passed down to the motor system, which in turn is thought to compute the most appropriate means to attain the specific goal (without performing it) through emulative reconstruction. In agreement with the direct matching account, action mirroring in the motor system implements a simulation function. However, the primary function of action mirroring is suggested to be predictive action monitoring rather than action understanding in terms of goals. In that sense, action understanding and goal inference are thought to precede action mirroring by the motor system, rather than following on from it (Csibra, 2007). Hence, action understanding is suggested to be predictive in nature, in that its goals are inferred in a top-down manner. This would allow for the anticipation of an ongoing action enabling action coordination with others, necessary for proper social interaction.

1.2.3 Predictive coding theory

According to the predictive coding theory (PCT; for a review, see Clark, 2013) — adapted from motor control theory (Sperry, 1950; Wolpert, Ghahramani, & Jordan, 1995) — actions are understood and predicted through the use of *generative internal models*. These generative models which comprise the expectation regarding a sensory input (unfolding of observed action) are probabilistic in nature and are emulators of the ongoing action, thus capable of estimating a state of the action before actual feedback from the observed action is available. It is suggested that generative forward models are learnt by action observers in the guise of optimal Bayesian observers (for detailed description, see Friston, 2010). According to the PCT, a generative model is used to predict the most probable action outcome (Friston et al., 2011; Kilner et al., 2007) via top-down modulation from higher to lower levels of the hierarchy (see Section 1.1.2) until it reaches the primary sensory areas. On each level, the generative model is compared to subordinate levels. If the predicted sensory info matches the actual sensory input, a goal can be inferred and the predictable sensory input (of the cortical motor system) can be suppressed (Keysers & Perrett, 2004; Schiffer, Ahlheim, Ulrichs, & Schubotz, 2013; Summerfield & Egner, 2009). On the other hand, a discrepancy between the predicted and actual sensory feedback creates a *prediction error*, which is processed upwards through the hierarchy in order to adjust the currently operating model, i.e., the top-down prediction (Clark, 2013). An adjustment at the goal prediction level is hence necessary and the next most likely goal is instead predicted. By minimizing the prediction error at all levels of the hierarchy, the most likely cause of an action can be inferred. Importantly, contextual information is integrated into the model (Kilner et al., 2007), thereby addressing the one-to-one mapping problem: the most likely intention of an action can be inferred even when the action is associated with multiple intentions, such as “grasping the cup to clean” or “grasping the cup to drink” (cf. Iacoboni et al., 2005).

It is worth mentioning that forward models might also be run offline (*internal emulation*). Hence, they could contribute to a range of cognitive tasks such as motor imaginary or action anticipation of others (Grush, 2004; Wolpert, Doya, & Kawato, 2003). It has been shown that the premotor cortex (PMC) is involved in such action prediction processes, making this region a likely candidate for the generation of forward models (Schubotz, Sakreida, Tittgemeyer, & von Cramon, 2004; Schubotz & von Cramon, 2002, 2003, 2004).

Taken together, the action reconstruction theory and the predictive coding theory have several parallels (Zentgraf, Munzert, Bischoff, & Newman-Norlund, 2011). Both use prediction as the main driving force during action observation and understanding, both emphasize the importance of empirical Bayes during goal inference, and both incorporate contextual information into their model (Csibra, 2007; Kilner et al., 2007; Kilner & Frith, 2008; Kilner, 2011). Indeed, in order to be able to infer the higher-level goals of an ongoing, observed action, it would be beneficial to consider all relevant perceivable information helping the observer to understand an observed action: that is, to take context into account. Therefore, I will review the relevant findings on the impact of contextual information during the observation of actions in the next section.

1.3 The impact of contextual information on action perception and understanding

The impact of context during action observation has been investigated within and across different domains (e.g., language), revealing facilitating and impeding effects depending on its compatibility with the target stimulus. Following on from this, the contextual setting in which an action takes place can provide relevant information that can help an observer to identify an observed action when core information is not fully available (e.g., for visibility reasons), or to infer higher-level meaning, i.e., overarching action goals. As such, it is important to emphasize contextual scenes as environments providing possibilities for action (cf. Malcolm et al., 2016),[5]

---

[5] Scenes are very complex stimuli and can be described at several levels ranging from basic stimulus features (e.g., edges and colours) to high-level characteristics (e.g., object identities). At any given moment, different visual aspects of an environment (i.e., scene) will be prioritized based on the current goal of an observer. Moreover, different properties of a scene can be processed to complete a single goal, while a single property can also be used to facilitate many different goals, depending on several factors such as the goal in question and the prior knowledge

as was initially provided by the *theory of affordances* put forward by Gibson (1977, 1979). According to this theory, affordances are not about mere physical properties, they rather incarnate the action opportunities that the environment has to offer to an individual who is able to perceive and use them.[6]

Taken together, contextual scenes should exert an impact on the recognition of objects and actions performed within them, whereas objects themselves — due to their functional use (e.g., a credit card is for paying) (Bach et al., 2014; Nicholson et al., 2017) — can serve as both: they can be part of the observed action itself, i.e., part of the primary action information, or a part of the contextual scene (e.g., a kitchen consists of a stove, a fridge and a sink). In the below section, I will summarize all relevant findings with regard to the impact contextual (scene) information can have on action perception.

1.3.1 Contextual information and object recognition

The influence of contextual information on object identification and recognition has been extensively investigated using a wide range of methods, including eye tracking, fMRI and electroencephalography (EEG) (Biedermann, Glass, & Stacy, 1973; Biedermann, Mezzanotte, & Rabinowitz, 1982; Boyce, Pollatsek, & Rayner, 1989; Palmer, 1975; Bar & Ulmann 1996; Bar, 2004; Davenport & Potter, 2004; Hayes, Nadel, & Ryan, 2007; Zimmermann, Schnier, & Lappe, 2010; Barenholtz, 2014; Wokke, Knot, Fouad, & Ridderinkhof, 2016). There is a general agreement across studies that objects appearing in a familiar or congruent scene are i) detected more accurately and ii) processed more quickly, compared to objects presented within an incongruent scene (cf. Bar, 2004; Oliva & Torralba, 2007). For example, Palmer (1975)

---

of the observer. Hence, scene processing is not entirely stimulus driven but is dependent on matching a percept to a prior experience. In this sense, scene understanding entails representing information about the properties and arrangement of the world in order to facilitate the ongoing needs of the observer (for a detailed review, see Malcolm et al., 2016).

[6] The theory of affordance is by definition the power of the environment to furnish the viewer with possible actions (Gibson, 1979).

found the accuracy of object categorization of a target object (e.g., loaf of bread) to be influenced by its presentation within congruent (e.g., kitchen counter) or incongruent (e.g., bathroom) contextual scenes, thereby facilitating or impairing accuracy, respectively. In a similar vein, studies on categorization have revealed that presenting objects within a scene facilitates categorization, especially of superordinate-level categories (e.g., animals) (see Murphy & Wisniewski, 1989; Borghi, Caramelli & Setti, 2005). Davenport and Potter (2004) found that coherence between an object and its respective scene revealed a mutual influence on recognition, i.e., objects influenced scene recognition and vice versa. Bar and Uhlmann (1996) found that in the presence of a clearly identifiable object, an additional ambiguous object's recognition was facilitated when both objects were semantically related. A more recent study even revealed that objects (e.g., a fridge) relating to a specific room category (e.g., kitchen) were recognized faster when participants were seated in that specific category of room during the experiment (Wokke et al., 2016).

To explain the observed effects of context on object recognition, Biedermann et al. (1982) postulated five contextual factors influencing object recognition: *interposition*, *support*, *probability, position* and *size*. While the first two respectively describe physical factors characterising the background of an object and the fact that an object is physically supported rather than free floating, the other three factors are semantic in nature. *Probability* refers to the likelihood that certain objects will be present in a specific scene, *position* refers to the typical position of a given object in a given scene, and *size* relates to the typical relative size that an object usually has. The authors found that a violation on each level increased reaction times (RTs) and error rates. Based on this study and related findings on object recognition, Bar (2004) concluded that the context of a scene might be extracted early enough to affect our perception of an individual object in a top-down manner. He proposed a specific mechanism underlying the contextual facilitation of object identification: *a context frame*. Context frames are defined by Bar (2004) as “a prototypical representation of unique contexts (for example, a library), which guide the formation of specific instantiations in episodic scenes (for example, our library)” (p. 618). Low-frequency visual information is quickly extracted from the visual scene (~ 80 ms), leading to the activation of a context frame which in turn provides expectations that facilitate other scenes that can be represented by the same context (Bar, 2004). Accordingly, the processing and recognition of objects that are sufficiently characteristic of a specific context will be facilitated while atypical objects for that given context require additional scrutiny. Other studies have also found that the gist of a scene (i.e., a coarse version of a context frame) can be extracted based on ‘one glance’ (Joubert et al., 2007), but they argue that contextual processing does not appear to be faster than object processing. Rather, object and contextual information are extracted in parallel by observers, allowing for the reciprocal influence of both sources: processing from context to object categorization, but also from objects to context categorization.

As demonstrated, the impact of contextual information on object recognition has been extensively researched. In contrast, the effects of contextual information on action perception and goal inferences have been less extensively addressed in the literature. I will give a summary of these findings in the next section.

1.3.2 Contextual information, action perception and goal anticipation

As with objects, actions naturally occur within a related contextual setting. Actions are not perceived in isolation, but are rather embedded in context, with objects, actors and the relationships amongst them merging into a unifying and meaningful scene (cf. Amoruso et al., 2014; Amoruso, Couto, & Ibanez, 2011). Therefore, the contextual information of an observed action should impact the understanding of that action similarly to how contextual information impacts object recognition: enhanced action recognition in cases of congruent contextual information as compared to hampered action recognition in cases of incongruent contextual information.

The impact of contextual information on action perception has attracted less attention than its impact on object recognition. Relevant studies have found that participants process contextual information spontaneously, i.e., without task requirements. While participants need longer to recognize an action when it takes place in an incompatible as opposed to a compatible or neutral room (Wurm & Schubotz, 2012), action-compatible room information can help when actions are difficult to recognize, leading to increased recognition accuracy (Wurm & Schubotz, 2017). Wurm and colleagues (2017) showed that children at the age of four to eight years already exploit contextual information during action observation: recognition rates for object-free pantomime actions were enhanced when taking place in compatible as opposed to incompatible or neutral rooms. This effect interacted with action familiarity: context effects were stronger for unfamiliar than for familiar actions. This effect suggests that the context of an observed action becomes especially important when actions are unfamiliar, i.e., there is no or little perceptual motor experience with a specific action.

Even where contextual information derived from the observed action scene is not necessary for the recognition of an ongoing action (e.g., when both aspects of core information are fully available), it might still be crucial for higher-level inferences of goals and the prediction of forthcoming action steps (Malcolm et al., 2016). For example, Iacoboni et al. (2005) have investigated how a background scene providing a meaningful context impacted on the observation of a “grasping a cup” action. The authors compared the cortical activity induced by viewing the action against a plain background to either a “reach and grasp to drink” or a “reach and grasp to clean” contextual background. Putting the action into a meaningful context increased activity in the premotor and inferior frontal cortices. Thus, an increase in neural activity was observed when visual information regarding an action goal was presented as compared to when no such additional information was available (plain background). This finding clearly shows that i) during action observation, context can provide meaningful information with regard to a goal, and ii) this additional information is indeed exploited by action observers, whereby ii) can serve as an indication of i).

Other evidence coming from neuroimaging studies investigating the effect of contextual information suggests the presence of interference effects for incompatible contextual information, rather than facilitation effects for compatible contextual information in action processing (Hrkać et al., 2014; Wurm & Schubotz, 2012). For example, an incompatibility between the room, the manipulated objects and the applied manipulation was accompanied by increased activity in brain regions associated with object and action processing (Wurm et al., 2012). When an action took place in an incompatible room, activity in the IFG[7] also increased (Wurm & Schubotz, 2012). Hrkać et al. (2014) found that actor information modulates brain activity during the observation of an action: they found a parametric increase in IFG activity when a single actor performed otherwise unrelated action steps, implying goal incoherence. Thus, contextual information in terms of actor information seems to crucially impact the interpretation of an observed action.

Finally, studies on corticospinal excitability suggest a top-down modulative role for contextual information during action observation (Amruso & Urgesi, 2016; Riach, Holmes, Franklin, & Wright, 2018). Providing actions with a congruent contextual setting facilitated corticospinal excitability (Riach et al., 2018), while an incongruent contextual setting did selectively decrease corticospinal excitability (Amuroso & Urgasi, 2016). These results indicate that motor resonance is not an entirely automatic process but can be modulated top-down via contextual information, underlining the predictive value that context can add to action observation.

---

[7] The IFG is a brain region associated with semantic integration; for a detailed description, see Section 1.4.3.

Taken together, these findings indicate that i) context is exploited by action observers, ii) context can not only provide but also activate meaningful information regarding the goal of an action, thereby providing access to corresponding motor representations, and iii) both congruent (Iacoboni et al., 2005) and incongruent (e.g., Wurm & Schubotz, 2012; Hrkać et al., 2014) contextual information can be meaningful with regard to goal inference: context can support the interpretation of actions by either constraining expectations in cases of high compatibility, or leading to interference for incompatibility. In the latter case, an extra effort would be needed to disentangle the meaning of a particular action scene.

As reviewed, contextual information can modulate action perception. Decoding an observed action hence involves a range of cognitive processes and the integration of these different sources of information into a common frame of reference. In the next section, I will consider in detail the underlying neural correlates of action perception with regard to the impact of contextual information on action understanding.

1.4 Neural correlates of action perception

The discovery of mirror neurons in the macaque brain (see Section 1.2.1) raised the question of whether a comparable system also exists in humans (e.g., Rizzolatti, Fogassi, & Gallese, 2001). Accordingly, action observation has become an intensively investigated and discussed research field in human cognitive neuroscience (for a review, see Caspers et al., 2010). It was suggested that the identified mirror neuron system corresponded to a fronto-parietal network in humans (Rizzolatti & Sinigaglia, 2010), activated when an individual performs an action as well as when the same action is observed in another individual. However, in the past few years, an extended network — termed the *action observation network* (AON) (see Cross, Kraemer, de Hamilton, Kelley, & Grafton, 2009) — has been identified that reaches far beyond the 'classical' mirror neuron areas within the ventral premotor (F5 in the macaque monkey) and inferior parietal cortices (PFG and AIP in the macaque monkey) (cf. Caspers et al., 2010; Van Overwalle & Beatens, 2009) and expands across both hemispheres, encompassing the posterior temporal, parietal and premotor areas, the IFG, the extrastriate visual areas and the fusiform gyrus (Caspers et al., 2010). Depending on the task at hand, specific brain areas are conjointly activated along with the AON, e.g., the dorsolateral prefrontal cortex (dlPFC) (Amuroso et al., 2018).

In the upcoming section, I will take a closer look at the specific brain areas engaged during action observation with regard to i) manipulable objects, ii) scene perception, and iii) exploiting contextual information en route to goals.

1.4.1 Neural correlates of manipulable objects

When visually perceiving an action, the initial requirement is the visual analysis of object and manipulation information. Visual information is distributed in a bottom-up manner from the primary visual cortex (V1) along two processing streams (Milner & Goodale, 1995; Ungerleider & Mishkin, 1982): *the ventral stream*, also termed the “what” or “semantic” stream; and *the dorsal stream*, also known as the “where” (Ungerleider & Mishkin, 1982) or “how” stream (Milner & Goodale, 1995). It has been proposed that the dorsal stream, which stretches from V1 to the parietal lobe, mediates spatial perception (“where”) and the sensorimotor transformation of visually guided action (“how”; see also Maranesi et al., 2014).[8] On the other hand, the ventral stream, which links V1 to the (inferior) temporal lobe via V2 and V4, has been associated with object recognition (Goodale & Milner, 1992; see also Maranesi et al. 2014), and on a broader level, with visual representation and knowledge (Buxbaum & Kalénine, 2010; Jeannerod, Arbib, Rizzolatti, & Sakata, 1995). A key region of the ventral

---

[8] A further division of the dorsal stream into the dorsodorsal and ventrodorsal stream has been proposed but is not described here, as it is beyond the scope of this thesis. For a detailed description see Rizolatti and Matelli (2003).

stream is the LOC, which consist of the lateral OTC and the fusiform cortex (Grill-Spector et al., 1999). These brain regions were found to be activated when seeing objects compared to scrambled objects (Grill-Spector et al., 2001). Other studies have suggested that the LOC represents semantic properties (Mahon et al., 2007; Park, Brady, Greene, & Oliva, 2011; Simons, Koutstaal, Prince, & Wagner, 2003). This is in line with the generally assumed function of the OTC: that is, the encoding of semantic object-and-action information (Bedny & Caramazza, 2011; Binder, Desai, Graves, & Conant, 2009; Noppeney, 2008). More recent studies propose that the lateral OTC also encodes action information on a more abstract level of representation that is object-, and orientation-independent (Wurm & Lingnau, 2015; Wurm, Ariani, Greenlee, & Lingnau, 2016).

Located within the OTC and anterior to the LOC, the *posterior middle temporal gyrus* (pMTG) has been found to be specific to the visual processing of tools and tool-specific motion patterns (Beauchamp et al., 2002; Beauchamp & Martin, 2007). It has therefore been named “the MTG-tool area” (Downing, Chan, Peelen, Dodds, & Kanwisher, 2006). Other studies suggest that the pMTG codes conceptual object properties (Buxbaum & Kalénine, 2010; Noppeney, 2008). Finally, a very crucial area for the perception of manipulable objects is the posterior parietal cortex (PPC) in the dorsal pathway. In particular, the *inferior parietal lobule* (IPL)[9] is engaged in the retrieval of specific ways of grasping tools (Chao & Martin, 2000) and manipulating objects (Kellenbach, Brett & Patterson, 2003), and might therefore be responsible for coding pragmatic object properties (Rumiati et al., 2004). Patients suffering from lesions in that area are often impaired in their ability to retrieve the correct manipulation for a given tool, whereas their ability to correctly name the tool stays intact (Rumiati, 2004; Buxbaum et al., 2005). The anterior part of the IPL (aIPL) is suggested to generate internal representations of actions and knowledge about actions (Creem-Regehr, 2009). In a similar vein, the left aIPL —

---

[9] The IPL consists of the supramarginal gyrus and the angular gyrus, and is divided from the superior parietal lobule by the intraparietal sulcus.

in combination with the ventral premotor cortex (PMv) and the pMTG— were found to positively co-vary with the number of object-implied action possibilities (Schubotz et al., 2014). This matches the fact that the IPL is anatomically connected to the PMv (Geyer et al., 2000; Schmahmann & Pandya, 2006).[10] Corroborating these findings, Wurm and Lingnau (2015) investigated the brain areas that decode actions (opening, closing) on *concrete, intermediate* and *abstract* levels of conceptualization. While they found the lateral OTC and IPL to decode actions at all levels, the PMv only decoded actions at the concrete level.

Taken together, each of these brain regions is suggested to crucially contribute to the processing of object manipulations (for a simplified visualisation of these brain areas, see Figure 1).

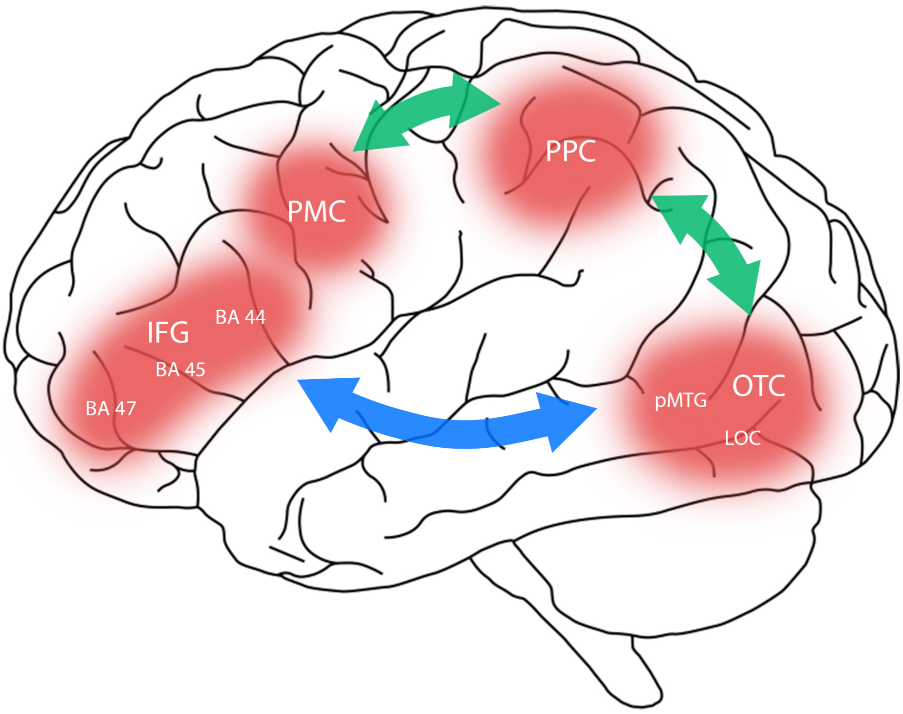


Figure 1. Brain regions underlying the processing of observed actions. Note, that this is a simplified visualisation: only regions relevant for this thesis are highlighted and their locations are schematized and simplified for clarity. The processing route via the dorsal stream is depicted in green, while the processing route via the ventral stream is depicted in blue.

---

[10] Note, that the PMv was found to be activated when perceiving manipulable objects but not for non-manipulable ones (Kellenbach, Brett, & Patterson, 2003).

1.4.2 Neural correlates of scene perception during action observation

Contextual information is normally represented in the form of a scene (cf. Malcolm et al., 2016). There are three scene-selective brain regions that are suggested to significantly contribute to scene processing in humans: the *occipital place area* (OPA) on the lateral occipital surface, the *parahippocampal place area* (PPA) on the ventral temporal surface, and the *retrosplenial complex* (RSC) on the medial parietal cortical surface (Epstein, 2014; Malcolm et al., 2016). Importantly, these regions show a peripheral visual field bias with relatively large population receptive fields, which makes them well positioned to capture summary information across large portions of the field of view (Silson, Chan, Reynolds, Kravitz, & Baker, 2015; Silson, Steel, & Baker, 2016; Silson, Groen, Kravitz & Baker, 2016). While the RSC is suggested to be sensitive to landmarks and is further associated with spatial memory and imagery (e.g., Vass & Epstein, 2017), especially in the context of navigation (Malcolm et al., 2016), the OPA codes local elements in a scene (Dilks, Julian, Paunov, & Kanwisher, 2013). Moreover, the OPA is involved in scene discrimination and categorization (Dilks et al., 2013; see also Malcolm et al., 2016). It has been suggested that the parahippocampal cortex is concerned with the local space information that is processed both when we perceive a given scene and when we perceive objects that are strongly suggestive of a particular environment (Harel, Kravitz & Baker, 2013). In a similar vein, it is important to stress the fact that a scene naturally comprises different sorts of information. As such, contextual information can include information about different entities such as actors or additional objects. Depending upon the different contextual elements of a particular scene, varying these aspects should engage additional respective brain areas, such as (for example) the OTC for additional object information.

1.4.3 Exploiting context en route to goals

As already indicated in Section 1.2, there is still an ongoing debate about the brain regions ascribed to the processing of action goals and intentions, which might in large part be owing to the diverging definitions and approaches employed to investigate goal understanding across studies.[11]

Initially, the inferring of goals has been ascribed to areas located within the dorsal stream. Previous studies focused on the IPS and the PMv due to their mirroring function (cf. Section 1.2). Recently, this was partly challenged by Wurm and Lingnau (2015), who argued for a cognitive account of action understanding (see also Marneweck & Vallence, 2015), whereby perceptual functions underlie action understanding. They emphasized the role of the lateral OTC and the IPL, as these areas seem to contain representations that are action specific and concurrently tolerate a wide range of perceptual variance, whereas the PMv only decoded actions at the concrete level and might therefore rather support specific action (outcome) anticipation (cf. Wurm et al., 2016). Regarding the IPL, there seems to be general agreement regarding its role in representing action outcomes, both in the sense of proximal end states or more distal long-term goals (Oosterhof, Tipper, & Downing, 2013; Rizzolatti, Fabbri-Destro, Rozzi, & Cattaneo, 2014), e.g., decorating a room (Leshinskaya & Caramazza, 2015).

A body of research has emphasized the role of the ventral pathway (Badre & D’Esposito, 2009, Binder et al., 2009) that links the pMTG and the anterior IFG (BA47) during goal inferences (Kilner, 2011; see also Nicholson et al., 2017). The idea is that a representation of a goal is generated through a process of semantic retrieval and selection, supported by the anterior IFG, resulting in the encoding of the representation of the most likely action that is required to

---

[11] Note that action understanding can occur at different levels of the action hierarchy (cf. Section 1.1.2 and 1.1.3). One could, for instance, distinguish between outcomes of single action steps that might relate to short-term goals (e.g., opening a bottle) and multistep actions (e.g., preparing breakfast) that might instead relate to higher-level overarching action goals.

achieve the most likely goal, encoded in the pMTG. Once this goal is estimated, a prediction of the sensory consequences of this action can be generated and the action's outcome can be anticipated by means of the dorsal pathway (see Figure 1; cf. Amoruso et al., 2013).

With regard to the inference of overarching long-term action goals, the incorporation of contextual factors during action observation is thought to play a crucial role (cf. Iacoboni et al., 2005; cf. Amoruso et al., 2013). The IFG is known for its central role in the retrieval and integration of action-relevant semantic information (Caspers et al., 2010; Kilner, 2011), and from a broader perspective, has a well-established role in effortful contextual integration in different cognitive domains, especially language (cf. Bookheimer, 2002; Liakakis, Nickel, & Seitz, 2011). Shan et al. (2003) have described the role of the IFG is "choosing, comparing, judging and retrieving" information. As such, it is a very likely candidate for integrating contextual action information en route to higher-level action goals.

The IFG is a brain structure which forms a large part of the ventrolateral prefrontal cortex (vlPFC). It consists of three different parts. Anterior to the ventral part of the precentral gyrus (BA 6) lies the posterior pars opercularis. This area is succeeded rostrally by the pars triangularis. Finally, ventrolateral to the pars triangularis lies the pars orbitalis, which occupies the most ventral part of the lateral frontal cortex and extends onto the orbital surface (Petrides, 2005). These areas roughly correspond to BAs 44, 45 and 47, respectively (Zilles & Amunts, 2010). The IFG belongs to the frontal component of the ventral pathway (see Section 1.4.1) (O'Reilly, 2010). The parcellation of the IFG, as well as various theories, suggest a hierarchical functional organization in the vlPFC (Amunts & Zilles, 2012; Badre & D'Esposito, 2009; Koechlin & Jubault, 2006).

Since 1861, when Paul Broca presented the case of aphemia (loss of speech due to damage to the left IFG), the IFG has been considered the centre of speech production. Consequently, it has been an area that has received special investigation with regard to language processing (cf. Bookheimer, 2002; Liakakis et al., 2011). In her review, Bookheimer (2002) summarizes findings on the function of the IFG pertaining to language processing as follows: the most posterior regions (BA 44/6) are involved in language tasks relating to phonological processing, whereas more anterior (BA 44/45) and anterior ventral regions (BA 45/47) code syntactic and semantic processing, respectively. This already points to a hierarchical functional organization of the inferior frontal gyrus.

In more recent years, the role of the IFG in other domains has been emphasized (Liakakis et al., 2011; Uddén & Balhmann, 2012),[12] and unifying functional models have been proposed, especially with an emphasis on executive (Koechlin & Summerfield, 2007) and cognitive control and memory (Badre & Wagner, 2007). In their cascade model of executive control, Koechlin and colleagues have proposed a hierarchical gradient for executive control, lying along the anterior-posterior axis in the lateral prefrontal cortex (lPFC) (e.g., Koechlin & Jubault, 2006; Koechlin & Summerfield, 2007). As implied by its name, the hierarchy is thought to be a cascade whereby top-down control is exerted from anterior to posterior frontal regions (Koechlin, Ody, & Kouneiher, 2003). This theory is not solely based on the IFG but includes other lateral prefrontal areas (Koechlin & Summerfield, 2007). Here, the premotor cortex lies at the base of the hierarchy and its function is to control stimulus-response patterns that are temporally in close relation, such as picking up the telephone when it is ringing (Koechlin & Summerfield, 2007). On the next level of the hierarchy, the posterior lPFC also includes contextual information and thus integrates additional information in which the stimulus occurs, e.g., not picking up another person's ringing telephone. The authors refer to this process as *contextual control*. The posterior lPFC could roughly corresponds to BA 44/45. On the next higher level, the anterior lPFC (~ BA 47) also takes preceding events into consideration in order

---

[12] The domains include music processing, action observation (Casper et al., 2010), creativity (Kröger et al., 2012), and emotion recognition and evaluation (Carr, Iacoboni, Dubeau, Mazzioatta, & Lenzi, 2003; Seitz et al., 2008).

to select a specific action, termed *episodic control*. Sticking with the above-mentioned example, although you are at your friend's house and it is their telephone that is ringing, you pick it up because your friend previously asked you to do so in the case of its ringing. At the highest level of the hierarchy lies the polar lPFC, which is suggested to be in charge of *branching* — when the execution of an action can be put on hold while another action is performed — thereby enabling long-term planning.

Badre and Wagner (2007) have proposed a *theory of cognitive control of memory* for the IFG. According to this account, two processes are important: *controlled retrieval*, activating goal-relevant information in a top-down manner when bottom-up cues are insufficient (i.e., there is no automatic activation), and *post-retrieval selection*, a process that resolves competition between simultaneously activated memory representations. While they suggest that the former process (controlled retrieval) is coded by the anterior BA 47, post-retrieval selection is suggested to take place in more posterior sites (BA 45).[13]

Badre and Wagner (2007) also propose a rostro-caudal axis of the IFG, reflecting a processing hierarchy en route to action (see also Petrides, 2005). Thus, they propose a hierarchical relationship between the IFG regions: the more posterior the area, the more it is concretely tied to an overt response, while more anterior sites represent more abstract action-related representations and are therefore associated with high-level goals retrieved from long-term memory (Badre & Wagner, 2007; Badre & D'Esposito, 2009; cf. Buckner, 2003).[14] In a similar vein, Uddén and Bahlmann (2012) have proposed the structured *sequence processing theory*: a linguistic theory also applicable to other domains. Again, the IFG is suggested to be

---

[13] Here, the proposed functions of BA 45 and BA 47 are not in conflict with those proposed by the cascadic model (Koechlin & Summerfield, 2007), i.e., contextual and episodic control, respectively. That is, in the case of post-retrieval selection, several memory representations are activated at the same time as the stimulus and are hence contextual. Moreover, semantic retrieval (from long-term memory) entails the consideration of previous information, concurring with the concept of episodic control.

[14] As such, the anterior BA 47 is linked to the temporal cortex, which is important in semantic knowledge storage and retrieval (Badre & Wagner, 2007).

structured by a rostro-caudal gradient for temporal abstraction and domain generality. While caudal areas are thought to process temporally less abstract stimuli, like phonemes and words, more rostral areas are thought to integrate these processes at a higher level of temporal abstraction and are domain general. Here, BA 44 is suggested to process sequences coming from premotor regions (Uddén & Bahlman, 2012).

Summarized and in line with the above-stated accounts on the IFG, the suggestion is that BA 44 is involved in structuring sequences to realize particular outcomes (Fiebach & Schubotz, 2006; Grafman, 2002). With regard to action understanding, it thus potentially supports the anticipation of upcoming action steps during action perception (Csibra, 2007; Fagg & Arbib, 1998; Friston et al., 2011; Schubotz & von Cramon, 2009). While BA 45 activation supports the selection between competitively activated semantic representations (Gold et al., 2006; Moss et al., 2005; Badre & Wagner, 2007), BA 47 is involved in the top-down semantic retrieval (Badre & Wagner, 2007) of goal-relevant knowledge. The proposed function of BA 47 is in line with studies concentrating on creativity, and more specifically, contextual expansion (Kröger et al., 2012). Here, increased activity of BA 47 was found when participants were asked to think of unusual functions of an object, i.e., using a shoe as a plant pot.

With regard to the IFG's role in integrating contextual information during action perception, it was found (as outlined in Section 1.3.2) that action–room incompatibility (e.g., squeezing a lemon in the bathroom) particularly challenges the (left) IFG, probably reflecting efforts for semantic integration (Wurm & Schubotz, 2012). The same is true when an actor performed unrelated action steps in succession — i.e., action steps that did not match their current goal (Hrkać et al., 2014) — or when otherwise unrelated action steps shared a common object (Hrkać, Wurm, Kühn & Schubotz, 2015). This points to an attempt to integrate the information under a common overarching action goal. However, the variety of paradigms in which the IFG is reported to function in different (albeit related) tasks shows that it is still a matter of debate as to how this area contributes to action perception. In particular, it remains uncertain at which levels and under which circumstances the IFG integrates contextual action information into action observation.

2. Objective of the Thesis

As reviewed in the above sections, action observers consider multiple sources of information when observing an action and inferring its goal. The present thesis aims at extending the existing work on the impact of contextual action information on action observation by investigating how *contextual objects* (COs), which are part of a particular action scene yet never integrated in the observed action, influence action observation. COs not only signify a specific room category (e.g., coffee machine implies kitchen), thereby telling the observer about the general probabilities of an action (e.g., cooking vs brushing teeth), but rather they also often give information about measures the actor has adopted before performing the observed manipulation. Thus, COs are often prepared for the achievement of — and hence inform an observer about — an overarching goal. For instance, cracking an egg next to a frying pan implies the actor is going to make scrambled eggs. This means that upcoming action steps are implied and the observer can adapt their expectation accordingly. For example, an actor who is cracking an egg may do so because he/she is engaged in baking a cake. Hence, he/she has prepared the kitchen worktop in front of them, which is now equipped with a flour bag, a sugar bowl, eggs in a carton and a measuring cup. Crucially, it is not only the selection of objects hinting at a particular room that is informative, but also the objects' condition and their spatial arrangement, both with regard to other objects and the actor.

I conducted several behavioural and two fMRI experiments to investigate the impact of COs on action observation. The basic approach was to let subjects attend to action videos that either contained a CO or did not, and to respond to verbal action descriptions referring to these actions by button press (accept or reject). Those videos containing a CO were varied according to two experimental factors: 1) *goal affinity* (GA), and 2) *location ergonomics* (LE). *Goal affinity* is defined as an object's semantic relation to the observed action, quantified by the probability with which the object will be employed in the same action. Accordingly, the higher the goal affinity of a CO, the more an observer will expect the CO to be soon integrated into the observed action. On the other hand, *location ergonomics*, defined as the spatial relation to an observed action, are meant to quantify the probability with which an object can be reached by an actor. In order to build expectations about the observed action, a CO's location in relation to the actor is meaningful to the extent that objects that are needed for an action must be easy to reach and grasp for subsequent efficient usage. By systematically varying these experimental factors, I aimed to investigate whether contextual objects are processed and integrated during the observation of an action, and if so, how this effect is grounded. The following specific *research questions* were addressed in Studies 1, 2 and 3, respectively:

1 Are COs processed with regard to the observed action?

2 Do action observers in a third-person-perspective (3pp) paradigm establish their own (virtual) action space, by perceiving COs in a scene not only with respect to the observed action, but also with reference to themselves?

3 Is the processing of COs driven by their informative value to the observed action and its predicted outcome?

In Study 1, I investigated the neural mechanisms underlying the integration of COs in semantic and spatial relation to observed actions. I hypothesized the goal affinity of COs to increase activity in the OTC, the IPL and the IFG. In addition, I expected that the goal affinity of a CO would have a specific effect when located in the preferred action space of the actor. In this case, COs would imply their upcoming usage particularly strongly and thus might trigger expectations of a concrete action sequence. I specifically tested the hypothesis that the IFG subserves this integration. To investigate whether COs are processed in terms of action probabilities, I modulated the factors *goal affinity* and *location ergonomics* as follows: I quantified goal affinity in an online survey (N = 500), measuring the subjective ratings of students regarding the associative strength of objects (N = 144) and actions (N = 72). Note that the COs and actions always belonged to the same room category (either kitchen or office), resulting in an overall compatibility between CO, action and context background (e.g., cracking an egg in the kitchen next to a frying pan). Here, the mismatch was solely evoked by the associative semantic strength between the CO and a specific action. Based on this pilot data, objects were assigned to four different levels of goal affinity ranging from "very low associated" to "very high associated" (Study 1). Subsequently, COs of level one ("very low associated") and level two ("rather low associated") were merged to form the level *low goal affinity COs* (Study 1), also termed *lowly congruent COs* in Study 3,[15] whereas level three ("rather high associated") and level four ("very high associated") were merged to form the level *high goal affinity COs* (Study 1), termed *highly congruent COs* in Study 3.[16] Location ergonomics was quantified by a pilot study (N = 24) in which the reachability of an object in 60 different positions on a table had to be rated. Based on these results, the table was subdivided into four quadrants: close-right (cr), close-left (cl), far-right (fr), and far-left (fl). Moreover, for each quadrant three positions were selected, which differed in regard to their reachability (high, middle, low) and, at the same time, mirrored the positions of the other quadrants. This resulted in a total of 12 positions for the possible placement of COs during the observation of actions (used in Studies 1, 2 and 3). Moreover, in this pilot study, a second task instructed participants to position an object in the most convenient location they could think of for subsequent use (Study 1). For the fMRI experiment, final videos were selected according to these two factors and piloted with regard to their recognizability.

In Study 2, I tested whether, depending on their location, COs can trigger an observer-referenced processing. If so, I expected that the more convenient the spatial location of the CO for the hypothetical grasp of the observer, the stronger the responses in the brain regions representing this observer perspective would be. I specifically expected a significant increase

---

[15] I will use the term *lowly congruent* COs hereafter.

[16] I will use the term *highly congruent* COs hereafter.

of activity in the aIPL, the PMv and the insula. To this end, I extended the *location ergonomic*s factor through a behavioural study in which 24 right-handed participants were asked to rate the ease with which they thought they could reach and grasp the CO as active counterparts in the final action videos of Study 1. Based on this pilot data, a reachability-by-observer map could be established. Subsequently, I used the behavioural data as a predictor to reanalyse the fMRI data set from Study 1.

Finally, in Study 3, I investigated whether the processing of COs was driven by their informative value in relation to the observed action. Extending my findings from Study 1, this second fMRI study specifically tested whether COs that matched neither the current action nor the corresponding room category, termed *incongruent COs*, strongly impact on action observation. I hypothesized that both highly congruent COs and incongruent COs provide rich information regarding potential upcoming action steps. Therefore, I expected an overlap of increased neural activity in the OTC and IPL, elicited by highly congruent and incongruent COs, when compared to lowly congruent COs. I concentrated particularly on the IFG, due to its established role in the retrieval and integration of action-relevant semantic information (Casper et al., 2010; Badre & Wagner, 2007). To this end, I extended the factor goal affinity by including the additional level *incongruent COs*. Incongruent COs neither match the observed action nor belonging room category. To determine the incongruent COs, a further pilot study was conducted where COs belonging to one contextual background or room category (e.g., kitchen) were tested for the probability of their occurring in another contextual background (office). 24 right-handed participants were asked to rate on a six-point Likert scale how strongly the presented CO (e.g., rolling pin) fitted into the alternative room category (e.g., office). Those objects creating the biggest mismatch according to these pilot data were chosen for the level *incongruent COs* (Study 3).

3. Research Articles

3.1 Study 1: Making sense of objects lying around: How contextual objects shape brain activity during action observation

Nadiya El-Sourani, Moritz F. Wurm, Ima Trempler, Gereon R. Fink, & Ricarda I. Schubotz (2018)

Neuroimage, 167, 429-437.

NeuroImage 167 (2018) 429–437

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Making sense of objects lying around: How contextual objects shape brain activity during action observation

Nadiya El-Sourani [a,b,*], Moritz F. Wurm [d], Ima Trempler [a], Gereon R. Fink [b,c], Ricarda I. Schubotz [a,b]

[a] Department of Psychology, Westfälische Wilhelms-Universität, 48149 Münster, Germany
[b] Institute of Neuroscience and Medicine (INM3), Cognitive Neuroscience, Research Centre Jülich, 52425 Jülich, Germany
[c] Department of Neurology, University Hospital Cologne, 50937 Cologne, Germany
[d] Center for Mind/Brain Sciences (CIMeC), University of Trento, Rovereto, TN, Italy

A R T I C L E I N F O

Keywords:
Semantic integration
Object perception
Action observation
fMRI
Inferior frontal gyrus

A B S T R A C T

Action recognition involves not only the readout of body movements and involved objects but also the integration of contextual information, e.g. the environment in which an action takes place. Notably, inferring superordinate goals and generating predictions about forthcoming action steps should benefit from screening the actor's immediate environment, in particular objects located in the actor's peripersonal space and thus potentially used in following action steps. Critically, if such *contextual objects* (COs) afford actions that are semantically related to the observed action, they may trigger or facilitate the inference of goals and the prediction of following actions.

This fMRI study investigated the neural mechanisms underlying the integration of COs in semantic and spatial relation to observed actions. Specifically, we tested the hypothesis that the inferior frontal gyrus (IFG) subserves this integration. Participants observed action videos in which COs and observed actions had common overarching goals or not (*goal affinity*) and varied in their location relative to the actor.

High goal affinity increased bilateral activity in action observation network nodes, i.e. the occipitotemporal cortex and the intraparietal sulcus, but also in the precuneus and middle frontal gyri. This finding suggests that the semantic relation between COs and actions is considered during action observation and triggers (rather than facilitates) processes beyond those usually involved in action observation. Moreover, COs with high goal affinity located close to the actor's dominant hand additionally engaged bilateral IFG, corroborating the view that IFG is critically involved in the integration of action steps under a common overarching goal.

Introduction

Ecologically, it is of utter importance that we observe and understand what people around us are doing. To properly interact with each other, the ability to derive intentions, to capture action goals and to predict future steps of an unfolding action is highly relevant. Consequently, a large body of research has focused on the question of how we capture, simply by observation, what others aim to do (Van Overwalle and Beatens, 2009; Caspers et al., 2010). However, since actions are complex and carry different sorts of information it is not surprising that it remains to be investigated i) how information is exploited during the observation of actions, and ii) how this information helps to constrain the observer's expectation.

When we perceive an object-directed action we are typically confronted with *manipulation information* including hand posture and movement trajectories, and *object information* including pragmatic and semantic object properties. Both information types are relevant for action recognition and are referred to as *core information,* hereafter. For example, after watching somebody cracking (manipulation movement) an egg (object information) one could easily name the action, i.e., cracking an egg. However, it is not before detecting the pan nearby that one would expect the actor to prepare scrambled eggs rather than baking a cake. Hence, information that is not necessary for the recognition of an on-going action might be crucial for higher-level inferences of goals and the prediction of forthcoming action steps (Malcolm et al., 2016). This information, which is referred to as *contextual information*, hereafter, potentially modulates action perception.

While the influence of contextual information on object recognition

---

* Corresponding author. Fliednerstr. 21, 48149 Münster, Germany.
E-mail address: n_elso02@uni-muenster.de (N. El-Sourani).

https://doi.org/10.1016/j.neuroimage.2017.11.047
Received 26 June 2017; Received in revised form 15 November 2017; Accepted 21 November 2017

has been investigated (Bar, 2004; Boyce et al., 1989; Hayes et al., 2007; Zimmermann et al., 2010), contextual influence on action perception has attracted little attention thus far. For this reason, in recent studies we concentrated on the impact of contextual information on action perception. We focused on the room and actor information during the observation of an action. As with actor information (Wurm et al., 2011; Hrkać et al., 2014), our data revealed that participants' brains process room information spontaneously, i.e., without instruction or task requirements (Wurm and Schubotz, 2012). Participants even process room information when it is in conflict with the action itself: subjects needed longer to indicate the recognition of an action when the action took place in an incompatible as compared to a compatible or neutral room (Wurm and Schubotz, 2012). On the other hand, action-compatible room information helped when actions were difficult to recognize (Wurm and Schubotz, 2017; Wurm et al., 2017). Moreover, a misfit between the room, the manipulated objects and the applied manipulation was associated with increased activity of brain regions associated with object and action processing, respectively, even when participants were not required to deal with this conflict (Wurm et al., 2012).

The present study takes the position that *contextual objects (COs)*, which are often part of an observed action scene albeit that they are not integrated into the observed action (Fig. 1), should also exert an impact on action observation. On the one hand, contextual objects often signify the room category (e.g., a knife block implies 'kitchen') and so inform the observer about the general probability of an action. On the other hand, contextual objects can also inform us about measures an actor has adopted before performing the observed manipulation. Thus, contextual objects are often prepared for the achievement of (and hence inform an observer about) an overarching goal (cf. Iacoboni et al., 2005), that is, a desired end state or outcome (Csibra and Gergerly, 2007). This should be especially the case when contextual objects are located close and right to the actor (i.e. actor's peripersonal space), as this area yields special expectations for subsequent usage of that object. In this case, contextual objects may undergo a shift in their meaning: from contributing to the general probability of actions to becoming a potential future target of the observed actor.

While we found that manipulated objects have a significant impact on action processing due to their implicating of possible actions (Schubotz et al., 2014; Hrkać et al., 2015), we know very little about the impact of contextual objects. If it is part of the information provided by an object which actions can be performed with it, contextual objects should also signal possible actions. The question hence arises: Do contextual objects also imply possible actions, and if so, how do observers deal with this additional information? We here focus on two factors that might modulate the impact of contextual object information on action perception: *goal affinity (GA)* and *location ergonomics (LE)*.

Goal affinity is defined as an object's semantic relation to the observed action, quantified by the probability with which the object becomes employed in the same action. Accordingly, the higher the goal affinity of a contextual object, the more an observer will expect the contextual object to be soon integrated into the observed action. Functional MRI studies have shown that the mere perception of objects implicates manipulation and action (Buxbaum et al., 2006; Johnson-Frey, 2004; Schubotz et al., 2014). Correspondingly, we expected the goal affinity of contextual objects to increase activity in brain areas that are related to object-related action representation, particularly the occipito-temporal cortex (OTC; Wigget and Downing, 2011) and the inferior parietal lobule (IPL; Buxbaum and Kalénine, 2010; Schubotz et al., 2014). Moreover, we expected that the goal affinity of contextual objects correlates with activity in the inferior frontal gyrus (IFG), as we found this

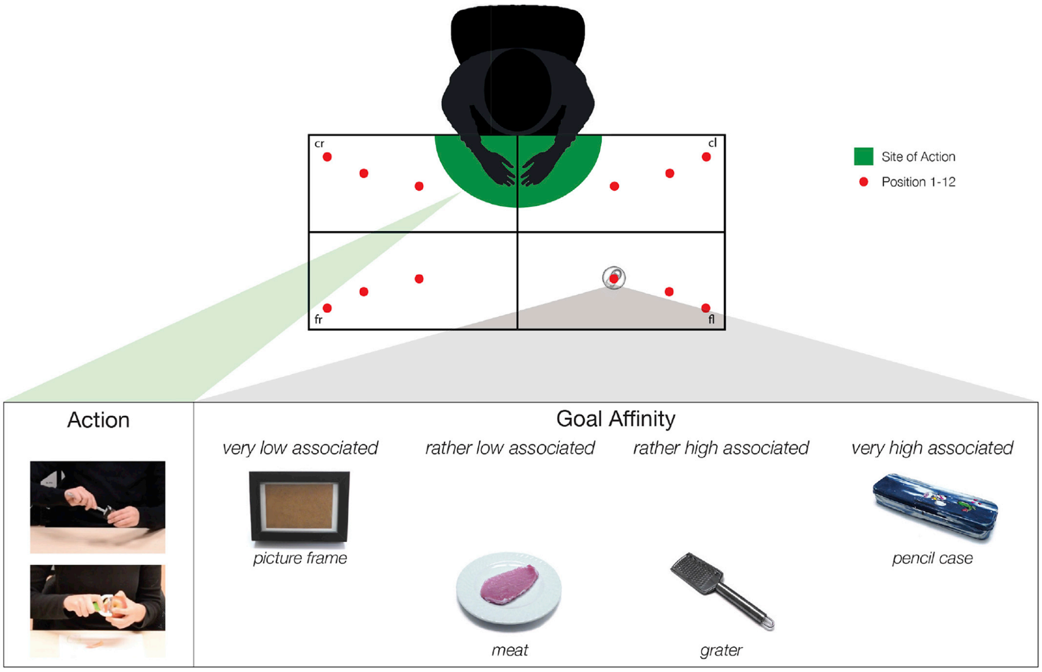


Fig. 1. Exemplary demonstration of experimental factors: Red dots refer to location ergonomics and include positions 1–12. Further positions are subdivided into quadrants from the actor's perspective: close-right (cr), close-left (cl), far-right (fr) and far-left (fl). Goal affinity levels of contextual objects for distinct actions are depicted in the box.

area to be modulated by other types of contextual information, specifically room and actor information, during action observation (Wurm and Schubotz, 2012; Hrkać et al., 2014; Hrkać et al., 2015). In action observation, the IFG is suggested to be engaged in the retrieval and integration of action-relevant semantic information (see also Caspers et al., 2010; Kilner, 2011).

On the other hand, *location ergonomics*, the spatial relation to an observed action, are meant to quantify the probability with which an object can be reached by an actor. For example, right-handers have a general bias for objects on their right-hand-side (Toney and Thomas, 2006; Rezaee et al., 2010; Bryden et al., 2000). This could be confirmed by a pilot study, in which we asked participants to place an object in the most convenient location they could think of for subsequent use. In order to build expectations about the observed action, the location of a contextual object in relation to the actor is meaningful in such a way that objects that are needed for an action have to be easily reached and grasped for subsequent efficient use. So far, object reachability has been studied only with regard to its effect on action execution, but not regarding the effect on goal inferences. It was found that grip facilitation depends on the object's availability to manipulation, pointing to an interaction between the object's pragmatic properties and the action goal implied by the contextual setting (Kalénine et al., 2013). Furthermore, current research on the effects of hand proximity in relation to objects showed that placing the hands near an object was found to initiate the visual system for the processing of visually guided actions (Gozli et al., 2012). Against this backdrop, we expected that goal affinity of contextual objects should have a specific effect when located close and right to the actor. Our own previous findings implied that objects with a low goal affinity would then lead to increased IFG activity as they would challenge attempts to be integrated into the currently observed action (Wurm and Schubotz, 2012; Hrkać et al., 2014). However, in principle, also the opposite effect is conceivable: IFG activity increases for contextual objects with both a strong semantic and spatial relation to the observed action, because when contextual objects score high on both dimensions, they might trigger expectations of a concrete action sequence. These opposite effects would speak in favor of alternative processing modes: in the former, the brain selects more information from the scene than is minimally needed to interpret the currently observed object manipulation; then, it strives to resolve potentially conflicting information. In contrast, the latter effect would be expected when the brain focuses only on core information (manipulated object, manipulating hands) and selects from the rest of the scene only compatible information to strengthen the current interpretation of the object manipulation with regard to upcoming action steps.

Materials and methods

Participants

Twenty-one right-handed subjects (12 females; 24.00 ± 3.25 years old; range, 20–30 years) with normal or corrected-to-normal vision participated in the study. None of them reported a history of medical, neurological or psychiatric disorders, or substance abuse. The study protocol was conducted in accordance with ethical standards of the Declaration of Helsinki and approved by the local Ethics Committee of the University of Münster. Each participant submitted a signed informed consent and received reimbursement or course credits for their participation afterwards.

Stimuli

During the scanning session participants were presented with 360 video clips showing actions (action trials) and with 72 written action descriptions referring to these actions (question trials) (presentation software: Presentation Version 13.1, Neurobehavioral Systems). Each trial (6 s) started with a video clip or a question (3 s), which was followed by a fixation phase (3 s). To enhance the temporal resolution of the BOLD response, a variable jitter (500, 1000, 1500 ms) was included after the fixation phase. In 5% ($N = 18$) of the trials a null event was implemented (6 s), which required participants to fixate a fixation cross.

Each video showed a single object-directed action performed by one and the same actress sitting at a table and was filmed from a frontal third person perspective (3pp). Each of the actions was performed in a typical setting (e.g., cracking an egg in the kitchen, writing a letter in the office), resulting in compatibility between action and context-background. Settings were either a kitchen background or an office background. In total, 39 action videos in the kitchen and 33 in the office were selected.

Each video involved two target objects (e.g., a lemon and a squeezer in „squeezing a lemon“). The actress's face was inclined downwards in order to minimize cognitive effects of person perception (Fig. 2). In a pilot study (C) all videos were investigated regarding their recognizability. Only videos were selected in which 100% of the participants recognized the action correctly.

Out of the 360 action videos presented to each participant, 288 included an additional contextual object, which was always positioned in front of the actress on a table (Fig. 2). In order to investigate whether those objects are processed in terms of action probabilities we modulated them due to two experimental factors: *location ergonomics* and *goal affinity*.

Location ergonomics was quantified by a pilot study A ($N = 24$) in which the reachability of an object at 60 varying positions on a table had to be rated. The mean standard deviation of this rating was 0.579. Based on these results the table was subdivided into four quadrants: close-right (cr), close-left (cl), far-right (fr), and far-left (fl). Furthermore, for each quadrant three positions were selected, which differed with their regards to reachability (high, middle, low) and, at the same time, mirrored the positions of the other quadrants. This resulted in a total number of 12 positions for the placement of contextual objects during the observation of action videos (Fig. 1). In a second task we asked participant to place a "power-grip" object in the most convenient location they can think of for subsequent use. Mean values of x and y coordinates on the table show that participants' preferred location for subsequent usage was close and right to the action site.

Goal affinity was quantified on the basis of subjective ratings of a large sample ($N = 500$) of students (pilot study B). The mean standard deviation of this rating was 0.645. Based on pilot data B, 144 different objects were assigned to four different levels of goal affinity ranging from "very low associated" to "very high associated". Each of the 72 actions was paired with two objects of two different goal affinity levels. However, these pairs were selected in such a way that the factor goal affinity was evenly distributed across all levels (Fig. 1). Videos were arranged and selected for the fMRI session so that each action was seen five times in total, once without and four times with one contextual object. Furthermore, those action videos containing a contextual object were arranged in a way that all goal affinity levels at all 12 positions occurred in an evenly distributed number (12 positions x 4 goal affinity levels x 6 occurrences = 288 action videos with a contextual object). Videos were presented in a pseudo-randomized fashion in order to avoid direct repetition of action, goal affinity, and location of the contextual object. All factor levels were presented in an evenly distributed manner.

Task

Participants were instructed to watch the video clips attentively. They were told that after some of the video clips an action description would appear that referred to the content of the preceding video, and that had either to be accepted or rejected by the participants. Participants were naïve with regard to the ratio of videos followed by a question trial. The ratio of action and catch trials was 1:5 (20% catch trials); 50% of trials were to be affirmed and 50% to be rejected. Action descriptions (e.g., squeezing a lemon) did not amount to an overarching action goal (e.g.,

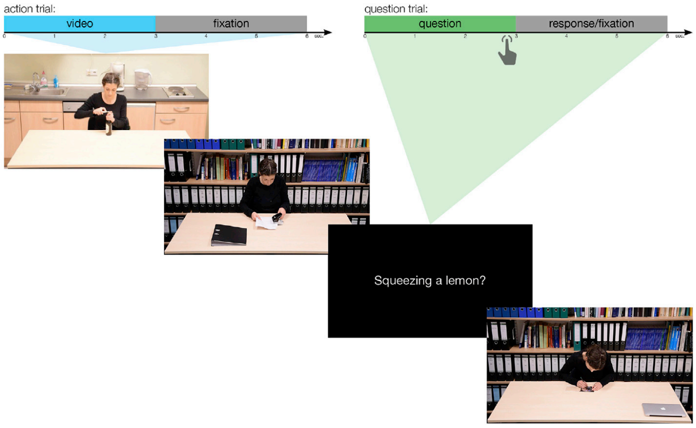


Fig. 2. Schematic diagram of the task. Action trials consisted of an action video (3s) and a fixation phase (3s). Question trials consisted of a question regarding the preceding video trial (n-1), followed by a response and fixation phase.

preparing a meal) but rather to each single action itself and hence a short-term goal. Action descriptions were presented in a pseudo-randomized fashion mixed with the experimental trials. Responses were given on a two-button response box, using the index finger to accept and the middle finger to reject the action description at hand. This method was used for the purpose of keeping the participants' attention on track while watching the videos (Wurm and Schubotz, 2012; Hrkać et al., 2015). Error rates were analyzed to assess participants' behavioral performance. Finally, a training phase of five minutes was included before the fMRI session in order to familiarize the participants with the task.

fMRI image acquisition

Imaging was performed on a 3 Tesla Siemens Magnetom Prisma MR tomograph using a 20 channel head coil. Participants were located in a supine position on the scanner bed with their right index and middle fingers positioned on the appropriate response buttons of a response box. To minimize head and arm motions, head and arms were tightly fixated with form-fitting cushions. Furthermore, participants were provided with earplugs in order to attenuate the scanner noise. Whole-brain functional images were acquired using a gradient T2*-weighted single-shot echo-planar imaging (EPI) sequence sensitive to blood oxygenation level dependent (BOLD) contrast (64 × 64 data acquisition matrix, 192 mm field of view, 90° flip angle, TR = 2000 ms, TE = 30 ms). Each volume consisted of thirty adjacent axial slices with a slice thickness of 4 mm and a gap of 1 mm, which resulted in a voxel size of 3 × 3 × 5 mm. Images were acquired in ascending order along the AC-PC plane to provide a whole-brain coverage. After functional imaging, structural data were acquired for each participant using a standard Siemens 3D T1-weighted MPRAGE sequence for detailed reconstruction of anatomy with isotropic voxels (1 × 1 × 1 mm) in a 256 mm field of view (256 × 256 matrix, 192 slices, TR = 2130, TE = 2.28).

In order to present the stimuli during the scanner session, a 45° mirror was fixated on the top of the head coil. A video-projector projected the experiment on a screen that was positioned behind the subject's head, so that participants could see the stimuli via the mirror. The mirror was adjusted for each participant to provide a perfect view (center of the field of vision). In a pilot study C we controlled for recognizability of actions and contextual objects in a similar experimental setting. Only action videos in which the action and the contextual object could be identified by at least 95% of the participants were employed in the present study.

fMRI data analysis

fMRI data preprocessing

Brain image preprocessing and basic statistical analyses were conducted using LIPSIA software package, version 3.0 (Lohmann et al., 2001). Initially, spikes in time series were corrected by interpolating them with adjacent time points. To correct for temporal offsets between the slices acquired in one scan, a cubic-spline interpolation was used. Furthermore, individual functional MR (EPI) images were motion-corrected with the first time-step as reference and six degrees of freedom of which three are rotational and three translational. Then, the average across all time points of this corrected data was used as reference scan for a second pass of motion-correction. Motion correction estimates were inspected visually. Coregistration was done using statistical parametric mapping package (SPM12, Wellcome Department of Imaging Neuroscience, London, UK). The images were coregistered and transformed into a standard stereotactic space using the intercommissural line as the reference plane for transformation (Ashburner and Friston, 1997). Anatomical datasets were normalized to the ICBM/MNI space by linear scaling. The resulting parameters were then used to transform all functional slices employing a trilinear interpolation. Resulting data had a spatial resolution of 3 × 3 × 3 mm (27 mm^3). Normalized functional images were spatially smoothed with a Gaussian kernel of 6 mm full width at half-maximum (FWHM). Finally, a temporal high-pass filter of 1/125 HZ was applied to the data in order to remove low-frequency noise such as scanner drift.

Design specification

The statistical evaluation was based on a least-squares estimation using the general linear model (GLM) for serially autocorrelated observations (Friston et al., 1995; Worsley and Friston, 1995). The design matrix was generated with delta functions and convolved with a canonical hemodynamic response function. Activations were analyzed time-locked to the onset of the videos and the analyzed epoch comprised the full duration (3s) of the presented videos, the duration of the null events (6s), and the reaction time in question trials (max. 3s). The GLM contained 12 regressors: eight predictors for the experimental conditions, one predictor for videos without contextual objects (noCO), one including all the null events (6s fixation phase), one predictor for question trials, and finally one parametric regressor for the iteration of action, which was included as a regressor of nuisance to control for effects of action repetition. The eight predictors for the experimental conditions were assigned with regard to location ergonomics in terms of the four quadrants. Furthermore, goal affinity levels were merged into high (level 3 + 4) and low level (level 1 + 2), which resulted in the following predictors: 1. High goal affinity at close-right (cr) positions; 2. Low goal affinity at close-right (cr) positions; 3. High goal affinity at close-left (cl) positions; 4. Low goal affinity at close-left (cl) positions; 5. High goal affinity at far-right (fr) positions; 6. Low goal affinity at far-right (fr); 7. High goal affinity at far-left (fl) positions; 8. Low goal affinity at far-left (fl) positions.

Influence of contextual objects on action perception. In order to investigate whether the brain automatically seeks to integrate contextual objects with the observed action, we contrasted action videos containing a contextual object with those, which did not (CO > noCO).

Influence of contextual objects goal affinity. For the main effect of goal affinity, all eight predictors regarding goal affinity were used in order to contrast high goal affinity with low goal affinity (GAhigh > GAlow) on a first level GLM.

For the interaction analysis of goal affinity and location ergonomics only close-right and close-left positions were included in the analysis reflecting high and low location ergonomics, respectively. This was done as we hypothesized close-right positions to have a specific effect on the processing of contextual objects, due to the preferences of actors to deposit objects that they plan to use in upcoming action steps at such locations (Pilot Study A).

Group analysis

To obtain group statistics, the resulting contrast images of all participants were entered into a second level random-effects analysis using a one-sample *t*-test across participants to test for significant deviation from zero. We then corrected for multiple comparisons across all voxels using the threshold-free cluster enhancement (TFCE) method (Smith and Nichols, 2009). The significance level for whole-brain activations was set to $p < 0.05$ TFCE-corrected. Default TFCE parameters $H = 2$ and $E = 0.5$ were used (Smith and Nichols, 2009).

SVC Analysis. To specifically test the hypothesis that the IFG is modulated as function of the interaction of the goal affinity and the location of the contextual object, a small volume correction (SVC) was performed on the interaction contrast goal affinity (high, low) x location ergonomics (close-right (cr), close-left (cl)) at $p < 0.05$ TFCE-SVC-corrected. Anatomical masks of left and right Brodmann area (BA) 44 and 45 (Amunts et al., 1999) were defined based upon the Anatomical Toolbox (Eickhoff et al., 2005).

Results

Behavioral results

Performance during the fMRI session was assessed by error rates and reaction times on correctly answered trials. The average response time was 1353 ± 65 ms and the average error rate was low (1.23 ± 2.16%) indicating that participants attentively observed and recognized the actions.

Moreover, in a post-fMRI interview four out of the 21 subjects reported not having noticed the CO during the course of the experiment, i.e. they did not consciously process these objects. It would be great, however, to find out whether implicit and explicit processing of these contextual objects would lead to differential activation patterns. Unfortunately the sample size was too small in order to conduct a valid statistical comparison.

fMRI results

Main effect of contextual objects

To investigate brain activity triggered by the mere presence of contextual objects, we contrasted action trials containing contextual objects compared with action trials, which did not (CO > noCO). After applying a TFCE-correction at $p < 0.05$, the inclusion of a contextual object revealed activation of the left intraparietal sulcus (IPS; descending branch, middle occipital gyrus), peristriate regions of the occipital cortex (BA 18, 19) and the fusiform gyrus (BA 19, 37) bilaterally, with stronger activation in the left than the right hemisphere (Fig. 3; Table 1).

Main effect of goal affinity

For the main effect of goal affinity, we tested for differences between high vs. low goal affinity trials (GAhigh > GAlow). We hypothesized that goal affinity of the contextual objects modulates activity of the OTC and the IPL. TFCE corrected at $p < 0.05$, we found the OTC (local maxima in posterior middle temporal gyrus and inferior temporal gyrus) and the IPL (IPS) to be activated bilaterally, for high compared to low goal affinity trials. In addition, the dorsolateral prefrontal cortex (dlPFC) was activated bilaterally, more specifically in BA 8, 9 and 10. Finally, we found activity in the precuneus and the cerebellum. The reverse contrast (GAlow > GAhigh) revealed no significant activation.

To test whether the activations found for high compared to low goal affinity were due to an increase in activation for contextual objects with a high goal affinity rather than a decrease for contextual objects with low goal affinity, we calculated a conjunction of the TFCE-corrected contrast GAhigh > GAlow and the TFCE-corrected contrast GAhigh > Rest (Null-Events), i.e. (GAhigh > GAlow) ∩ (GAhigh > Rest). When corrected at $p < 0.05$, the conjunction analysis revealed bilateral activation of the IPS (including aIPS), the dlPFC, the OTC, the precuneus, and the cerebellum (Fig. 4; Table 2). The results indicate that the effect of $GA_{high} > GA_{low}$ did in fact become apparent due to an increase in activation for contextual objects with a high goal affinity. The conjunction analysis for low GA $((GA_{low} > GA_{high}) \cap (GA_{low} > Rest))$ revealed no significant activations.

Interaction of goal affinity by actor side

In order to investigate whether goal affinity effects differed depending on the contextual objects location, we conducted a SVC-based analysis of the left and right IFG for the interaction contrast goal affinity (GAh, GAl) x location ergonomics (close-right (cr), close-left (cl)). The analysis revealed significant engagement of the left and right IFG (Fig. 5), when TFCE-corrected at $p < 0.05$. Mean beta values of the significantly activated clusters can be found in Fig. 5.

Discussion

The present study aimed at gaining insight into the question if and how the brain processes nearby objects (contextual objects, COs) that are not involved in an observed action. We manipulated the contextual objects' semantic relation (goal affinity, GA) as well as their spatial relation (location ergonomics, LE) to the observed action. We expected both factors to influence action observation such that, first, the higher the contextual object's goal affinity the stronger the observing brain's readiness to integrate it into the observed action. Second, we expected that goal affinity would have an additional effect for contextual objects presented close and right to the actor, i.e., those contextual objects that were potentially easy to reach for the actor (Pilot study A).

We found that contextual objects are processed by brain areas that transcend simple visual processing of man-made objects. In particular, activity increased at several sites when the contextual objects' semantic relationship to the observed action was strong. Notably, when such contextual objects with a high goal affinity were presented close and to the right side of the actor in contrast to close left, we observed additional effects. We had hypothesized modulation of the IFG, which is known to be involved in the retrieval and integration of action-relevant semantic information. In the following, we will outline the view that the impact of the actor's location on the contextual object's goal affinity suggests that objects lying close to the actor's dominant hand are processed as being potential future targets of the observed actor.

Our findings show that the brain processes objects in the vicinity of an observed action, and it does so in a telling manner. We here have to

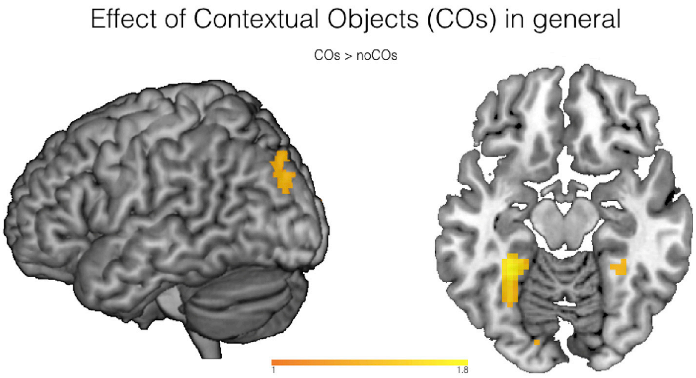


Fig. 3. Effects of contextual objects, at p < 0.05, TFCE-corrected. P-values are scaled to –log10 (p).

Table 1
fMRI activations for the presence of contextual objects (CO > noCO).

Region[a]	Side	BA	MNI Coordinates			*p* values[b]/local maxima
			x	*y*	*z*	
Fusiform gyrus[c]	L	37/19	−24	−76	−7	1.9
	R	37/19	30	−46	−10	1.5
Middle occipital gyrus	L	–	−27	−85	32	1.4
	R	19	36	−85	26	1.4
Cuneus	R	18	18	−91	8	1.3

R, Right; L, Left; x, y, z, MNI coordinates of peak voxel activation.
[a] According to the Anatomy Toolbox (Amunts et al., 1999).
[b] TFCE-corrected for multiple comparison: p-values are scaled to –log10 (p).
[c] Extending into left BA 18 + 19.

distinguish between the effects induced by the mere presence of a contextual object on the table (vs. no contextual object), and the effects induced by the goal affinity this contextual object has with regard to the observed object manipulation. Regarding the presence of contextual objects, in our experiment contextual objects were from trial to trial located at different sites on a table including locations that were not easily accessible to the actor. Independent of this accessibility, the mere presence of an object anywhere on the table increased neural activation in areas known for visual, haptic, and action-related object processing, including the left IPS (Creem-Regehr, 2009; Grefkes and Fink, 2005; Singh-Curry and Husain, 2009; Ramsey et al., 2011), the fusiform gyrus (cf. Bar et al., 2001; Tyler et al., 2004), and visual association areas, i.e., BA 18 and 19 (Chawla et al., 1999). Activity in the visual association areas, especially BA 19, corresponds to the occipital place area (OPA), which is suggested to represent local elements of an action scene (Kamps et al., 2016).

Remarkably, an additional set of brain areas was engaged in processing *the goal affinity* of the contextual object in relation to the currently observed action, indicating an in-depth processing for this particular object property. This set of areas included the posterior components of the action observation network (AON), i.e., the right IPS (in addition to the left IPS already active for contextual objects vs. no objects) and the OTC. We additionally found activation of the precuneus and the dlPFC, which we had not a priori hypothesized and only shortly speculate upon in the following. While AON activity is already triggered by the observation of short and simple object manipulations, dlPFC and precuneus are often seen to join this network for longer and multi-step actions (e.g., Wurm et al., 2014). The posterior dorsal precuneus, which we found in our contrast, is strongly connected to the co-activated middle frontal gyrus (dlPFC), and both are engaged in visuomotor imagery and action

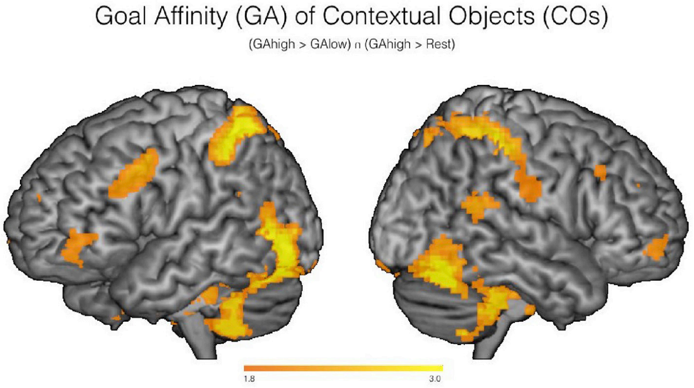


Fig. 4. Effects of goal affinity at p < 0.05, TFCE-corrected. P-values are scaled to –log10 (p).

Table 2
fMRI activations for Goal Affinity (GAhigh > GAlow) ∩ (GAhigh > Rest).

Region[a]	Side	BA	MNI Coordinates x	y	z	p values[b]/local maxima
IPL	L	–	−54	−34	44	3
			−42	−52	59	3
	R	–	45	−40	53	3
			36	−46	59	2.7
MTG/ITG	L	21/22/37	−57	−61	−1	3
	R	21/22/37	52	−66	−6	3
Fusiform gyrus	L	20/36	−27	−7	−37	2.15
	R	20/36	27	3	−44	1.8
Precuneus	L	7	−12	−73	56	2.4
	R	7	12	−67	59	2.3
Middle orbital gyrus	L	10	−45	56	−5	1.9
	R	10	45	59	−7	2.0
MFG	L	8/9	−51	14	38	2.4
	R	8/9	51	23	35	1.86
Cerebellum (Crus 1)	L		−14	−79	−22	3
Cerebellum (V1)	R		15	−73	−22	3

R, Right; L, Left; IPL, inferior parietal lobule; MTG, middle temporal gyrus; ITG, inferior temporal gyrus; MFG, middle frontal gyrus; x, y, z, MNI coordinates of peak voxel activation.

[a] According to the Anatomy Toolbox (Amunts et al., 1999).

[b] TFCE-corrected for multiple comparison: p-values are scaled to –log10 (p).

planning (Zhang & Chiang-shan, 2012). Balser et al. (2014a) reported enhanced precuneus and cerebellar activity for experts vs. novices during the anticipation of sports movements. In a further study, these areas' activity increased as a function of sport experts' anticipatory performance (Balser et al., 2014b). In a similar vein, Calvo-Merino et al. (2005) reported an almost identical network as ours to increase its activity when expert dancers watched previously learned versus novel movements. These activations were interpreted as a recall of acquired action representations, enabling greater skills in anticipating these actions' outcomes (cf. Aglioti et al., 2008). In sum, we take the network for high vs. low goal affinity to reflect enhanced action anticipation in the observers' brains, and thus to corroborate the notion that contextual objects are processed with regard to their potential upcoming usage in an observed action.

To further explore the impact of contextual objects on brain activity during action observation, we tested the hypothesis that goal affinity would have a specific effect when contextual objects were located close and right to the actor. This idea derived from the fact that actors have a preference to deposit objects that they plan to use in upcoming action steps at this location (Pilot study A). We reasoned that the observers' brains should be tuned to integrate particularly those contextual objects in the currently observed action that are endorsed by their ergonomic location; moreover, if these contextual objects score particularly high on goal affinity to the ongoing action, semantic integration should be especially facilitated. Based on former studies, we expected IFG to be the area where such integration effects could surface (Wurm and Schubotz, 2012; Hrkać et al., 2014; see also Kilner, 2011). Indeed, the comparison

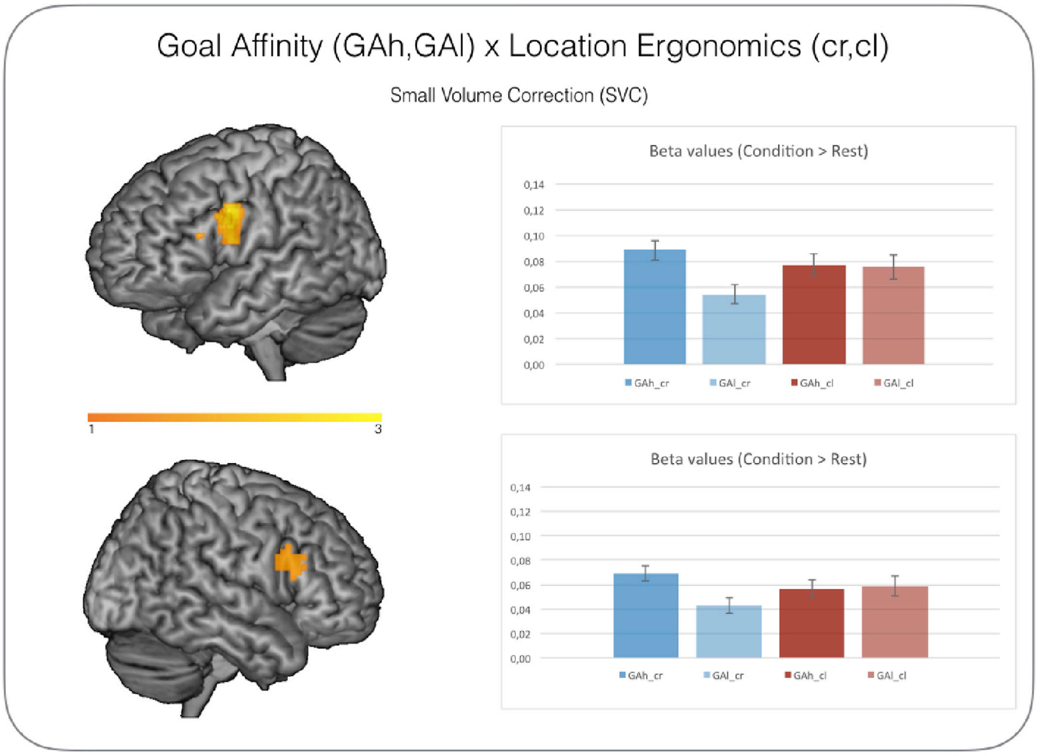


Fig. 5. Effects of IFG SVC, TFCE-corrected at $p < 0.05$. P-values are scaled to –log10 (p). The bar graphs depict beta values of the significantly activated clusters, in order to provide a better understanding of the interaction effect. The IFG was enhanced by contextual objects with high goal affinity when presented close to the actor's right (cr) vs. left (cl).

between close right and close left contextual objects revealed an enhanced goal affinity effect in both right and left IFG: while high and low goal affinity objects induced comparable activity in this brain region when contextual objects were presented to the actor's left hand side, they differed significantly when presented to the actor's right (Fig. 5). The engagement of the IFG in close-left positions of the contextual objects could be interpreted as baseline activation. Hence, when presented close and right to the actor, contextual objects with high goal affinity received a boost in IFG engagement, while low goal affinity contextual objects led to a suppression of IFG activation. This finding supports the view that contextual objects were processed in their relation to the actor and the action: not only with regard to their goal-relatedness, but also with regard to their spatial reachability. The influence of the actor's location on the goal affinity effect implies that contextual objects are perceived as being possible upcoming targets of the actor: the stronger the spatial and semantic relation of the contextual object to the observed manipulation, the more information regarding a specific upcoming action sequence is provided, and possibly, the more a specific action goal can be anticipated. This result is in good accordance with the presumed core function of the IFG in action observation, i.e., the retrieval and integration of action-relevant semantic information (Caspers et al., 2010; Kilner, 2011). Beyond its role in action observation, the IFG has also been linked to response inhibition (e.g. Menon et al., 2001; Rubia et al., 2003; Aron et al., 2004) and to the detection of important cues (Hampshire et al., 2010). We cannot exclude that, and therefore it remains to be elucidated in future studies, whether such processes might have taken part to enhance focus on the currently observed action, as required by our task.

We suggest that, as long as contextual objects are not presented in the close right area of the actor, they are processed with regard to their goal affinity (relative to the observed action), whereas when entering the preferred action zone, this information is used as current impact on upcoming overarching goals.

In previous studies of our own (Hrkać et al., 2014; Wurm and Schubotz, 2012) we found interference rather than enhancement effects for contextual information in action processing. This apparent discrepancy can be resolved by the fact that in Wurm and Schubotz (2012) we investigated the effect of compatibility and incompatibility of room information on action perception, whereas the present study emphasized the effect of compatibility, ranging from unspecific compatibility (low goal affinity) to very specific compatibility (high goal affinity). Thus, in the present study no strong conflict or mismatch emerged by the presence of contextual objects with low goal affinity. Going into more detail, in Wurm and Schubotz (2012) we created compatible room information by employing a set of non-specifically related contextual objects in order to form a room category (e.g., kitchen: stove, kettle, bottle of milk). Hence, contextual objects were processed as part of the 'room information'. As we are used to be surrounded by room-compatible objects not directly fitting our action goals, compatible objects can be neglected by an observer, whereas objects revealing incompatible room information are given more weight. In contrast, in the present study contextual objects were presented in isolation and in front of the actor and action site, while additional room information provided behind the actor was kept constant. Therefore, contextual objects received a special emphasis with regard to their implications for probable goals or – at least – subsequent action steps. Due to their potential to soon become core information, an interference effect of low goal affinity contextual objects was still conceivable, especially when positioned close and right to the actor. The fact that we did not find an effect of low goal affinity on IFG activation suggests that – in line with Wurm and Schubotz (2012) - contextual objects with a non-specific compatibility to the room category are in fact treated as non-informative regarding a specific action goal. Hence, we assume that low goal affinity contextual objects are perceived as part of the room category comparable to the compatibility condition in Wurm and Schubotz (2012). Indeed, in everyday life most of the objects that surround us are room-compatible but have a low goal affinity to the object manipulation we are about to perform. On the contrary, contextual objects with a high goal affinity are perceived as highly informative in such a way that a specific overarching action goal and hence a concrete action sequence can be generated, especially when presented close and right to the actor.

Conclusion

Our study shows that objects provide a meaningful context for observed actions, even when they are neither involved in the action nor relevant to the observer's task. Brain activation elicited by a contextual object varied as a function of the currently observed action, i.e., the contextual object's goal affinity. Moreover, when highly goal-related objects were presented close to the actor's preferred action space, additional brain activity signified efforts to integrate the contextual object into potentially upcoming action steps.

Acknowledgments

We would like to thank Monika Mertens for her great help in the data acquisition. Further we thank Sarah Fromme, Haris Orucevic, Daniel Westen and Daniel Kluger for assistance with the creation of the stimulus material. Finally, we want to thank Christiane Ahlheim and the workgroup AE Schubotz for fruitful discussions.

References

Aglioti, S.M., Cesari, P., Romani, M., Urgesi, C., 2008. Action anticipation and motor resonance in elite basketball players. Nat. Neurosci. 11 (9), 1109–1116.

Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H.B.M., Zilles, K., 1999. Broca's region revisited: cytoarchitexture & intersubject variability. J. Comp. Neurol. 412, 319–341.

Ashburner, J., Friston, K.J., 1997. Multimodal image coregistration and partitioning - a unified framework. Neuroimage 6, 209–217. https://doi.org/10.1006/nimg.1997.0290.

Aron, A.R., Robbins, T.W., Poldrack, R.A., 2004. Inhibition and the right inferior frontal cortex. Trends Cogn. 8, 170–177.

Balser, N., Lorey, B., Pilgramm, S., Naumann, T., Kindermann, S., Stark, R., et al., 2014a. The influence of expertise on brain activation of the action observation network during anticipation of tennis and volleyball serves. Front. Hum. Neurosci. 8, 35–47. https://doi.org/10.3389/fnhum.2014.00568.

Balser, N., Lorey, B., Pilgramm, S., Stark, R., Bischoff, M., Zentgraf, K., et al., 2014b. Prediction of human actions: expertise and task-related effects on neural activation of the action observation network. Hum. Brain Mapp. 35 (8), 4016–4034. https://doi.org/10.1002/hbm.22455.

Bar, M., Tootell, R.B.H., Schacter, D.L., Greve, D.N., Fischl, B., Mendola, J.D., Rosen, B.R., Dale, A.M., 2001. Cortical mechanisms specific to explicit visual object recognition. Neuron 29 (2), 529–535.

Bar, M., 2004. Visual objects in context. Nat. Rev. Neurosci. 5 (8), 617–629.

Boyce, S.J., Pollatsek, A., Rayner, K., 1989. Effect of background information on object identification. J. Exp. Psychol. Hum. Percept. Perform. 15 (3), 556–566.

Buxbaum, L.J., Kyle, K.M., Tang, K., Detre, J.A., 2006. Neural substrates of knowledge of hand postures for object grasping and functional object use: evidence from fMRI. Brain Res. 1117 (1), 175–185.

Buxbaum, L.J., Kalénine, S., 2010. Action knowledge, visuomotor activation, and embodiment in the two action systems. Ann. N. Y. Acad. Sci. 1191 (1), 201–218.

Bryden, P.J., Pryde, K.M., Roy, E.A., 2000. A performance measure of the degree of hand preference. Brain Cogn. 44, 402–414.

Calvo-Merino, B., Glaser, D.E., Grezes, J., Passingham, R.E., Haggard, P., 2005. Action observation and acquired motor skills: an fMRI study with expert dancers. Cereb. Cortex 15 (8), 1243–1249.

Caspers, S., Zilles, K., Laird, A.R., Eickhoff, S.B., 2010. ALE meta-analysis of action observation and imitation in the human brain. Neuroimage 50 (3), 1148–1167.

Chawla, D., Rees, K., Friston, J., 1999. The physiological basis of attentional modulation in extrastriate visual areas. Nat. Neurosci. 2 (7), 671–676. https://doi.org/10.1038/10230.

Creem-Regehr, S.H., 2009. Sensory-motor and cognitive functions of the human posterior parietal cortex involved in manual actions. Neurobiol. Learn. Mem. 91 (2), 166–171. https://doi.org/10.1016/j.nlm.2008.10.004.

Csibra, G., Gergerly, G., 2007. 'Obsessed with goals': functions and mechanisms of teleological interpretations of actions in humans. Acta Psychol. 124 (1), 60–78.

Eickhoff, S.B., Stephan, K.E., Mohlberg, H., Grefkes, C., Fink, G.R., Amunts, K., Zilles, K., 2005. A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. NeuroImage 25 (4), 1325–1335.

Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S.J., 1995. Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2 (4), 189–210.

Gozli, D.G., West, G.L., Pratt, J., 2012. Hand position alters vision by biasing processing through different visual pathways. Cognition 124 (2), 244–250.

Grefkes, C., Fink, G.R., 2005. The functional organization of the intraparietal sulcus in humans and monkeys. J. Anat. 207 (1), 3–17. https://doi.org/10.1111/j.1469-7580.2005.00426.x.

Hayes, S.M., Nadel, L., Ryan, L., 2007. The effect of scene context on episodic object recognition: parahippocampal cortex mediates memory encoding and retrieval success. Hippocampus 17 (9), 873–889.

Hampshire, A., Chamberlain, S.R., Monti, M.M., Duncan, J., Owen, A.M., 2010. The role of the right inferior frontal gyrus: inhibition and attentional control. Neuroimage 50 (3–3), 1313–1319. http://doi.org/10.1016/j.neuroimage.2009.12.109.

Hrkać, M., Wurm, M.F., Schubotz, R.I., 2014. Action observers implicitly expect actors to act goal- coherently, even if they do not: an fMRI study. Hum. Brain Mapp. 35 (5), 2178–2190. https://doi.org/10.1002/hbm.22319.

Hrkać, M., Wurm, M.F., Kühn, A.B., Schubotz, R.I., 2015. Objects mediate goal integration in ventrolateral prefrontal cortex during action observation. PloS one 10 (7), e0134316.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J.C., Rizzolatti, G., 2005. Grasping the intentions of others with one's own mirror neuron system. PLoS Biol. 3 (3), e79.

Johnson-Frey, S.H., 2004. The neural bases of complex tool use in humans. Trends Cognit. Sci. 8 (2), 71–78.

Kalénine, S., Shapiro, A.D., Flumini, A., Borghi, A.M., Buxbaum, L.J., 2013. Visual context modulates potentiation of grasp types during semantic object categorization. Psychonomic Bull. Rev. 21 (3), 645–651. https://doi.org/10.3758/s13423-013-0536-7.

Kamps, F.S., Julian, J.B., Kubilius, J., Kanwisher, N., Dilks, D.D., 2016. The occipital place area represents the local elements of scenes. NeuroImage 132, 417–424. https://doi.org/10.1016/j.neuroimage.2016.02.062.

Kilner, J.M., 2011. More than one pathway to action understanding. Trends Cognit. Sci. 15 (8), 352–357.

Lohmann, G., Müller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., et al., 2001. LIPSIA—a new software system for the evaluation of functional magnetic resonance images of the human brain. Comput. Med. Imaging Graph. 25 (6), 449–457.

Malcolm, G.L., Groen, I.I., Baker, C.I., 2016. Making sense of real-world scenes. Trends Cognit. Sci. 20 (11), 843–856.

Menon, V., Adleman, N.E., White, C.D., Glover, G.H., Reiss, A.L., 2001. Error-related brain activation during a Go/NoGo response inhibition task. Hum. Brain Mapp. 12, 131–143.

Ramsey, R., Cross, E.S., Hamilton, A.F.D.C., 2011. Eye can see what you want: posterior intraparietal sulcus encodes the object of an Actor's gaze. J. Cognit. Neurosci. 23 (11), 3400–3409.

Rezaee, M., Shojaee, M., Ghasemi, A., Moghaddam, A., Momeni, M., 2010. Which factors affect hand selection in Adults? Combined effects of ocular dominance, task demand and object location. J. Hum. Kinet. 26, 39–44.

Rubia, K., Smith, A.B., Brammer, M.J., Taylor, E., 2003. Right inferior prefrontal cortex mediates response inhibition while mesial prefrontal cortex is responsible for error detection. Neuroimage 20, 351–358.

Schubotz, R.I., Wurm, M.F., Wittmann, M.K., von Cramon, D.Y., 2014. Objects tell us what action we can expect: dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI. Front. Psychol. 5, 636. https://doi.org/10.3389/fpsyg.2014.00636.

Singh-Curry, V., Husain, M., 2009. The functional role of the inferior parietal lobe in the dorsal and ventral stream dichotomy. Neuropsychologia 47 (6), 1434–1448.

Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44 (1), 83–98.

Toney, A., Thomas, B.H., 2006. Considering reach in tangible and table top design. In: First IEEE International Workshop on Horizontal Interactive Human-computer Systems.

Tyler, L.K., Stamatakis, E.A., Bright, P., Acres, K., Abdallah, S., Rodd, J.M., Moss, H.E., 2004. Processing objects at different levels of specificity. J. Cognit. Neurosci. 16, 351–362.

Van Overwalle, F., Beatens, K., 2009. Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. Neuroimage 48, 564–584.

Wigget, A.J., Downing, P.E., 2011. Representation of action in occipito-temporal cortex. J. Cognit. Neurosci. 23 (7), 1765–1780. https://doi.org/10.1162/jocn.2010.21552.

Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited—Again. Neuroimage 2, 173–181.

Wurm, M.F., von Cramon, D.Y., Schubotz, R.I., 2011. Do we mind other minds when we mind other minds' actions? A functional magnetic resonance imaging study. Hum. Brain Mapp. 32 (12), 2141–2150. https://doi.org/10.1002/hbm.21176.

Wurm, M.F., Schubotz, R.I., 2012. Squeezing lemons in the bathroom: contextual information modulates action recognition. Neuroimage 59 (2), 1551–1559. https://doi.org/10.1016/j.neuroimage.- 2011.08.038.

Wurm, M.F., von Cramon, D.Y., Schubotz, R.I., 2012. The context-object-manipulation triad: cross talk during action perception revealed by fMRI. J. Cognit. Neurosci. 24 (7), 1548–1559.

Wurm, M.F., Hrkać, M., Morikawa, Y., Schubotz, R.I., 2014. Predicting goals in action episodes attenuates BOLD response in inferior frontal and occipitotemporal cortex. Behav. Brain Res. 274, 108–117. https://doi.org/10.1016/j.bbr.2014.07.053.

Wurm, M.F., Schubotz, R.I., 2017. What's she doing in the kitchen? Context helps when actions are hard to recognize. Psychonomic Bull. Rev. 24 (2), 503–509. Advance online publication. https://doi.org/10.3758/s13423-016-1108-4.

Wurm, M.F., Artemenko, C., Giuliani, D., Schubotz, R.I., 2017. Action at its place: contextual settings enhance action recognition in 4-8 year old children. Dev. Psychol. 53 (4), 662–671.

Zhang, S., Chiang-shan, R.L., 2012. Functional connectivity mapping of the human precuneus by resting state fMRI. Neuroimage 59 (4), 3548–3562.

Zimmermann, E., Schnier, F., Lappe, M., 2010. The contribution of scene context on change detection performance. Vis. Res. 50 (20), 2062–2068.

3.2 Study 2: Is it mine or is it yours? Proximal unused objects invite third-person perspective action observers to create their own action space

Nadiya El-Sourani, Ima Trempler, Moritz F. Wurm, Gereon R. Fink, & Ricarda I. Schubotz (2018)

Manuscript is under review

Title: Is it mine or is it yours? Proximal unused objects invite third-person perspective action observers to create their own action space

Authors: Nadiya El-Sourani[1,2], Ima Trempler[1], Moritz F. Wurm[3], Gereon R. Fink[2,4], Ricarda I. Schubotz[1,2]

[1]Department of Psychology, Westfälische Wilhelms-Universität, 48149 Münster, Germany
[2]Department of Neurology, University Hospital Cologne and University of Cologne, 50937 Cologne, Germany
[3]Center for Mind/Brain Sciences (CIMeC), University of Trento, Rovereto, TN, Italy
[4]Institute of Neuroscience and Medicine (INM3), Cognitive Neuroscience, Research Centre Jülich, 52425 Jülich, Germany

Corresponding author:
Nadiya El-Sourani
Fliednerstr. 21,
48149 Münster
n_elso02@uni-muenster.de

Abstract

Acting in a social world not only involves understanding the intention of others but at the same time considering one's ability and opportunity to act. The present study investigated whether in a third-person perceptive (3pp) action observation paradigm participants establish their own action space by perceiving unused contextual objects (COs) in a scene not only regarding the observed action but also with reference to themselves. Participants were presented with action videos that contained COs at varying locations on a table. We conducted a behavioural study (N = 24) assessing reachability scores of the COs from an observing perspective. In a separate sample (N = 21), we employed fMRI while participants were presented with the same action videos as in the behavioural study. Reachability scores of the behavioural study were used to model the BOLD responses in the fMRI data analysis. Reachability scores of the COs assessed by the behavioural study revealed that in addition to the actor's space, participants created their own action space in the scene. The fMRI analysis revealed significant activations in premotor, cingulate, inferior parietal, and insular cortex parametrically increasing with reachability of the presented CO for the observer. Results suggest that depending on where COs are located, even merely observing an action from 3pp can trigger the involvement of an observer-referenced perspective: proximal located unused COs furnish action observers with their own action possibilities.

Keywords

Action observation, fMRI, perspective, insula, object perception

Introduction

In everyday life, we frequently observe other people acting and easily infer their goals (cf. Van Overwalle & Beatens, 2009; Caspers, Zilles, Laird, & Eickhoff, 2010). This inference is significantly influenced by the contextual setting in which the action takes place (Costantini, Ambrosini, Sinigaglia, Committeri, 2010; Borghi, Flumini, Natraj, & Wheaton, 2011; Wurm & Schubotz, 2012; Hrkać, Wurm, & Schubotz, 2014; El-Sourani, Wurm, Trempler, Fink, & Schubotz, 2018). Context provides information about the (Wurm & Schubotz, 2012), the actor (Hrkac et al., 2014), and the location of objects (Costantini et al., 2010; El-Sourani et al., 2018). We recently showed that *contextual objects (COs)*, i.e., objects which are part of an observed scene but which are not yet integrated into the observed action, shape brain activity during action observation although being task-irrelevant (El-Sourani et al., 2018). By modulating COs' semantic (*Goal Affinity*) and spatial (*Location Ergonomics*) relation to the observed action, we found that COs were spontaneously processed in reference to the current goal of the agent. Brain activation in action observation network (AON) nodes implied that the stronger the semantic relation of a CO to the observed action, the more a specific action sequence and hence a specific action outcome could be anticipated. This effect was further increased when the CO was presented close to the agent's right hand (*close-right*, hereafter; cf. *Figure 1*). This area is where right-handed actors preferred to place objects that they planned to reach for subsequent manipulation, according to own pilot data (unpublished). Specifically, we took inferior frontal gyrus (IFG) engagement to show that a strong spatial relation in combination with a strong semantic relation to the observed action leads action observers to perceive the contextual object as a potential future target of the actor (Wurm & Schubotz, 2012; Hrkac et al., 2014).

Building on these findings, we here aimed to extend this idea and conjectured that acting in our social world comprises situations which not only involve the understanding of the other person's intention to act but at the very same time one's ability and opportunity to act and interact with one another and the environment. Previous explorative analyses showed us that the observer's perspective might play a crucial role in action interpretation in the present paradigm: We found shortest reaction times for the recognition of actions including highly semantically related COs that were either close-right to the actress or far-left to the actress, the latter being equivalent to close-right to the observer (180 degrees rotated). The most parsimonious explanation for these findings is that observers are fastest in processing those COs that are particularly easy to reach either for the actor or the observer (cf. Costantini, Committeri, & Sinigaglia, 2011), According to the concept of peripersonal space (Rizzolatti, Fadiga, Fogassi, & Gallese, 1997; Bufacchi & Iannetti, 2018), it is conceivable that in our paradigm, action observers represented the space proximal to the actor, and objects within this space, as referenced to the actor (*actor-referenced map*) (cf. El-Sourani et al., 2018). Likewise, they most probably represented the space proximal to the 3pp camera position – the virtual location of the observer – as belonging to their own peripersonal space and thus as being referenced to themselves, including objects therein (*observer-referenced map*). The overlap between the actor's and observer's peripersonal space, the *shared action space* (Pezzulo, Iodice, Ferraina, & Kessler, 2013), is convenient to reach for both. We reasoned that COs located either in the observer-referenced map or in the shared action space should trigger brain activity associated with observer-involvement, as both incorporate the engagement of the observer's perspective (*Figure 3*, heatmaps).

Against this backdrop, we conducted the present follow-up study to explore (i) whether 3pp action observation can trigger an involvement of the observer's perspective (*observer-referenced)* in addition to an *actor-referenced* perspective that we witnessed in the previous study (El-Sourani et al., 2018), and if so, (ii) how this effect is represented in the brain.

Typically, in 3pp action observation paradigms, it is not possible to account for the observer's perspective, as there are no means to measure it. Importantly, the introduction of COs enabled us to take the observer's perspective into account. In a first step, we conducted a behavioural study (in a separate group of N = 24 subjects) assessing reachability scores for COs from the observer's perspective. Participants were presented with the original action videos from the 3pp (El-Sourani et al., 2018), with a camera perspective that led to the impression one would be sitting at the table opposite to the actress. We asked participants how reachable COs would be for them as observers. Subsequently, these scores were used as a regressor to model the BOLD response in a re-analysis of the fMRI dataset of El-Sourani et al. (2018).

In El-Sourani et al. (2018) we already showed that COs in the preferred action zone of the actress were processed concerning the actress and the observed action. In contrast, in the present study, depending on their localization, some COs were hypothesized to trigger an observer-referenced action representation rather than an actor-referenced one. We expected that the more convenient the spatial location of the CO for a hypothetical grasping of the observer, the stronger the responses in brain areas representing this observer perspective, specifically the AON, which mediates the goal-representation governing our behavioural space (Metzinger & Gallese, 2003). Among the AON components, left anterior intraparietal lobule (aIPL) and ventral premotor cortex (PMv) were previously found to positively co-vary with the number of object-implied action possibilities (Schubotz, Wurm, Wittmann, & von Cramon, 2014), and with the number of sub-goals required to achieve a final overarching goal (Newmann-Norlund, Noordzij, Meulenbroek, & Bekkering, 2007). It is conceivable that these areas are especially engaged with increasing observer involvement, as the CO might even imply additional actions and action goals for the observer, irrespective of the observed action.

Moreover, we assumed that the involvement of the observer-referenced action space also entails processes related to agency and self-other distinction. These effects could be particularly pronounced in the shared action space where actress and observer could cooperate or compete for the presented COs. Previous findings indicate that on a neural level, the insula and the right IPL may be crucial in distinguishing self from other (Newman-Norlund et al., 2007). Specifically, activation of the right IPL was negatively correlated with a sense of agency (Blakemore, Oakley, & Frith, 2003; Blanke & Arzy, 2005; Farrer et al., 2003), while activation in the insula was positively correlated with a feeling of control (Farrer & Frith, 2002; Farrer et al., 2003). In a similar vein, the insula has been related to issues of agency in joint action tasks (Newman-Norlund et al., 2007). Accordingly, we expected increased activity in the insula to accompany AON activity for COs at observer-convenient positions.

2. Materials and Methods

In this follow-up study the identical participant sample, stimuli and task were used as in El-Sourani et al. (2018), and will therefore only briefly be described in the below section. For a detailed description please see El-Sourani et al. (2018).

2.1 Participants

Twenty-one right-handed subjects (12 females; 24.00 ± 3.25 years old; range, 20-30 years) with normal or corrected-to-normal vision participated in this study. Participants neither reported a history of neurological or psychiatric disorders nor of substance abuse. The study protocol was conducted following the ethical standards of the Declaration of Helsinki and approved by the Ethics Committee of the University of Münster. Before participation, each subject submitted a signed informed consent. Subjects either received course credits or reimbursement after participating.

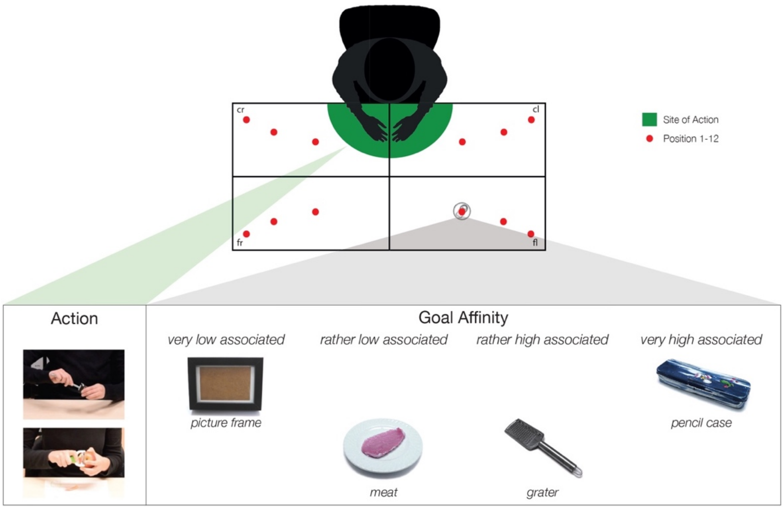


Figure 1. Illustration of the experimental factors: Red dots refer to location ergonomics (positions 1-12). Further positions are subdivided into quadrants from the actor's perspective: close-right (cr), close-left (cl), far-right (fr) and far-left (fl). Goal affinity levels of contextual objects for distinct actions are depicted in the box**.** Retrieved from El-Sourani et al. (2018).

2.2 Stimuli

During the scanning session, participants were presented with 360 action videos, of which 280 included an additional contextual object positioned in front of the actress on the table. Each action video depicted a single object-directed action with two target objects and with two varying factors regarding the additional contextual object *(Figure 1): Goal Affinity,* depicting the semantic relation to the observed action, and *Location Ergonomics,* depicting the spatial relation to the observed action (*Figure 1*).

Actions were performed by the same actress throughout the experiment and were filmed from a third person perspective (3pp). Actions were performed in their typical setting (either a kitchen or an office), resulting in compatibility between action and context. Videos were piloted regarding their recognizability, and only those videos were employed, which were recognized by all participants.

Each trial (6 s) started with a video clip or a question (3 s), which was followed by a fixation phase (3 s). A variable jitter (500, 1000, 1500 ms) was included after the fixation phase. In 5% (i.e., N = 18) of the trials a null event was implemented (6 s), which required participants to fixate a fixation cross (*Figure 2*). During the scanning session, action trials were intermixed with question trials (20%), that is, 72 written action descriptions referring to these actions were presented. Videos were presented in a pseudo-randomized fashion to avoid direct repetition of action, goal affinity, and location of the contextual object. All factor levels were presented in an evenly distributed manner. Stimuli were presented using Presentation 13.1 (Neurobehavioral Systems, San Francisco, CA, USA).

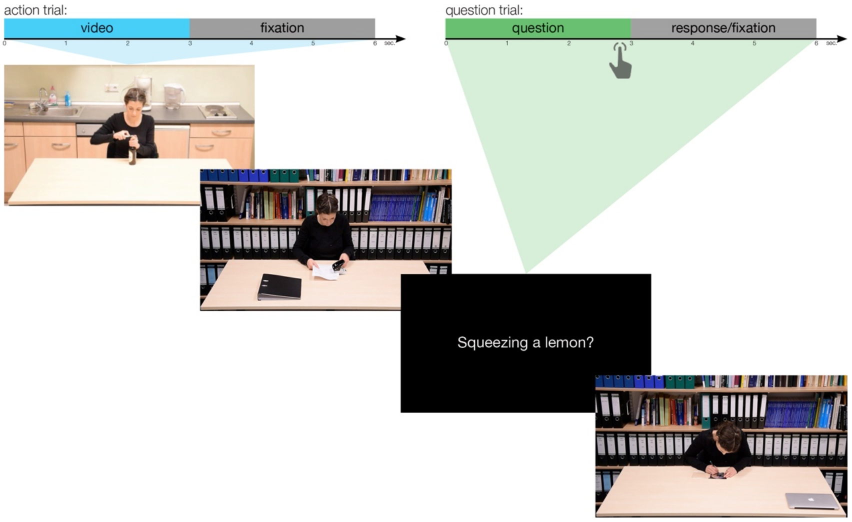


Figure 2. Schematic illustration of the task. Action trials consisted of an action video (3s) and a fixation phase (3s). Question trials consisted of a question regarding the preceding video trial (n-1), followed by a response and fixation phase. Retrieved from El-Sourani et al. (2018).

2.3 Task

Participants were instructed to observe the video clips and to respond to the action description (20%) that either referred to the content of the preceding video (50%) or not (50%). Participants had to either accept or reject the action description using a two-button response box. This task was chosen to ensure the participants' attention throughout the experiment (Wurm & Schubotz, 2012; Hrkać et al., 2015).

2.4 Pilot study on perceived observer and actor space

As we were interested in whether participants were actively engaging in the action scene, we used the final video selection for an additional behavioral study (*Pilot "Observer"*). We asked 24 right-handed individuals to rate on a 5-point Likert Scale the ease with which they thought they could reach and grasp the contextual object as active counterparts in the videos. This was done in order to establish a reachability-by-observer map.

Moreover, to represent the *actor-referenced map* (*Pilot "Actor"*), data from the original pilot study to determine Location Ergonomics for the actor were used from El-Sourani et.al (2018). Here, mean values for the 12 selected locations on the table according to the final video selection of El-Sourani et al. (2018) were used (*Figure 1*). Ratings of the Pilot "Observer" (depicted in *Figure 4 A/B)* indicated that action observers perceived the space on the table in a 180-degree rotated fashion, for example, far left positions become close-right (*Figure 3 A/B,* heatmaps*)*.

Data were analyzed using the statistic software package R (R Foundation for Statistical Computing, Vienna, Austria). In order to compare the results of actor-referenced and observer-referenced maps, data were z-standardized before statistical analysis. To test for differences in reachability of the COs depending on their location within and between studies, we conducted repeated measures ANOVAs. The standardized mean values per position were used as parametric regressors for the fMRI analysis. For the *observer-referenced map,* we used the standardized data set of *Pilot "Observer"* per CO location. In order to determine the regressor for the *shared action space,* we added up the values of the observer-map and the actor-map (derived from the original Location Ergonomics Pilot Study in El-Sourani et al. (2018)). Accordingly, the highest values depict locations which are most convenient to reach for both the actor and the observer. *Figure 3* provides an illustration of these maps.

2.5 fMRI Image Acquisition

Imaging was performed on a 3 Tesla Siemens Magnetom Prisma MR tomograph using a 20channel head coil. Participants were located in a supine position on the scanner bed with their right index and middle fingers positioned on the appropriate response buttons of the response box. The head and arms were tightly fixated with form-fitting cushions to minimize motion. Furthermore, participants were given earplugs and headphones to attenuate the scanner noise. Whole-brain functional images were acquired using a gradient T2*-weighted single-shot echo-planar imaging (EPI) sequence sensitive to blood oxygenation level dependent (BOLD) contrast (64 x 64 data acquisition matrix, 192 mm field of view, 90° flip angle, TR = 2000 ms, TE = 30 ms). Each volume consisted of thirty adjacent axial slices with a slice thickness of 4 mm and a gap of 1 mm, which resulted in a voxel size of 3 x 3 x 5 mm. Images were acquired in ascending order along the AC-PC plane to provide a whole-brain coverage. After functional imaging, structural data were acquired for each participant using a standard Siemens 3D T1-weighted MPRAGE sequence for detailed reconstruction of anatomy with isotropic voxels (1 x 1 x 1 mm) in a 256 mm field of view (256 x 256 matrix, 192 slices, TR = 2130, TE = 2.28).

In order to present the stimuli during the scanner session, a 45° mirror was fixated on the top of the head coil. A video-projector projected the experiment on a screen that was positioned behind the subject's head so that participants could see the stimuli via the mirror. The mirror was adjusted for each participant to provide a perfect view (center of the field of vision). In a pilot study, we controlled for recognizability of actions and contextual objects in a similar experimental setting. Only action videos in which the action and the contextual object could be identified by at least 95% of the participants were employed in the study.

2.6 fMRI Data Analysis

2.6.1 fMRI data preprocessing

Brain image preprocessing and statistical analyses were conducted using SPM12 (Wellcome Department of Imaging Neuroscience, London, UK; see: http://www.fil.ion.ucl.ac.uk/spm/software/spm12/). Functional images were slice-timed to the middle slice to correct for differences in slice acquisition time. To correct for three-dimensional motion individual functional MR (EPI) images were realigned to the mean epi image and further motion correction estimates were inspected visually. The anatomical scan was co-registered (rigid body transformation) to the mean functional image. Each subject's co-registered anatomical scan was segmented into native space tissue components. The parameters obtained were applied to normalize the subject's functional scans to the template brain MNI space. Finally, the normalized images were spatially smoothed using a Gaussian kernel of 8mm3 full width at half-maximum (FWHM). A 128 s temporal high-pass filter was applied to the data to remove low-frequency noise.

2.6.2 Design specification

The statistical evaluation was based on a least-squares estimation using the general linear model (GLM) for serially autocorrelated observations (Friston et al., 1995; Worsley & Friston, 1995). The design matrix was generated with delta functions and convolved with a canonical hemodynamic response function. The subject-specific six rigid-body transformations obtained from residual motion correction were included as covariates of no interest. Activations were analyzed time-locked to the onset of the videos and the analyzed epoch comprised the full duration (3s) of the presented videos and the reaction time in question trials (max. 3s). Our GLM contained six regressors in total: One predictor containing all videos including an object (VidOb), one predictor containing all videos without a contextual object (noOb), one regressor for question trials (Que), and parametric regressors for the experimental manipulation. The three parametric regressors depicted the standardized mean values of the respective behavioral maps per position for videos containing a contextual object: *observer-referenced (parObs)*, *shared space overlay (parShared),* and *actor-referenced (parAc)*. Here, parObs and parShared served as regressors of interest, whereas parAc was included as a regressor of no interest to control for the effect of the actor space. Resting periods (null events) were not modeled and served as the implicit baseline (Pernet, 2014).

2.6.3 Group analysis

To obtain group statistics, the resulting contrast images of all participants for our regressors of interest involving the observer's perspective, i.e., *parObs* and *parShared*, were entered into a second level random-effects analysis using a one-sample t-test across participants to test for significant deviation from zero. We then corrected for multiple comparisons using the false discovery rate (FDR) method with $p < 0.01$. Activation maps were superimposed on a ch2better.nii.gz atlas using MRIcron software (https://www.nitrc.org/projects/mricron).

3. Results

3.1. Behavioral results

3.1.1 Perceived reachability by the observer

In a first step, mean rating scores for the observer-referenced hypothetical reachability were calculated and separately aggregated according to four quadrants on the table: close-right, close-left, far-right and far-left. Descriptive results are shown in *Figure 4A*. As evident from the heatmaps (*Figure 3*), observer-referenced scores were roughly inverse to the actor-referenced scores. As we were interested in exploring the potential role of an observer, we next conducted a repeated measures ANOVA with the four-level within-subject factor CO Location (close-right$_{\text{observer}}$, close-left$_{\text{observer}}$, far-right$_{\text{observer}}$, far-left$_{\text{observer}}$) for the observer-referenced reachability rating. This effect was significant (F(3, 69) = 40.3, $p < 0.001$), showing that for an observer the reachability of the CO differed significantly depending on where it was located.

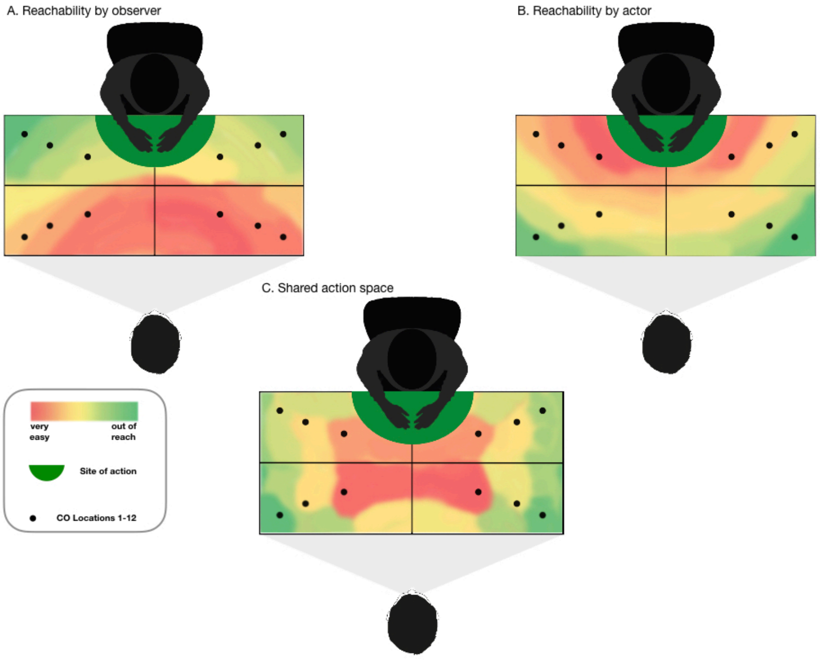


Figure 3. Heat-maps showing the mean reachability ratings according to 12 positions on the table. Dark red indicates very easy to reach locations, while dark green indicates locations out of reach. A. Reachability-by-observer B. Reachability-by-actor C. Shared action space, an overlay of both ratings.

3.1.2 Comparison between reachability by observer versus reachability by actor

To test for differences between observer-referenced and actor-referenced reachability, we conducted a 2 x 2 between-subjects factorial ANOVA on rating scores with the factors REFERENCE (actor, observer) and CO LOCATION (far, close). The interaction of these factors was significant ($F(1,112) = 29.83$, $p < 0.001$; *Figure 4C*).

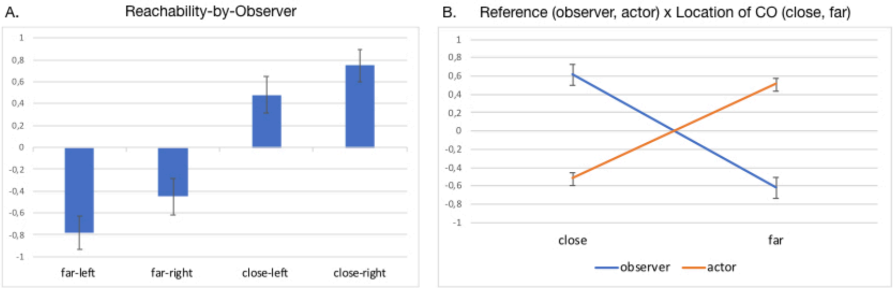


Figure 4. Behavioral results of CO reachability. (A) Mean reachability-by-observer scores, including SE values per condition. (B) Interaction of REFERENCE (actor/observer) by CO LOCATION (close/far).

3.2. fMRI results

To investigate evidence for a potential involvement of an observer-referenced processing perspective, we investigated the parametric effect of the observer-referenced map of CO locations (FDR-correction at $p < 0.01$). The observer-referenced map revealed activation in several areas to increase with increasing reachability-by-observer, including the left aIPS, the left insula, bilateral PMv, PMd, (pre)supplementary motor area (SMA), and middle cingulate cortex (MCC), with an overall more pronounced effect in the left hemisphere (Figure 5; Table 1). The reverse effect did not show any significant activation patterns after correction.

In order to test whether brain areas with a preference for the area which is convenient to reach for both the observer and the actor, we additionally tested for the parametric effect of the shared action space. This map was generated by adding up the z-standardized values of the actor-referenced map and the observer-referenced map and thus highlights areas that are best reachable for both the actor and the observer, as rated by the observer. After FDR correcting at $p < 0.05$, no significant activations were revealed by this analysis.

.

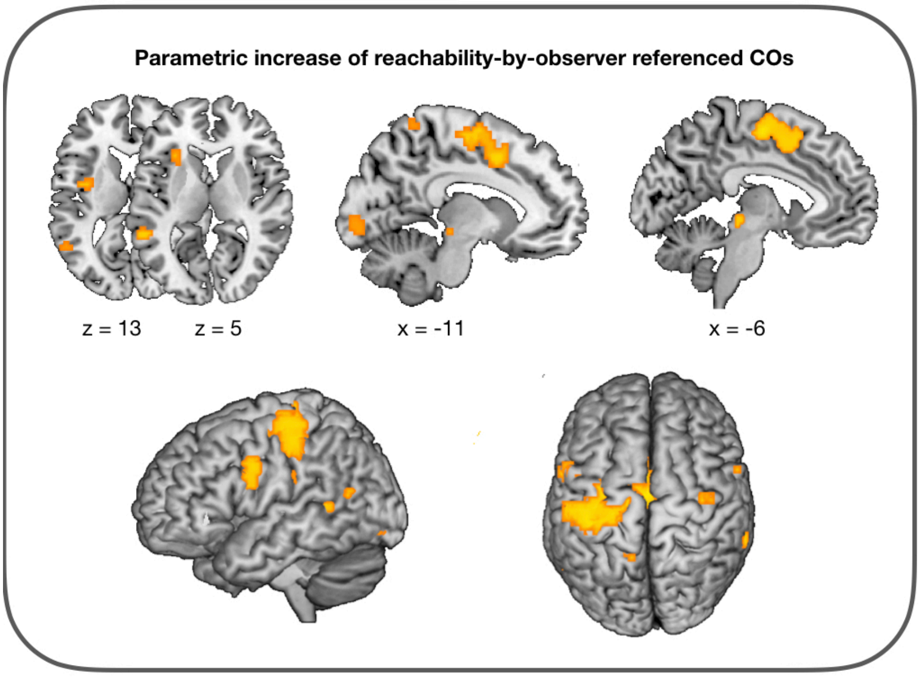


Figure 5. Parametric effect of the observer-referenced CO map, at $p < 0.01$ FDR corrected. The parametric effect of reachability-by-observer revealed increased activation in left aIPS and left insula, left pMTG in addition to bilateral (pre)-SMA, MCC, PMd, and PMv.

Table 1. fMRI activations for the observer-referenced map (parObs)

			MNI Coordinates			
Region	*Hemisphere*	*BA*	*x*	*y*	*z*	*t-scores*
SMA	l	6	-3	-10	56	6.44
	r	6	1	-7	55	5.64
pre-SMA	l	6	-4	6	52	4.91
aIPS	l	-	-49	-24	42	5.31
PMd	l	6	-33	-10	58	4.81
	r	6	39	-13	59	4.95
PMv	l	6	-57	5	32	5.52
	r	6	60	8	35	4.31
IFG	l	44	-51	11	23	4.92
MCC	l	32	-9	9	42	4.62
	r	32	11	20	38	4.34
Insula	l	48	-27	20	5	4.58
		48	-36	-4	14	4.73
Visual association area/ lingual gyrus**	l	18	-15	-91	-4	6.1
Supramarginal gyrus/ pSTS	l	40/22	-57	-25	23	4.7
	r	22/40	66	-40	26	4.78
(p)MTG	l	21	-54	-58	14	4.70
	l	21/22	-57	-43	5	4.88
MI	l	4	-42	-19	53	6.29

Cerebellum	r	-	24	-46	-25	7.18
	r	-	21	-58	-46	4.67

r, right; l, left; x, y, z, MNI coordinates of peak voxel activation; **ranging into left fusiform gyrus; PMv, ventral premotor cortex; PMd, dorsal premotor cortex; IFG, inferior frontal gyrus; pMTG, posterior middle temporal gyrus; SMA, supplementary motor area; pSTS, posterior superior temporal sulcus; $p < 0.01$ FDR-corrected for multiple comparison

4. Discussion

Observers perceive unused contextual objects (COs) in an action scene as goal-referenced to the actor's current action, especially when they are easy to reach by the actor (El-Sourani et al., 2018). The present fMRI study investigated whether participants concurrently, merely by watching an action video from 3pp, also establish their own action space in the presented scene. To this end, subjects rated the location of COs concerning their reachability by themselves (i.e., the observers). Using these rating scores to analyse BOLD responses during action observation of a separate sample in a fMRI study, we found that activity in premotor, inferior parietal, cingulate, and insula cortex parametrically increased with the CO location's (virtual) reachability for the observer. We propose that depending on where COs are located observers grant either the actor or themselves with potential action possibilities.

Our findings showed that subjects were able to represent COs' reachability both concerning themselves and the actor. Thus, close-right$_{\text{observer}}$ located COs were particularly easy to reach for the observer, but difficult to reach for the actor, comparable to far-left$_{\text{actor}}$ located COs. As evidenced by previous pilot studies, this effect appears unlikely to be due to mere visual characteristics, such as better visibility of the COs when they were located close to the observer, only. The object size, and thus visibility and recognisability of the objects, differed only subtly between close and far positions.

As hypothesized for the parametric effect of reachability-by-observer, we found activation at several AON sites. We specifically expected the aIPS in combination with the PMv to increase with reachability-by-observer, as activation of these areas co-varies both with the number of object-implied action possibilities (Schubotz et al., 2014) and of sub-goals required to achieve a final overarching goal (Newmann-Norlund et al., 2007). About the present finding, we suggest that COs in the observer's operational space provided independent action possibilities such as handing over the CO to the actor or manipulating it on one's own. Speaking in favour of this interpretation, we also found preSMA/SMA and mid-cingulate cortex (MCC) activity to increase with COs' increasing reachability-by-observer. SMA supports the execution of sequential actions and bimanual movements, while the pre-SMA has been associated with response selection as well as response inhibition (Chouinard & Paus, 2010; Nachev, 2008). SMA in combination with left lateral premotor cortex (Filimon, 2010; Fagg & Arbib, 1998; Rizzolatti & Luppino, 2001; Brochier & Umiltà, 2007), as found in the present study, suggests that participants imagined or planned to initiate a right-hand reach and grasp towards an object. Thus, the more comfortable a CO could be reached by the participants, the more consistently their brains engaged in such a subliminal imagined reach and grasp action towards the CO. Also, the recruitment of the pre-SMA might point to inhibition of this motor initiation (Chen et al., 2009). This interpretation is supported by the fact that we found preSMA/SMA activity to coincide with increased engagement of the adjacent MCC (Vogt, Berger, & Derbyshire, 2003; Vogt, 2005; Medford & Critchley, 2010).

Neuroimaging studies have attributed activation of the cingulate cortex to a wide range of attentional control functions, such as conflict monitoring, motor control, and motor inhibitory processes (Rubia et al., 2001; Brandeis, Brandys, & Yehuda, 1989; Botvinick, Braver, Barch, Carter, & Cohen, 2001; Hoffstadter et al., 2014). The same is true for increased insular activity (Kana, Keller, Minshew, & Just, 2007), and together with the anterior cingulate cortex (ACC), the MCC and insula have been suggested to operate as a (response) inhibition network (Kana et al., 2007). Other evidence of a direct functional relationship comes from a resting state connectivity study. Here, Taylor and colleagues (2009) proposed two distinct insula-cingulate systems, with anterior insula (aI) showing functional connectivity with ACC and MCC, and mid/posterior insula (m/pI) being connected with only the MCC. The authors suggested that the former network is involved in emotional salience monitoring, whereas the latter serves general environmental monitoring, response selection, and skeletomotor body orientation. Since we found increased engagement of both, the aI and pI, the observed pI-MCC activation suggests inhibition of motor initiation by the observer, necessitated by observer-conveniently located COs.

Finally, due to its role in autonomic control, interoceptive awareness (Critchley, Wiens, Rotshtein, Ohman, & Dolan, 2004), and distinguishing the self from others (Farrer & Frith, 2002; Farrer et al., 2003; Newman-Norlund et al., 2007), we hypothesized the insula to be significantly activated for observer-convenient located COs. This was indeed the case. Thus, we take (anterior) insular engagement as an indication for an involvement of the observer in the present scene. We argue that the more comfortable the CO's location concerning the observer’s potential action, the stronger their involvement apparent by the CO's afforded additional action possibilities. In support of our line of argumentation, Ruby & Decety (2001) found a comparable pattern of activation including inferior parietal, SMA, aI, precentral gyrus in the left hemisphere, and right cerebellum when participants simulated an action for themselves (first-person perspective) in a task where they either had to mentally simulate an action performed by themselves or by another individual.

Interestingly, we only found significant brain responses for reachability-by-observer, but no such effects for the shared action space. This finding is particularly interesting against the backdrop of the precursor study where we found, based on an analysis of the same data set, that COs are preferably processed when they are semantically related to the observed action and, crucially, when they are located close and right to the observed actor, that is, in the preferred actor space (El-Sourani et al., 2018). Together with these findings, the present results show more specifically that observers perceive not only the spatial and goal-relation between the CO and the observed action but at the same time also their action options at the level of COs' reachability. Thus, observers appear to separate the table into two operational spaces: one for the actor (El-Sourani et al., 2018) and one for the observer. Thereby, it seems as if objects in the shared action space are neither processed concerning the actor nor the observer. Future studies are warranted that explicitly investigate potential representations of a shared action space for joint actions of an observed actor and the observer.

We acknowledge that our results might be specific to the present paradigm and its contextual constraints and, thus, may not be generalizable to all 3pp action observation paradigms. Only by the introduction of COs, it was possible to account for observer-referenced processing. We propose that action observers carried their own action possibilities into the observed scene as their task was to watch and understand the observed action. However, we cannot completely rule out the possibility that the present results can solely be traced to the CO’s location convenience, independent of the observed action. Thus, future studies are required to investigate whether the observed action in addition to the contextual object is necessary to elicit the brain activations elicited.

In sum, the current findings suggest that depending on an observed action’s context, e.g., unused objects in the scene, action observers carry their action options into an observed scene. While previous analyses of the same data set showed that observers processed the COs with reference to the actor and the observed action (El-Sourani et al., 2018), the present study revealed that action observers simultaneously processed COs regarding their own action space and action possibilities.

5. Conclusion

Our study assessed the potential involvement of observer-referenced processing in a 3pp action observation paradigm. Results point out that the contextual setting of an observed action can trigger the involvement of the observer’s referenced perspective. Thereby, observers appear to be furnished with their action options triggered by objects located within the observer’s potential action space.

Acknowledgments

We thank Monika Mertens for her great help with data acquisition. Further, we thank Katharina Thiel for her assistance in behavioral data analysis and Haris Orucevic for his assistance with the creation of the figures.

Literature

Blakemore, S. J., Oakley, D. A., & Frith, C. D. (2003). Delusions of alien control in the normal brain. *Neuropsychologia*, 41, 1058-1076.

Blanke, O., & Arzy, S. (2005). The out-of-body experience. Distributed self-processing at the temporoparietal junction. *The Neuroscientist*, 11, 16-24.

Borghi, A. M., Flumini, A., Natraj, N., & Wheaton, L. A. (2012). One hand, two objects: emergence of affordance in contexts. *Brain Cogn.* 80, 64–73.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D., (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624-652.

Brandeis, R., Brandys, Y., & Yehuda, S. (1989). The use of the Morris water maze in the study of memory and learning. *International Journal of Neuroscience*, 48(1-2), 29-69.

Brochier, T., & Umiltà, M. A. (2007). Cortical control of grasp in non-human primates. *Current opinion in neurobiology*, 17(6), 637-643.

Bufacchi, R. J., & Iannetti, G. D. (2018). An action field theory of peripersonal space. *Trends in cognitive sciences*.

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B., (2010). ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage*, 50(3), 1148-1167.

Chen, C. Y., Muggleton, N. G., Tzeng, O. J., Hung, D. L., & Juan, C. H. (2009). Control of prepotent responses by the superior medial frontal cortex. *Neuroimage* 44, 537–545.

Chouinard, P. A, & Paus, T. (2010). What have we learned from"perturbing" the human cortical motor system with transcranial magnetic stimulation? *Frontiers in Human Neuroscience*. 4, 173.

Costantini, M., Ambrosini, E., Tieri, G., Sinigaglia, C., & Committeri, G. (2010). Where does an object trigger an action? An investigation about affordances in space. *Exp. Brain Res.* 207, 95–103.

Costantini, M., Committeri, G., and Sinigaglia, C. (2011). Ready both to your and to my hands: mapping the action space of others. *PLoS ONE* 6:e17923. doi: 10.1371/journal.pone.0017923

Critchley, H. D., Wiens, S., Rotshtein, P., Ohman, A., & Dolan, R. J., (2004) Neural systems supporting interoceptive awareness. *Nat Neurosci,* 7, 189 –195.

Decety, J., & Sommerville, J. A. (2003). Shared representations between self and other: a social cognitive neuroscience view. *Trends in cognitive sciences*, 7(12), 527-533.

El-Sourani, N. Wurm, M. F., Trempler, I., Fink, G. R., & Schubotz, R. I. (2018). Making sense of objects lying around: How contextual objects shape brain activity during action observation. *Neuroimage*, 167, 429-437.

Fagg, A. & Arbib, M. (1998). Modeling parietal-premotor interactions in primate control of grasping. *Neural Netw*, *11(7-8)*, 1277–1303.

Farrer, C., Franck, N., Georgieff, N., Frith, C. D., Decety, J., & Jeannerod, M. (2003). Modulating the experience of agency: A positron emission tomography study. *NeuroImage*, 18, 324-333.

Farrer, C., & Frith, C. D. (2002). Experiencing oneself vs. another person as being the cause of an action: The neural correlates of the experience of agency. *NeuroImage*, 15, 596-603.

Filimon, F. (2010). Human cortical control of hand movements: parietofrontal networks for reaching, grasping and pointing. *Neuroscientist* 16, 388–407.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4), 189–210.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, 41(10-11), 1409-1422.

Hoffstaedter, F., Grefkes, C., Caspers, S., Roski, C., Palomero-Gallagher, N., Laird, A. R., & Eickhoff, S. B. (2014). The role of anterior midcingulate cortex in cognitive motor control: evidence from functional connectivity analyses. *Hum. Brain Mapp.* 35 (6), 2741–2753.

Hrkać, M., Wurm, M. F., & Schubotz, R. I. (2013). Action observers implicitly expect actors to act goal- coherently, even if they do not: An fMRI study. *Human Brain Mapping*, 35(5), 2178-2190.

Hrkać, M., Wurm, M. F., Kühn, A. B., & Schubotz, R. I. (2015). Objects Mediate Goal Integration in Ventromedial Prefrontal Cortex during Action Observation. *PLOS ONE*, 10(9): e0137813.

Kana, R. K., Keller, T. A., Minshew, N. J., & Just, M. A. (2007). Inhibitory control in high-functioning autism: decreased activation and underconnectivity in inhibition networks. *Biological psychiatry*, 62(3), 198-206.

Medford, N., & Critchley, H. D. (2010).Conjoint activity of anterior insular and anterior cingulate cortex: awareness and response, *Brain Struct. Funct,* 214, 535-549.

Metzinger, T., & Gallese, V., (2003). *Consciousness and Cognition*, 12(4), 549-571.

Nachev, P., Kennard, C., & Husain, M. (2008). Functional role of the supplementary and pre-supplementary motor areas. *Nat Rev Neurosci,* 9, 856–869.

Newman-Norlund, R. D., Noordzij, M. L., Meulenbroek, R. G. J., & Bekkering, J. (2007). Exploring the brain basis of joint action: co-ordination of actions, goals and intentions.

Soc. Neurosci., 2, 48–65.

Pernet, C. R. (2014). Misconceptions in the use of the General Linear Model applied to functional MRI: a tutorial for junior neuro-imagers. *Frontiers in neuroscience, 8*, 1.

Pezzulo, G., Iodice, P., Ferraina, S., & Kessler, K. (2013). Shared action spaces: a basis function framework for social recalibration of sensorimotor representations supporting joint action. Fron. Hum. Neurosci., 7, 800.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1997). The space around us. *Science*, *277*(5323), 190-191

Rizzolatti, G., & Luppino, G. (2001).The cortical motor system. *Neuron*, 3, 889–901.

Rubia, K., Russell, T., Overmeyer, S., Brammer, M. J., Bullmore, E. T., Sharma, T., Simmons, A., Williams, S. C. R., Giampietro, V., Andrew, C. M., & Taylor, E. (2001). Mapping motor inhibition: conjunctive brain activations across different versions of go/no-go and stop tasks. *Neuroimage*, *13*(2), 250-261.

Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat. Neurosci*. 4, 546–50.

Schubotz, R. I., Wurm, M. F., Wittmann, M., & von Cramon, D. Y. (2014). Objects tell us what action we can expect: Dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI. *Frontiers in Psychology*, 5, 636.

Taylor, K. S., Seminowicz, D. A., & Davis, K. D. (2009). Two systems of resting state connectivity between the insula and cingulate cortex. *Hum. Brain Mapp*., 2731-2745.

Van Overwalle, F., & Beatens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, 48, 564-584.

Vogt, B. A., Berger, G. R., & Derbyshire S.W. (2003). Structural and functional dichotomy of human midcingulate cortex**.** *Eur. J. Neurosci*., 18 (11), 3134-3144.

Vogt, B. A. (2005). Pain and emotion interactions in subregions of the cingulate gyrus. *Nat Rev Neurosci,* 6,533–544.

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI Time-Series Revisited – Again. *NeuroImage*, 2(3), 173-181.

Wurm, M. F., & Schubotz, R. I. (2012). Squeezing lemons in the bathroom: contextual information modulates action recognition. *Neuroimage*, *59*(2), 1551-1559.

3.3 Study 3: Predictive impact of contextual objects during action observation: evidence from fMRI

Nadiya El-Sourani, Ima Trempler, Moritz F. Wurm, Gereon R. Fink, & Ricarda I. Schubotz
(2019)

Manuscript submitted to journal

Title: Predictive impact of contextual objects during action observation: evidence from fMRI

Authors: Nadiya El-Sourani[1,2], Ima Trempler[1], Moritz F. Wurm[3], Gereon R. Fink[2,4], Ricarda I. Schubotz[1,2]

[1]Department of Psychology, Westfälische Wilhelms-Universität, 48149 Münster, Germany
[2]Department of Neurology, University Hospital Cologne and University of Cologne, 50937 Cologne, Germany
[3]Center for Mind/Brain Sciences (CIMeC), University of Trento, Rovereto, TN, Italy
[4]Institute of Neuroscience and Medicine (INM3), Cognitive Neuroscience, Research Centre Jülich, 52425 Jülich, Germany

Corresponding author:
Nadiya El-Sourani
Fliednerstr. 21,
48149 Münster
n_elso02@uni-muenster.de

Abstract:

The processing of congruent stimuli, such as an object or action in its typical location, is usually associated with reduced neural activity, probably due to facilitated recognition. However, in some situations, congruency increases neural activity, e.g., when objects next to observed actions are likely vs unlikely to be involved in forthcoming action steps. Here, we investigated using fMRI whether the processing of contextual cues during action perception is driven by their (in-)congruency and, thus, informative value to make sense of an observed scene. Specifically, we tested whether both highly congruent contextual objects (COs), which strongly indicate a future action step, and highly incongruent COs, which require updating predictions about possible forthcoming action steps, provide more anticipatory information about the action course than moderately congruent COs. In line with our hypothesis that especially the inferior frontal gyrus (IFG) subserves the integration of the additional information into the predictive model of the action, we found highly congruent and incongruent COs to increase bilateral activity in action observation nodes, i.e., the IFG, the occipitotemporal cortex (OTC), and the intraparietal sulcus (IPS). Intriguingly, Brodmann Area 47 was significantly stronger engaged for incongruent COs reflecting updating of prediction in response to conflicting information. Our findings imply that the IFG reflects the informative impact of COs on observed actions, by using contextual information to supply and update the currently operating predictive model. In the case of an incongruent CO, this model has to be reconsidered and extended towards a new overarching action goal.

Keywords:
fMRI, action observation, context, congruency, IFG

1. Introduction

In daily life, it is essential to understand what people around us are doing, that is, to predict the goal they currently pursue (cf. van Overwalle & Beatens, 2009; Caspers, Zilles, Laird, & Eickhoff, 2010). To this end, action observers can exploit various sources of information, including not only moving body parts (i.e., manipulation movements) and manipulated objects, but also various contextual factors, such as the room (Wurm & Schubotz, 2012), the actor (Hrkać, Wurm, & Schubotz, 2014), additional objects in a scene (*contextual objects*, COs; El-Sourani, Wurm, Trempler, Fink, & Schubotz, 2018), and spatial relations between objects and agents. While the influence of contextual information on object recognition has been intensively investigated (Bar, 2004; Boyce, Pollatsek, & Rayner, 1989; Hayes, Nadel, & Ryan, 2007; Zimmermann, Schnier, & Lappe, 2010; Barenholtz, 2014), its impact on action understanding has so far been addressed by only a few studies. The results of these studies suggest that participants process contextual information spontaneously, i.e., without task requirements: while participants need longer to recognize an action when it takes place in an incompatible vs a compatible or a neutral room (Wurm & Schubotz, 2012), action-compatible room information can help when actions are difficult to recognize, leading to increased recognition accuracy (Wurm & Schubotz, 2017). Moreover, brain activation during action recognition suggested interference effects of action-incompatible contexts rather than facilitation effects of action-compatible contexts (Hrkać et al., 2014; Wurm & Schubotz, 2012). For example, when the manipulated object does not fit room and manipulation, or when the manipulation does not fit object and room, neural activity increased in brain regions associated with object and manipulation processing, respectively (Wurm, von Cramon, & Schubotz, 2012). In particular, the inferior frontal gyrus (IFG) has been frequently linked to action-incompatible information processing, for instance when actions took place in incompatible rooms (Wurm & Schubotz, 2012). Also, an increase in IFG activity was found when participants observed an actor performing actions that did not match their current goal, supposedly reflecting attempts to integrate incoherent action steps into a common goal (Hrkać et al., 2014). So far, the involvement of the IFG underlines its central role in the integration of contextual information during action perception (Badre & Wagner, 2007; Kilner, 2011), and, from a broader perspective, its role in effortful contextual integration in different cognitive domains, including language (Poldrack et al., 1998; Smirnov et al., 2014; van Schie, Toni, & Bekkering, 2006). These findings indicate that different types of contextual information may impact on the processing of observed actions.

In a recent fMRI study (El-Sourani et al., 2018), we focused on *contextual objects (COs),* which are part of an observed action scene, yet not part of the action itself. By modulating the semantic relation (*Goal Affinity*) as well as the spatial relation (*Location Ergonomics*) of the CO to the observed action, we investigated under which conditions such task-irrelevant objects modulate an action observer's brain activity. We argued that such effects may reflect attempts to incorporate these COs into an internal model of the observed action in order to anticipate an overarching action goal. Functional MRI results confirmed that COs are processed during action observation, even though the participants' attention was tied to the observed action and considering the COs was not necessary to identify the action at hand. Contrary to the previously observed interference effects for action-incompatible information, for instance, an action-incompatible room (Wurm & Schubotz, 2012), we found significant engagement of brain areas associated with object-related action representation when COs were *highly compatible* with the observed action, e.g., a frying pan next to cracking an egg. Specifically, Brodmann Area (BA) 44 and BA 45 of the IFG showed increased activation when the location of the CO and its semantic relation to the observed action strongly implied its use in (immediately) upcoming action steps (El-Sourani et al., 2018). This apparent discrepancy of brain activation in response to manipulating congruency may be explained by the different operationalization of context-action incompatibility in the different studies, and point towards a more specific interpretation of the IFG's function in action observation. For example, Wurm and Schubotz (2012) investigated the effect of compatibility and incompatibility of room information on action perception (e.g., squeezing lemons in the bathroom). By contrast, in El-Sourani et al. (2018) the observed action (e.g., cracking an egg) was generally compatible with the room information implied by the CO but was more or less associated with the CO itself (frying pan vs wine opener). Hence, no strong *conflict* was induced by lowly congruent COs, suggesting that this object category was processed as part of the room, or room category, rather than concerning a potential usage by the actor. As we are used to be surrounded by room-compatible objects with low congruency to our currently performed action, such lowly congruent COs can usually be ignored. In contrast, COs with a high congruence (e.g., a frying pan) are probably perceived as comparatively *highly informative* (and thus relevant) for action observers in such a way that a specific overarching action goal (e.g., preparing scrambled eggs) can be inferred. Importantly, a similar degree of informativeness and thus relevance conceivably also applies to highly incongruent contextual information, as mismatches between the observed action and contextual information signal the need to reconsider the action's anticipated outcome (cf. Wurm &

Schubotz, 2012; Hrkać et al., 2015). In this case, the current predictive model of the observed action should be revised (cf. Kilner, Friston, & Frith, 2007; Kilner, 2011).

Extending upon these findings, the present fMRI study aimed at investigating if COs that neither match the current action nor the belonging room category exert a substantial impact on action observation. More specifically, when COs point to actions associated with a different room category, they should generate a real conflict for, and hence complicate, goal inference.

To test this assumption, participants watched action videos containing COs that varied with regard to three levels of the factor goal affinity: they either matched the currently observed action and the contextual background (i.e., highly congruent CO), only the contextual background but not the action (i.e., lowly congruent CO), or neither the contextual background nor the action (i.e., incongruent CO). To replicate previous findings, we also implemented the factor LOCATION ERGONOMICS, with varying positions of the COs on the table on which the action was performed (cf. El-Sourani et al., 2018).

We hypothesized that particularly high compatibility between an observed action and its context (highly congruent COs) as well as particularly high action-context incompatibility (incongruent COs) both provide rich information regarding potentially upcoming next action steps. Therefore, we expected brain activity to initially reflect an in-depth processing of these two object categories, demonstrated by an assumed overlap of neural activity elicited by highly congruent and incongruent COs, as compared to lowly congruent COs. Based on previous findings, this effect was expected to be reflected in brain areas linked to object-related action representations, especially the occipito-temporal cortex (OTC; Wigget & Downing, 2011) and the inferior parietal lobule (IPL; Buxbaum & Kalénine, 2010; Schubotz, Wurm, Wittmann, & von Cramon, 2014), as the perception of an object can already imply its manipulation and action (Buxbaum, Kyle, Tang, & Detre, 2006; Johnson-Frey, 2004; Schubotz et al., 2014). We particularly focused on the IFG due to its role in the retrieval and integration of action-relevant semantic information (Casper et al., 2010; Badre & Wagner, 2007). As outlined above, we argue that previous findings can be reconciled if the IFG not simply reflects integration attempts but rather signals how informative a contextual object is concerning an observed action's anticipated outcome. If so, IFG activity should be low for COs with a low congruence to the observed action, but high for action scenes with (a) COs that are highly congruent to the observed action, and (b) COs that neither match the observed action nor the room category (incongruent COs) in which the action is observed. More specifically, within the IFG, BA 44 is suggested to be involved in structuring sequences to realize particular outcomes (Fiebach & Schubotz, 2006; Grafman, 2002) thereby potentially supporting the anticipation of upcoming action steps during action perception (Csibra, 2007; Fagg & Arbib, 1998; Friston, Mattout, & Kilner, 2011; Schubotz & von Cramon, 2009). While BA 45 activation supports the selection among competitively activated semantic representations (Badre, Poldrack, Pare-Blagoev, Insler, & Wagner, 2005; Gold et al., 2006; Moss et al., 2005), BA 47 is suggested to be involved in top-down semantic retrieval of goal-relevant knowledge (e.g., when participants are asked to think of unusual functions of an object (Kröger et al., 2012). Hence, we expected BA 44 and 45 to be engaged to a similar degree for both highly congruent COs and incongruent COs, whereas BA 47 might be stronger engaged by incongruent COs as compared to highly congruent COs.

2. Materials and Methods

2.1 Participants

Thirty-five right-handed subjects (20 females; 24,6±3.1 years old; range, 19-30 years) with normal or corrected-to-normal vision participated in the study. Three of these participants were excluded due to either poor performance or strong head motion (more than 3mm between two scans). None of the remaining thirty-two participants reported a history of medical, neurological/psychiatric disorders or substance abuse. The study protocol was conducted following the ethical standards of the Declaration of Helsinki and approved by the local Ethics Committee of the University of Münster. Each subject submitted a signed informed consent before they participated in the study. Afterwards, participants either received course credits or reimbursement.

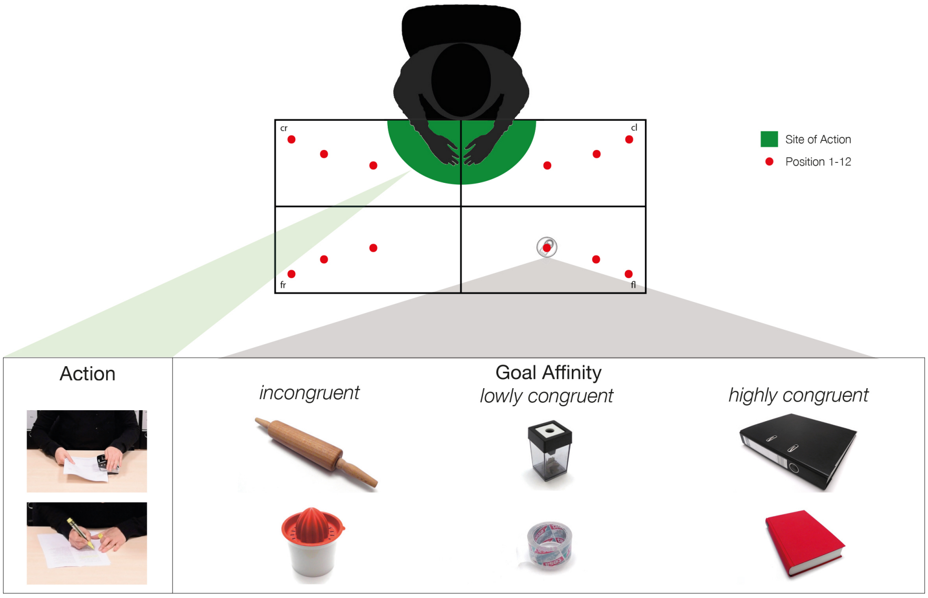


Figure 1. Example stimuli for implementing the three factor levels for goal affinity of COs, depicted for two different actions (punching and writing)**.** Red dots refer to possible CO locations on the table (Location Ergonomics).

2.2 Stimuli

Stimuli were presented using Presentation 18.1 (Neurobehavioral Systems, San Francisco, CA, USA). In total, participants were shown 360 action videos (action trials). Action trials were intermixed with 72 questions trials (20%), i.e., written action descriptions that referred to these actions (see 2.3 Task). Action and question trials had a duration of 6s and consisted of either an action video (3s) or a question (3s), followed by a fixation phase (3s). A variable jitter (500, 1000, 1500 ms) was included after the fixation phase in order to enhance the temporal resolution of the BOLD response. Finally, in 5% of the trials, a null event (fixation cross) was implemented (6s).

All action videos were performed by the same actress throughout the experiment and were filmed from a third person perspective (3pp). 72 actions were used. Each of the actions was performed in its typical setting that was either a kitchen (39 actions) or an office (33 actions), i.e., action and contextual background were always compatible (cf. El-Sourani et al. 2018). Each action video depicted a single object-directed action with two target objects. Out of the 360 action videos, 288 contained an additional contextual object (CO) that was positioned in front of the actress on the table (*Figure 2*). In a pilot study (N = 24), action videos with and without a CO were tested for their recognizability, and only those actions and COs were employed that were recognized by all participants.

COs varied according to two experimental factors: GOAL AFFINITY (GA), i.e., the semantic relation of the CO to the observed action, and LOCATION ERGONOMICS, i.e., the spatial relation of the CO to the observed action. Note that the latter factor was not relevant for testing the hypotheses of the present study, but was employed to replicate findings from the precursor study (El-Sourani et al., 2018). The factor GOAL AFFINITY had three levels:

1. *Highly congruent CO* (GA_{high}), depicting COs that are compatible with the contextual background and the action,
2. *Lowly congruent CO* (GA_{low}), depicting COs that are compatible with the contextual background but not the action,
3. *Incongruent CO* (GA_{no}), depicting COs that are neither compatible with the contextual background nor with the action.

Goal affinity was initially quantified based on subjective ratings of a large sample (N = 500) of students, in which participants had to rate the associative strength of objects (N = 144) and actions (N = 72). Based on this pilot data, objects were assigned to four different levels of goal affinity ranging from “very low associated” to “very high associated”. Subsequently, COs of level one (“very low associated”) and level two (“rather low associated”) were merged to form the level *lowly congruent CO* in the present study, whereas level three (“rather high associated”) and level four (“very high associated”) were merged to form the level *highly congruent CO,* corresponding to the categories of our previous study (El-Sourani et al., 2018).

To determine *incongruent COs,* we conducted a further pilot study, where COs belonging to the one contextual background (kitchen) were tested for their probability of occurrence in the other contextual background (office), and vice versa. Twenty-four right-handed participants rated on a 6-Point Likert scale how strongly the presented CO (e.g., rolling pin) fitted into the other room category (e.g., office). Objects creating the biggest mismatch according to these pilot data were chosen for the level *incongruent CO*.

As in our previous study (El-Sourani et al., 2018), the factor LOCATION ERGONOMICS was implemented by varying the locations of the CO on the table, corresponding to close-right (cr), close-left (cl), far-right (fr) and far-left (fl) concerning the action site (*Figure 1*).

Subsequently, each of the 72 actions was paired (using Adobe Premiere Pro CS, Adobe Photoshop and/or Matlab) with two COs of two different goal affinity levels. This ensured a balanced distribution of the GOAL AFFINITY levels, that is, videos containing a contextual object were arranged in a way that all goal affinity levels at all 12 positions occurred in an evenly distributed number (12 positions x 3 goal affinity levels x 8 occurrences = 288 action videos with a contextual object). In addition, each action was once seen without a contextual object, resulting in a total of 360 action videos.

Resulting videos were presented in a pseudo-randomized fashion by avoiding direct repetition of the presented action, the goal affinity, and the location of the CO. Levels of both factors were presented in an evenly distributed manner.

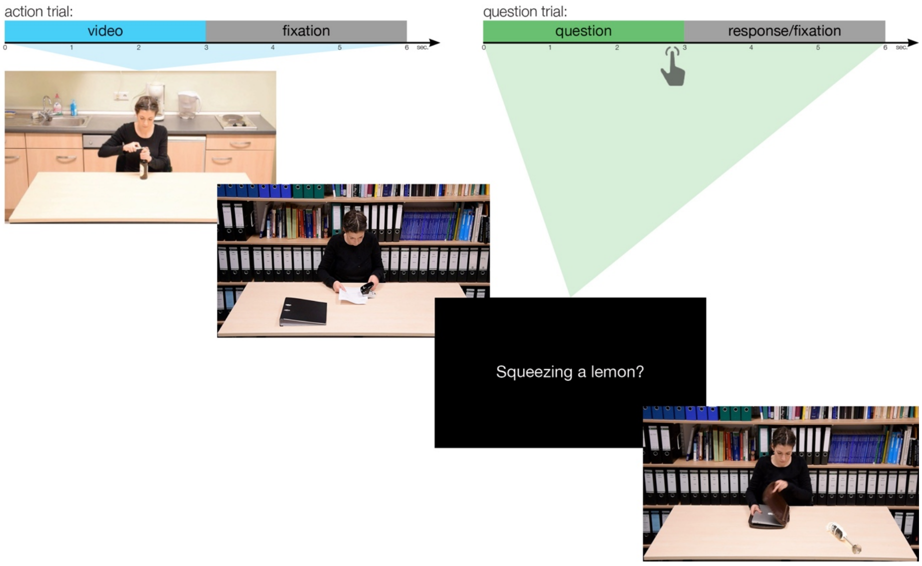


Figure 2. Schematic diagram of the task. Action trials consisted of an action video (3s) and a fixation phase (3s). Question trials consisted of a question regarding the previous video trial (n-1), followed by a response and fixation phase. Retrieved from El-Sourani et al. (2018) and partly modified.

2.3 Task

To keep the participants' attention focused on the videos, we asked them to watch the video clips attentively and to respond to the action description (20%) that either referred to the content of the preceding video (50%) or not (50%). Participants had to either accept or reject the action description using a two-button response box.

2.4 fMRI Image Acquisition

Imaging was performed on a 3 Tesla Siemens Magnetom Prisma MR tomograph using a 20channel head coil. Participants were located in a supine position on the scanner bed with their right index and middle fingers positioned on the appropriate response buttons of the response box. To minimize head and arm movements, head and arms were tightly fixated with form-fitting cushions. Furthermore, participants were provided with earplugs and headphones to attenuate the scanner noise. Whole-brain functional images were acquired using a gradient T2*-weighted single-shot echo-planar imaging (EPI) sequence sensitive to blood oxygenation level dependent (BOLD) contrast (64 x 64 data acquisition matrix, 192 mm field of view, 90° flip angle, TR = 2000 ms, TE = 30 ms). Each volume consisted of 33 adjacent axial slices with a slice thickness of 3 mm and a gap of 1 mm, which resulted in a voxel size of 3 x 3 x 4 mm.

Images were acquired in interleaved order along the AC-PC plane to provide whole-brain coverage. After functional imaging, structural data were acquired for each participant using a standard Siemens 3D T1-weighted MPRAGE sequence for detailed reconstruction of anatomy with isotropic voxels (1 x 1 x 1 mm) in a 256 mm field of view (256 x 256 matrix, 192 slices, TR = 2130, TE = 2.28).

For stimuli presentation, a 45° mirror was fixated on the top of the head coil. A video-projector projected the experiment on a screen that was positioned behind the participant's head so that they could see the stimuli via the mirror. The mirror was adjusted for each participant to provide a perfect view (center of the visual field). In a pilot study, we controlled for recognizability of actions and contextual objects using the final video selection. Only action videos in which the action and the contextual object could be identified by at least 95% of the participants were employed.

2.5 fMRI Data Analysis

2.5.1 fMRI data preprocessing

Brain image preprocessing and basic statistical analyses were conducted using SPM12 (Wellcome Department of Imaging Neuroscience, London, UK; www.fil.ion.ucl.ac.uk/spm/software/spm12/). Functional images were slice-timed to the middle slice to correct for differences in slice acquisition time. To correct for three-dimensional motion, individual functional MR (EPI) images were realigned to the mean EPI image and further motion correction estimates were inspected visually. The anatomical scan was co-registered (rigid body transformation) to the mean functional image. Each subject's co-registered anatomical scan was segmented into native space tissue components. The parameters obtained were applied to normalize the subject's functional scans to the template brain MNI space. Finally, the normalized images were spatially smoothed using a Gaussian kernel of 8mm3 full width at half-maximum (FWHM). A 128 s temporal high-pass filter was applied to the data to remove low-frequency noise.

2.5.2 Design specification

The statistical evaluation was based on a least-squares estimation using the general linear model (GLM) for serially autocorrelated observations (Friston et al., 1995; Worsley & Friston, 1995). The design matrix was generated with delta functions and convolved with a canonical hemodynamic response function. The subject-specific six rigid-body transformations obtained from residual motion correction were included as covariates of no interest. Activations were analyzed time-locked to the onset of the videos and the analyzed epoch comprised the full duration (3s) of the presented videos and the reaction time in question trials (max. 3s). In order to make results as comparable as possible between the current study and our previous study (El-Sourani et al., 2018), we aimed at having a similar design regarding our regressors. Therefore, our GLM contained 15 regressors in total: 12 predictors for the experimental conditions, one predictor for videos without a contextual object (noCO), one including all the null events (6s fixation phase), and one predictor for question trials. The 12 experimental regressors were assigned to the level combination of the factor LOCATION ERGONOMICS (close-right, close-left, far-right, and far-left) as well as the factor GOAL AFFINITY (highly congruent, lowly congruent, incongruent). To test for the effects of the factor GOAL AFFINITY, lowly congruent COs served as a control condition. Thus, to test for the effect of incongruent COs, all predictors containing incongruent COs were contrasted with all predictors containing lowly congruent COs ($GA_{no} > GA_{low}$) on a first level GLM. We contrasted highly congruent CO regressors with lowly congruent CO regressors ($GA_{high} > GA_{low}$) to replicate the main effect of GOAL AFFINITY as found in El-Sourani et al. (2018).

2.5.3 Group analysis

To obtain group statistics, the resulting contrast images of all participants for our contrasts of interests ($GA_{high} > GA_{low}$; $GA_{no} > GA_{low}$; $GA_{no} > GA_{high}$) were entered into a second-level random-effects analysis using a one-sample *t*-test across participants to test for significant deviations from zero. Subsequently, we corrected for multiple comparisons using the false discovery rate (FDR) method with $p < 0.01$. Significant activation maps were superimposed on a ch2better.nii.gz atlas using MRIcron software (https://www.nitrc.org/projects/mricron).

2.5.4 ROI analysis

To test whether different areas of the IFG are differentially modulated by the compatibility of the contextual information, we performed a ROI analysis of the contrasts of interest ($GA_{high} > GA_{low}$; $GA_{no} > GA_{low}$). Thus, lowly congruent COs served as baseline. Anatomical masks of left and right IFG were defined according to the automated anatomical labeling (AAL) atlas, implemented in SPM12 (Mazoyer & Joliot, 2002). To this end, we extracted mean beta scores per ROI and entered them into two-sided one-sample t-tests. Note that we aggregated beta values across hemispheres, as we did not hypothesize differential activation patterns regarding left and right IFG. To specifically test for a difference between incongruent and highly congruent COs regarding the different IFG areas (pars opercularis (BA 44), pars triangularis (BA 45), pars orbitalis (BA 47)), we performed one-sided paired-sample *t*-tests.

3. Results

3.1. Behavioral results

The performance was assessed by error rates and reaction times (of correctly answered trials). The average response time was 1257.69 ± 39ms, and the average error rate was low (2.47 ± 1.14%) indicating that participants attentively observed and recognized the actions.

3.2. fMRI results

3.2.1 Highly congruent COs (GA_{high} > GA_{low})

To test for the effect of highly congruent COs, we contrasted action videos containing highly congruent COs with lowly congruent COs, irrespective of their location on the table. Largely replicating previous findings (El-Sourani et al. 2018), highly congruent COs increased activity in the posterior parietal cortex (PPC) and the OTC (posterior temporal gyrus (pMTG), fusiform gyrus, and the lateral occipital complex (LOC)) bilaterally. Also, the right middle and superior frontal gyri were significantly activated (*Figure 3*). In contrast to our previous study, where we found an increase of IFG activity for highly congruent objects only when they were positioned close and to the right of the actress, the IFG here now became significantly activated independently of the position of the CO on the table (cf. *Figure 3*). The reverse contrast did not reveal any significant activation patterns after FDR-correction.

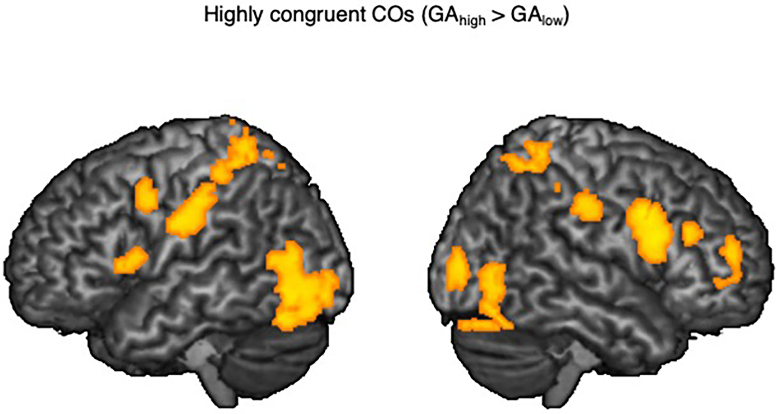


Figure 3. Brain activations for highly congruent vs lowly congruent COs (GA_{high} > GA_{low}), FDR corrected at $p < 0.01$.

Table 1. fMRI activations for highly congruent COs ($GA_{high} > GA_{low}$)

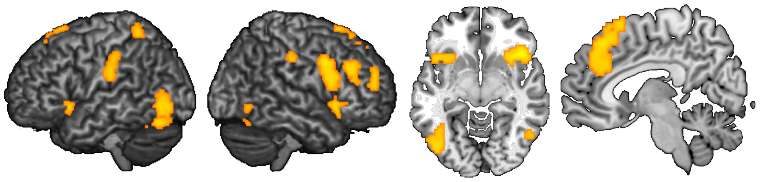
			MNI Coordinates			
Region	*Hemisphere*	*BA*	*x*	*y*	*z*	*t-scores*
Inferior/middle occipital gyrus	l	18/19	-39	-70	-4	6.25
	r		33	-82	5	5.04
Fusiform gyrus	l	37	-33	-58	-18	5.01
	r		36	-61	-16	5.22
MTG /ITG	l	19	-51	-64	5	4.38
	r		41	-64	5	4.71
PMv	l	6	-54	5	41	4.61
	r		60	11	35	5.91
IFG (pars opercularis)	l	44	-60	11	9	4.13
	r		60	17	20	5.58
IPL /IPS	l	7	-30	-46	53	4.35
	r		37	-49	56	4.64
SPL/IPS	l	7	-27	-49	68	4.59
	r		33	-49	60	5.38
Supramarginal gyrus	l	40	-60	-22	41	4.99
	r		54	-24	39	4.68
Postcentral gyrus	l		-36	-40	65	4.63
Precuneus	l	7	-9	-52	74	4.13
Superior medial frontal gyrus	l	9	-3	56	29	4.97
Superior frontal gyrus	r	10	32	54	11	4.29
Middle frontal gyrus	r	9	45	32	23	4.43
Cerebellum	r	-	18	-55	-43	4.94

r, right; l, left; x, y, z, MNI coordinates of peak voxel activation; MTG, middle temporal gyrus, ITG, inferior temporal gyrus; PMv, ventral premotor cortex; IFG, inferior frontal gyrus; IPL, inferior parietal lobule; SPL, superior parietal lobule; $p < 0.01$ FDR-corrected for multiple comparisons. *extending into left IFG (BA 44).

3.2.2 Incongruent COs ($GA_{no} > GA_{low}$)

Incongruent versus lowly congruent COs yielded significantly increased activation in the bilateral anterior dorsal insula, ACC, SMA and IFG (BA47). Moreover, we found a significant engagement of the left PPC, OTC as well as the right middle and superior frontal gyri (*Figure 4A*). As expected, brain areas partly overlapped with those engaged for highly congruent COs (*Figure 4B*). Again, the reverse contrast did not reveal any significant activation patterns after applying FDR-correction. Finally, directly contrasting incongruent COs with highly congruent COs did not reveal any significant whole-brain effects after FDR-correction.

a) Incongruent COs ($GA_{no} > GA_{low}$)



b) Overlay of highly congruent and incongruent COs

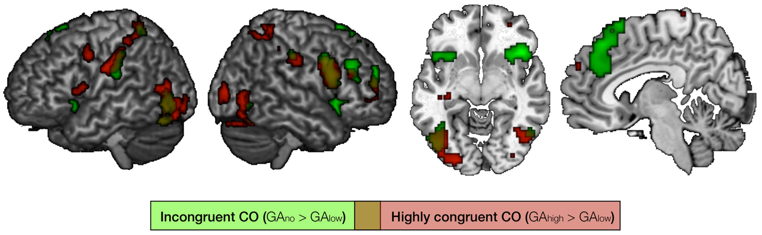


Figure 4. a) depicts brain activation patterns for incongruent COs, FDR corrected at $p < 0.01$. b) shows an overlay of activations for highly congruent (red) and incongruent COs (green). Activations of both object categories overlap in the IPL and OTC as well as in the right PMv and right IFG.

Table 2. fMRI activations for incongruent COs ($GA_{no} > GA_{low}$)

			MNI Coordinates			
Region	*Hemisphere*	*BA*	*X*	*y*	*z*	*t-scores*
Inferior/middle occipital gyrus	l	19	-45	-73	-10	5.65
Fusiform gyrus	l	37	-30	-52	-19	5.78
	r		33	-58	-19	4.5
MTG/ITG	l	37	-51	-61	-4	4.51
	r		52	-58	-5	4.71
IPL/ IPS	l	40	-54	-25	38	4.91
		7	-33	-52	55	4.35
SPL/ IPS	l	7	-28	-49	68	4.67
Supramarginal gyrus /Postcentral gyrus	l	40/1	-51	-25	38	5.17
	r		54	24	41	5.27
Insula	l	13	-30	17	-7	4.99
	r		42	23	-7	7.44
SMA	l	6	-3	20	66	4.71

	r		3	20	65	5.46
ACC/MCC	l	24/32	-3	11	23	4.16
	r		5	40	31	4.87
PMv	r	6	57	11	35	6.59
IFG (pars opercularis)	r	44	60	14	26	5.85
IFG (pars orbitalis)	l	47	-54	-25	38	5.09
	r		47	23	-4	5.62
Superior medial frontal gyrus	l	8	-2	34	51	5.27
	r	8	3	35	50	6.81
Superior frontal gyrus	r	10	30	47	9	4.79
Middle frontal gyrus	r	46	27	56	23	6.10
*	r	9	42	35	30	6.00
Cerebellum	l		-38	-72	-22	4.73

r, right; l, left; x, y, z, MNI coordinates of peak voxel activation; MTG, middle temporal gyrus, ITG, inferior temporal gyrus, PMv, ventral premotor cortex; IFG, inferior frontal gyrus; IPL, inferior parietal lobule; SPL, superior parietal lobule; SMA, supplementary motor area; IPS, intraparietal sulcus; ACC, Anterior cingulate cortex; MCC, middle cingulate cortex; $p < 0.01$ FDR-corrected for multiple comparisons; * extending into IFG (pars triangularis; BA45).

3.2.3 ROI analysis of CO congruency effects in the IFG

To assess a putative differential contribution of different sub-regions of the IFG to the processing of incompatible vs highly compatible contextual information, we performed ROI analyses of the pars opercularis, pars triangularis and pars orbitalis roughly corresponding to BA 44, 45 and 47, respectively (*Figure 5*). We extracted beta values for highly congruent COs ($GA_{high} > GA_{low}$) and incongruent COs ($GA_{no} > GA_{low}$). As we did not hypothesize differences in activation between left and right IFG we aggregated beta values of both hemispheres. One sample *t*-tests revealed significant activations for all conditions of interest (GA_{high}, GA_{no}) as compared to baseline (GA_{low}) (*Figure 5*). Finally, one-sided paired sample *t*-tests revealed a significant difference between incongruent and highly congruent COs in BA 47 ($t_{(63)} = 2.479$; $p < 0.01$).

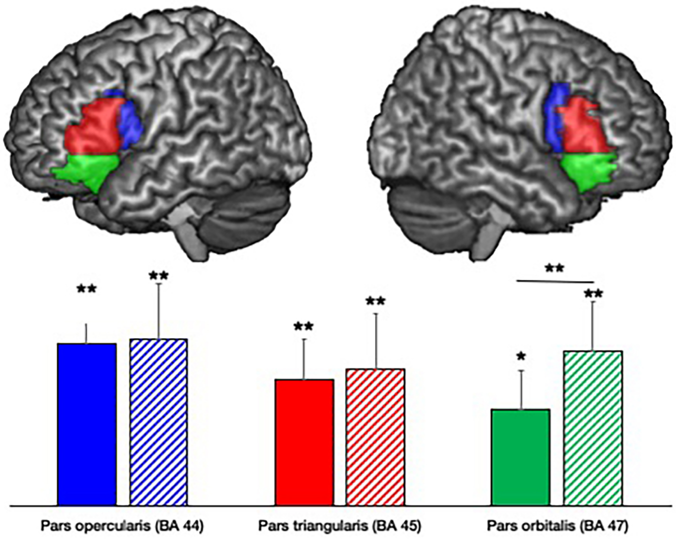


Figure 5. ROI analysis of IFG sub-regions according to the AAL atlas. Applied masks are illustrated in blue, red, and green for BA 44, 45, and 47, respectively. Corresponding beta values for highly incongruent versus lowly congruent COs are depicted in full colored bars, whereas beta values for incongruent COs versus lowly congruent COs are depicted in striped colored bars. Beta values for right and left were aggregated. The two conditions significantly differed from zero in all sub-regions of the IFG. Regarding BA 47 (pars orbitalis), paired sample *t*-tests revealed a significant increase of activation for incongruent as compared to highly congruent COs. None of the other pairwise comparisons revealed a significant difference. * corresponds to a significance level of $p < 0.05$ and ** to a significance level of $p < 0.01$.

4. Discussion

When we observe others' actions, specific brain regions are involved in integrating this action with contextual information to enable the inference of action goals. This contextual information includes not only the environment and the actor but as recently found also unused objects nearby (El-Sourani et al., 2018). In the current study, we aimed to better understand the processes underlying the latter effect. Specifically, we tested the assumption that the brain's engagement in processing contextual objects (CO) is not driven by the COs congruency or incongruency to the observed action, but rather to the CO's potential to inform expectations towards upcoming action steps.

Relative to COs with a low congruency to the observed action, both highly congruent and entirely incongruent COs were accompanied by increased brain activation at several action observation network (AON) sites, amongst these, as hypothesized, the OTC and the PPC (especially the IPS). The same effect was found for the IFG, the area we focused on due to its well-established role in the processing of semantic information. Interestingly, BA 47 of the IFG was especially engaged in the processing of incongruent COs.

These findings support the view that when observing an action, the brain is particularly tuned to highly informative context. Contextual information may exert its impact via probabilistic associative knowledge about rooms, in which certain classes of actions are frequently observed (Wurm & Schubotz, 2012), or about objects that are frequently used in the same sequence of actions (El-Sourani et al., 2018). In support of the latter assumption, posterior AON areas associating objects with actions, including the OTC (Grill-Spector, Kourtzi, & Kanwisher, 2001; Wigget & Downing, 2011) and the IPS (Creem-Regehr, 2009; Grefkes & Fink, 2005; Singh-Curry & Husain, 2009; Ramsey, Cross, & Hamilton, 2011), were significantly more active for both highly congruent and incongruent COs as compared to lowly congruent COs. Moreover, processing incongruent COs engaged a set of brain areas related to conflict and error processing: the ACC (Botvinick, Cohen, & Carter, 2004) and the anterior insula (Klein, Ullspenger, & Danielmeier, 2013). More precisely, these areas have been suggested to operate as a (response) inhibition network (Kana, Keller, Minshew, & Just, 2007, see also Botvinick, Braver, Barch, Carter, & Cohen, 2001; Hoffstadter et al., 2014), indicating that the processing of this object category also entailed processing its conflict to the observed action. Put into broader context, the efficient processing of an observed action scene includes the selection of sensory input that is crucial for informing the expectation of potential outcomes of the observed action (cf. Csibra, 2007; Kilner et al., 2007), irrespective of whether the information fits or contradicts the action.

The current study focused particularly on BOLD effects in the IFG, which is known for its role in retrieval and integration of semantic information (Caspers et al., 2010; Kilner, 2011). We here aimed to extend this role and hypothesized that the IFG is sensitive to the informative impact of COs concerning a potential refinement of expectations of action outcomes. As expected, ROI analyses revealed significant engagement of all IFG compartments - BA 44, BA 45 and BA 47 - for both incongruent and highly congruent COs as compared to lowly congruent COs. The three sub-regions of the IFG have been associated with different functions across different domains, including language (cf. Bookheimer, 2002; Liakakis, Nickel, & Seitz, 2011), emotion processing (Carr, Iacoboni, Dubeau, Mazzioatta, & Lenzi, 2003; Seitz et al., 2008), and creativity (Kröger et al., 2012). According to different accounts regarding the IFG function, (Koechlin & Summerfield, 2007; Udden & Bahlmann, 2012; Badre & Wagner, 2007), the IFG is hierarchically organized in a functional stepwise gradient along the rostro-caudal axis, where top-down control is exerted from anterior to posterior regions (Koechlin, Ody, & Kouneiher, 2003). In specific reference to action observation, the more posterior the area the more it is suggested to contribute to constraining the immediate action requirements or options. More anterior sites, in turn, are more content-independent and associated with high-level goals (Badre, Poldrack, Pare-Blagoev, Insler, & Wagner, 2005; Badre & D'Esposito, 2009; cf. Buckner, 2003). More specifically, BA 44 (together with premotor regions) supports structured sequence processing (cf. Udden & Bahlman, 2012) in order to realize a particular outcome (Fiebach & Schubotz, 2006; Grafman, 2002). Regarding BA 45 and 47, Badre and Wagner (2007) associated strategic memory retrieval with BA 47 and post-retrieval selection among competing memories with BA 45. While semantic retrieval is necessary when bottom-up cues are not sufficient to activate goal-relevant information, post-retrieval selection is necessary to resolve the competition between simultaneously activated memory representations (e.g., grasping to clean vs grasping to drink).

The assumed functions of the IFG compartments concur with the activation pattern observed in our current study: incorporating relevant CO information into an observed action scene in order to anticipate its outcome draws on all IFG sub-regions. However, while BA 44 and 45 were recruited to an equivalent degree by highly congruent and incongruent COs, BA 47 was significantly stronger engaged by incongruent COs. This underlines the ascribed function of the BA 47 in the controlled retrieval, i.e., a top-down process activating goal-relevant knowledge especially in the face of contradicting representations. Thus, increased BA 47 in response to incongruent vs highly congruent COs can be explained by the increased demand to retrieve an action outcome when confronted with conflicting action-related information. More specifically, incorporating incongruent COs and the observed action under a common overarching goal requires a much higher level of abstraction, evoked by the associative strength of the CO to an incompatible room category and hence to actions associated with this room category. Importantly, the observed pattern of activation in the IFG does not simply reflect demands on integrating more or less compatible contextual information (here: contextual objects) in the observed action. In that case, one would see a parametric increase of IFG activation with increasing incongruence, i.e., lowest IFG engagement for highly congruent COs. Instead, the IFG rather appears to respond to contextual information that specifies and/or enriches the interpretation of an observed action and ignores contextual information that is less informative for action interpretation.

Taken together, our findings imply that the brain considers the informative value of contextual objects when observing an action. More specifically, our results suggest that the neural activity of the IFG reflects the informational impact of COs on an observed action at several circumstances: either when the contextual information depicts a strong match so that the currently operating predictive model can be updated and specified towards a particular outcome; or when the contextual information reveals a strong conflict with the observed manipulation, in which case the currently operating predictive model has to be reconsidered and possibly extended towards a new overarching action goal.

Literature

Badre, D., Poldrack, R.A., Pare-Blagoev, E.J., Insler, R.Z., Wagner, A.D., 2005. Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral pre-frontal cortex. *Neuron 47*, 907–918.

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, *45*(13), 2883-2901.

Badre, D., & D'esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical?. *Nature Reviews Neuroscience, 10*(9), 659.

Bar, M. 2004. Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617-629.

Barenholtz, E. (2014). Quantifying the role of context in visual object recognition. *Visual Cognition, 22*(1), 30-56.

Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annual review of neuroscience*, *25*(1), 151-188.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D., (2001). Conflict monitoring and cognitive control. *Psychological Review*, 108(3), 624-652.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in cognitive sciences, 8*(12), 539-546.

Boyce, S.J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. Journal of Experimental Psychology: *Human Perception Performance, 15*(3), 556-566.

Brandeis, R., Brandys, Y., & Yehuda, S. (1989). The use of the Morris water maze in the study of memory and learning. *International Journal of Neuroscience*, 48(1-2), 29-69.

Buckner, R. L. (2003). Functional–anatomic correlates of control processes in memory. *Journal of Neuroscience*, *23*(10), 3999-4004.

Buxbaum, L. J., & Kalénine, S. (2010). Action knowledge, visuomotor activation, and embodiment in the two action systems. *Annals of the New York Academy of Sciences*, *1191*(1), 201-218.

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceedings of the national Academy of Sciences*, *100*(9), 5497-5502.

Carter, C.S. et al. (1998) Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science 280*, 747–749.

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B., (2010). ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage*, 50(3), 1148-1167.

Creem-Regehr S. H. (2009). Sensory-motor and cognitive functions of the human posterior parietal cortex involved in manual actions. *Neurobiology of Learning and Memory, 91*(2), 166–171.

Csibra, G., 2007. Action mirroring and action interpretation: An alternative account. In: P. Haggard, Y. Rosetti, & M. Kawato (Eds.), *Sensorimotor Foundations of Higher Cognition. Attention and Performance XXII* (pp. 435-459). Oxford: Oxford University Press

El-Sourani, N. Wurm, M. F., Trempler, I., Fink, G. R., & Schubotz, R. I. (2018). Making sense of objects lying around: How contextual objects shape brain activity during action observation. *Neuroimage*, 167, 429-437.

Fagg, A. H., & Arbib, M. A. (1998). Modeling parietal–premotor interactions in primate control of grasping. *Neural Networks*, *11*(7-8), 1277-1303.

Fiebach, C. J., & Schubotz, R. I. (2006). Dynamic anticipatory processing of hierarchical sequential events: a common role for Broca's area and ventral premotor cortex across domains?. *Cortex*, *42*(4), 499-502.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., & Frackowiak, R. S. J. (1995). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2(4), 189–210.

Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological cybernetics*, *104*(1-2), 137-160.

Gold, B. T., Balota, D. A., Jones, S. J., Powell, D. K., Smith, C. D., & Andersen, A. H. (2006). Dissociation of automatic and strategic lexical-semantics: functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. *Journal of Neuroscience*, *26*(24), 6523-6532.

Grafman, J., 2002. The structured event complex and the human prefrontal Cortex. In: Stuss DT. & Knight RT. Principles of frontal lobe function. Oxford University Press, New York.

Grefkes, C., & Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy, 207*(1), 3–17 10.1111/j.1469-7580.2005.00426.x

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision Research*, 41(10-11), 1409-1422.

Hayes, S. M., Nadel, L., & Ryan, L. (2007). The effect of scene context on episodic object recognition: parahippocampal cortex mediates memory encoding and retrieval success. *Hippocampus, 17*(9), 873-889.

Hoffstaedter, F., Grefkes, C., Caspers, S., Roski, C., Palomero-Gallagher, N., Laird, A. R., ... & Eickhoff, S. B. (2014). The role of anterior midcingulate cortex in cognitive motor control: evidence from functional connectivity analyses. *Human brain mapping*, *35*(6), 2741-2753.

Hrkać, M., Wurm, M. F., & Schubotz, R. I. (2014). Action observers implicitly expect actors to act goal-coherently, even if they do not: An fMRI study. *Human brain mapping*, *35*(5), 2178-2190.

Hrkać, M., Wurm, M. F., Kühn, A. B., & Schubotz, R. I. (2015). Objects mediate goal integration in ventrolateral prefrontal cortex during action observation. *PloS one*, *10*(7), e0134316.

Kana, R. K., Keller, T. A., Minshew, N. J., & Just, M. A. (2007). Inhibitory control in high-functioning autism: decreased activation and underconnectivity in inhibition networks. *Biological psychiatry*, 62(3), 198-206.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive processing*, *8*(3), 159-166.

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in Cognitive Sciences*, *15*(8), 352-357.

Klein, T. A., Ullsperger, M., & Danielmeier, C. (2013). Error awareness and the insula: links to neurological and psychiatric diseases. *Frontiers in human neuroscience*, *7*, 14.

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. Trends in cognitive sciences, 11(6), 229-235.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, *302*(5648), 1181-1185.

Kröger, S., Rutter, B., Stark, R., Windmann, S., Hermann, C., & Abraham, A. (2012). Using a shoe as a plant pot: neural correlates of passive conceptual expansion*.* *Brain research, 1430*, 52-61.

Liakakis, G., Nickel, J., & Seitz, R. J. (2011). Diversity of the inferior frontal gyrus—a meta-analysis of neuroimaging studies. *Behavioural brain research*, *225*(1), 341-347.

Moss, H. E., Abdallah, S., Fletcher, P., Bright, P., Pilgrim, L., Acres, K., & Tyler, L. K. (2005). Selecting among competing alternatives: selection and retrieval in the left inferior frontal gyrus. *Cerebral Cortex*, *15*(11), 1723-1735.

Petrides, M. (2005). Lateral prefrontal cortex: architectonic and functional organization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 781-795.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. N*euroimage*, 10(1), 15-35.

Ramsey R., Cross, E. S., & Hamilton, A. F. D. C. (2011). Eye Can See What You Want: Posterior Intraparietal Sulcus Encodes the Object of an Actor's Gaze. *Journal of Cognitive Neuroscience*, *23*(11), 3400-3409

Rubia, K., Russell, T., Overmeyer, S., Brammer, M. J., Bullmore, E. T., Sharma, T., Simmons, A., Williams, S. C. R., Giampietro, V., Andrew, C. M., & Taylor, E. (2001). Mapping motor inhibition: conjunctive brain activations across different versions of go/no-go and stop tasks. *Neuroimage*, *13*(2), 250-261.

Smirnov, D., Glerean, E., Lahnakoski, J. M., Salmi, J., Jääskeläinen, I. P., Sams, M., & Nummenmaa, L. (2014). Fronto-parietal network supports context-dependent speech comprehension. *Neuropsychologia, 63*, 293-303.

Schubotz, R. I., & von Cramon, D. Y. (2009). The case of pretense: Observing actions and inferring goals. *Journal of Cognitive Neuroscience*, *21*(4), 642-653.

Schubotz, R. I., Wurm, M. F., Wittmann, M. K., & von Cramon, D. Y. (2014). Objects tell us what action we can expect: Dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI. *Frontiers in Psychology, 5*, 636.

Singh-Curry, V., & Husain M. (2009). The functional role of the inferior parietal lobe in the dorsal and ventral stream dichotomy. *Neuropsychologia*, *47*(6):1434-1448

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage, 15*(1), 273-289.

Uddén, J., & Bahlmann, J. (2012). A rostro-caudal gradient of structured sequence processing in the left inferior frontal gyrus. *Philosophical Transactions of the Royal Society B: Biological Sciences, 367*(1598), 2023-2032.

Van Overwalle, F., & Beatens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, 48, 564-584.

Van Schie, H. T., Toni, I., & Bekkering, H. (2006). Comparable mechanisms for action and language: neural systems behind intentions, goals, and means. Cortex, *42*(4), 495-498.

Wigget, A. J., & Downing, P. E. (2011). Representation of Action in Occipito-temporal Cortex. *Journal of Cognitive Neuroscience, 23*(7), 1765-1780.

Worsley, K. J., & Friston, K. J. (1995). Analysis of fMRI Time-Series Revisited – Again. *NeuroImage*, *2*(3), 173-181.

Wurm, M. F., von Cramon, D. Y., & Schubotz, R. I. (2012). The context-object-manipulation triad: cross talk during action perception revealed by fMRI. *Journal of Cognitive Neuroscience, 24*(7), 1548-1559.

Wurm, M. F., & Schubotz, R. I. (2012). Squeezing lemons in the bathroom: contextual information modulates action recognition. *Neuroimage*, *59*(2), 1551-1559.

Wurm, M. F., & Schubotz, R. I. (2017). What's she doing in the kitchen? Context helps when actions are hard to recognize. *Psychonomic Bulletin & Review, 24(*2), 503-509.

Zimmermann, E., Schnier, F., Lappe, M., (2010). The contribution of scene context on change detection performance. *Vision Research, 50*(20), 2062-2068.

4. General Discussion and Future Directions

The present work investigated how contextual object (CO) information modulates action perception and interpretation. More specifically, I examined the neural mechanisms underlying the integration of COs in semantic and spatial relation to observed actions. Study 1 tested whether action observers process COs with regard to the observed action. In Study 2, I considered whether, depending on the CO's location, action observers also establish their own action space by perceiving COs not only with respect to the observed action but also with reference to themselves. Finally, by adding incongruent COs to the experimental manipulation, Study 3 examined whether the processing of COs is driven by their informative value to the observed action with regard to its potential outcome. In the following I will initially summarize and discuss findings of these studies with regard to i) general effects of the CO, ii) effects evoked by modulating the goal affinity of the CO and iii) effects evoked by modulating the location of the CO. Subsequently I will incorporate the findings of this thesis with the current state of research by proposing a model of action prediction in which the IFG selectively uses relevant contextual information to update currently operating predictive models. I will close this section by providing future directions.

4.1. General effects of COs

Study 1 revealed that COs are indeed processed by action observers, although no a priori attention was drawn towards them. More specifically, the presence of a CO was not even necessary to accomplish the task, i.e., naming the observed object manipulation. The mere presence of an additional object anywhere on the table increased activation in both the left IPS (Creem-Regehr, 2009; Grefkes & Fink, 2005; Singh-Curry & Husain, 2009; Ramsey, Cross, & Hamilton, 2011) and the fusiform gyrus (cf. Bar et al., 2001; Bar, 2004; Tyler et al., 2004).

These areas are known for visual, haptic and action-related object processing. I also found increased activation of the OPA: a scene-selective region that codes local elements in a scene (Kamps, Julian, Kubilius, Kanwisher, & Dilks, 2016; see also Malcolm et al., 2016). The question could be raised as to whether the increased neural activity for actions including a CO — compared to those which did not — might have solely been evoked by the additional visual information. In this case, however, activity should only have impacted activation in the (early) visual areas and not, for example, in the PPC: an area related to the visuomotor transformation of visually guided actions (cf. Section 1.4.1). In addition, as will be discussed in the upcoming sections, the goal affinity (Studies 1 and 3) and location (Studies 1 and 2) of COs were also processed by action observers, implying that action observers do indeed exploit the CO information of an observed action, beyond their mere visual information. In the following sections, the results and limitations of these studies are discussed in more detail.

4.2 COs' semantic relation (goal affinity) to the observed action

Study 1 investigated if and how COs are processed by action observers. I employed COs that were either highly congruent or lowly congruent with the observed action and varied in their location on the table. The study revealed that highly congruent COs increased bilateral activity in AON nodes, i.e., the OTC and the IPS, but also in the precuneus and middle frontal gyri. This finding was interpreted in terms of enhanced action anticipation with regard to a specific overarching action goal, provided by the informative value of highly associated COs. When these highly matching COs were presented in the preferred action zone of the actor (close and to the right of the action site), an additional increase of IFG activation signified efforts to integrate the CO into potential upcoming action steps. On the contrary, lowly congruent COs did not reveal any significant activation patterns, despite their incompatibility with the observed action. However, as both highly congruent and lowly congruent COs were part of the same room category as the observed action, it was suggested that lowly congruent COs were processed as part of the room but not with regard to their potential usage by the actor. These objects neither gave information about upcoming action steps to allow higher-level goal anticipation nor did they give rise to a strong conflict through their presence. Thus, they could be ignored with regard to further processing. In line with this argumentation, Wurm and Schubotz (2012) did not find a significant increase in IFG activation for actions occurring in compatible rooms as opposed to incompatible rooms. More specifically, here compatible room information (e.g., a kitchen) was operationalized by placing loosely related objects around the actor (e.g., a fridge, a kettle, a stove and a bottle of milk). Hence, the unspecific compatibility of lowly congruent COs in the present studies is very similar to that of the compatible room information in Wurm & Schubotz (2012). Following on from this, in Study 3, the goal affinity factor was extended by implementing *incongruent COs.* Incongruent COs neither match the observed action nor the corresponding room category, thereby posing a real conflict to the observed action scene. I aimed to investigate whether COs are processed as a function of their informative value to the observed action, rather than simply by either compatibility or incompatibility. The potential conflict that arises from the presence of incongruent COs could signal the need to reconsider the action's anticipated outcome and hence also provide action observers with relevant information. Indeed, Study 3 revealed an increased activation of posterior AON sites (OTC and IPS) for both highly congruent and incongruent COs, compared to lowly congruent COs. The same effect was observed for the engagement of the IFG. Interestingly, BA 47 of the IFG was especially engaged for incongruent COs, accompanied by the activation of the anterior cingulate cortex (ACC) and the anterior dorsal insula.

In the following subsections, I will discuss the findings for highly congruent and incongruent COs in more detail, before taking location ergonomics into closer consideration in Section 4.3.

4.2.1 Highly congruent COs

Initially, the effect of highly congruent COs seems to contrast with previous findings on the impact of contextual information on neural activity (Wurm & Schubotz, 2012; Hrkać et al., 2014; Hrkać et al., 2015; Wurm et al., 2014). These studies revealed either no increased activation or attenuation (Wurm et al., 2014) for compatible contextual information. However, this discrepancy can be resolved when looking at the studies more closely.[17] In Wurm et al. (2014), participants watched videos of action steps that either amounted to an overarching action goal or did not, and were performed by either one or several actors. OTC and inferior frontal activation decreased as a function of the predictability of a long-term goal, i.e., a long-term goal becomes more and more evident in the unfolding of an episode.

In fact, action goal anticipation *should* trigger the activation of the predicted sensory consequences of an action (internal representation of subsequent action steps). In this sense, attenuation *should not occur* before the predicted action steps are actually perceived (cf. Waszak, Cardoso-Leite, & Hughes, 2012). While this was indeed the case in Wurm et al. (2014), in my studies the effect of COs on the processing of a single action step was investigated but no further action steps as part of an episode were presented.

Highly congruent COs thus provided valuable meaning by potentially informing action observers about the *upcoming* action steps towards an overarching action goal and thus supporting action prediction (cf. Kilner, 2007) by *triggering* the corresponding action representations. However, these upcoming action steps were never observed by the participants, which might explain the increase as opposed to the attenuation of activation of the AON. In accordance with the PCT, the accurate prediction of upcoming action steps should suppress (or

---

[17] In the above Section 4.2, the difference in the operationalization of Wurm & Schubotz (2012) is described, explaining that it is the effect of lowly congruent COs and not that of highly congruent COs that is comparable to the compatibility effect of Wurm & Schubotz (2012).

attenuate) *upcoming* sensory input (Friston et al., 2011; Kilner et al., 2007). Following this line of argumentation, Iacoboni et al. (2005) found increased activation of the premotor and inferior frontal cortices, when a "reach and grasp" action was put into a meaningful context ("to clean" or "to drink"), but not when the action was observed against a plain background. Presumably, the additional congruent and meaningful information provided action observers with a higher-level goal (cleaning, drinking), thereby not only activating the corresponding motor act of the observed action itself (grasping), but also activating additional motor acts to accomplish the higher-level goal, i.e., activating the motor representation of the corresponding action steps. The same logic could possibly apply to actions involving highly congruent COs in the present studies, which imply their upcoming usage and thus provide more anticipatory information for the inference of higher-level action goals. In a similar vein, Calvo-Merino and co-workers (2005) reported an almost identical network as was found for highly congruent COs in Study 1, when expert dancers watched previously learned vs novel movements. These activations were suggested to surface a recall of acquired action representations, reflecting greater skills in anticipating these actions' outcomes (cf. Aglioti, Cesari, Romani, & Urgesi, 2008). Finally, it is worth noting that the engagement of the dlPFC and precuneus, which accompanied the posterior AON activation in Study 1 for highly congruent COs, has been associated with i) anticipatory performance (Balser et al., 2014; Wurm et al., 2014) and ii) visuomotor imagery and action planning (Zhang & Chiang-Shan, 2012). Finally, I found increased IFG engagement (Study 1) when highly congruent COs were presented close to and on the right side of the actor: an area in which participants place objects they plan to use shortly. The IFG's presumed core function during action observation is for the retrieval and integration of action-relevant semantic information (Caspers et al., 2010; Kilner, 2011), and the IFG has been consistently found to play a role in studies investigating the potential impact of contextual information on action observation and goal anticipation (Wurm et al., 2012; Wurm et al., 2014; Hrkać et al., 2014; Hrkać et al., 2015). I will discuss the role of the IFG in light of the presented studies in more detail in Section 4.4.2.

Taken together, the findings for highly congruent vs lowly congruent COs imply that these COs provide meaningful information for action observers. They enable the anticipation of a specific overarching action goal by triggering (rather than facilitating) representations of corresponding and potentially imminent action steps.

4.2.2 Incongruent COs

In Study 3, incongruent COs were included in the experimental manipulation. The increased pattern of activation for incongruent as opposed to lowly congruent COs is in good accordance with previous studies on the impact of contextual information on action observation (Wurm & Schubotz, 2012; Hrkać et al., 2014; Hrkać et al., 2015). In all of these studies, the processing of incompatible contextual information was interpreted as an attempt to integrate the conflicting information under a common overarching goal, entailing a reconsideration of the currently predicted action outcome. Engagement of brain regions included areas related to the specific stimulus category (e.g., the OTC and IPS in Study 3 of the present studies), as well as the IFG.

Following this line of argumentation, incongruent COs should provide action observers with valuable additional information regarding the potential inference of an overarching action goal. However, compared to highly congruent COs, the incorporation of incongruent COs and an observed action under a common goal requires a much higher level of abstraction. These objects belong to a different room category and are therefore strongly associated with actions usually occurring in this other specific setting. In support of this assumption, activity in BA 47 increased for incongruent COs, compared to highly congruent COs. This area is suggested to retrieve long-term memory representations when bottom-up cues are not sufficient for the interpretation of an observed scene (cf. Badre & Wagner, 2007).

But is the observed action indeed necessary to elicit the observed pattern of activation for incongruent COs? In order to unambiguously show that the observed effect for incongruent COs can(not) be solely traced back to room–object incompatibilities, fMRI experiments are needed that can investigate CO–room associations without the employment of additional primary action information.[18] Here, one cannot completely rule out that incongruent COs are preferentially processed due solely to the conflict arising from their presence within an incompatible room category. In support of this argumentation, behavioural results indicate that object recognition is facilitated by strong context–object associations and hampered by low context–object associations (Biederman et al., 1982; also see Oliva & Torralba, 2007). Objects are recognized faster than action kinematics, as the latter only develop over time. While the manipulated objects might have been partly occluded by the performed manipulation (primary action information), COs were simultaneously presented in isolation somewhere on the table and might have impacted scene processing (or vice versa; cf. Joubert et al., 2007) by producing incompatibility effects before effective action recognition could be achieved. Nevertheless, I instead speculate that incongruent COs are processed with regard to the observed action. First, the main task of the participants was to watch and understand the presented object manipulation. No a priori attention was drawn to the presence of COs, nor were they necessary to accomplish the task. Second, in Study 1, I was able to show that the specific semantic match of a CO to the observed action led action observers to consider the COs in depth. Lowly congruent COs did not demonstrate this effect. Here, both lowly congruent and highly congruent COs were

---

[18] Additionally, if objects were the main driver of action understanding as proposed by the affordance matching hypothesis (Bach et al., 2014), it would be possible for the effect of incongruent COs to be traced back to an object–object conflict between the action objects (manipulated objects) and the CO, irrespective of the room and the actor (i.e., action kinematics). Further studies are warranted that can explicitly test this possibility.

strongly associated with the room category, ruling out the possibility that the effect (for highly congruent COs) was a mere compatibility effect between the CO and the observed room category. Third, both highly congruent and incongruent COs engaged the IFG: an area which has been related to the integration of action-related semantic information en route to goals (Badre & Wagner, 2007; Kilner, 2011) and is thus sensitive to the processing of action goals (Schubotz & Cramon, 2009; Wurm et al., 2014; Hrkać et al., 2014; Hrkać et al., 2015; Iacoboni et al., 2005).

Incorporating my findings into the current accounts of the IFG’s function, one has to consider whether the IFG might react to the informative value of contextual information for action selection, rather than simply reflecting incompatible or compatible information. Following on from this, the IFG’s function should be interpreted and discussed in light of action prediction and the updating of currently operating generative models when observing an action. Before discussing these issues in detail in Section 4.4, I will first consider how the location of a CO impacted the processing of that CO in relation to the observed action.

4.3 The impact of the CO’s location on action observers

The CO’s location on a table was consistently and systematically varied with respect to the observed action. In the final action videos, the table was subdivided into four quadrants: close-right (cr), close-left (cl), far-right (fr), and far-left (fl). For each quadrant, three positions were selected, which differed in regard to reachability (high, middle, low) and, at the same time, mirrored the positions of the other quadrants. The camera perspective chosen for the final videos gave the impression that one was sitting at the table opposite the actor. This experimental setup enabled reachability to be investigated for an actor from an observing perspective (3pp) as well as for the observer from an egocentric frame of reference (1pp). In the following section, the results concerning the CO’s location with reference to i) the actor and ii) the observer will be discussed in detail.

4.3.1 COs located in the peripersonal space of an actor

In Study 1, we did not find any significant brain activation for the main effect of location ergonomics (close, far). However, we did find an interaction between the goal affinity of the CO and its location relative to the actor: when highly congruent COs were positioned close to and to the right of the actor, we found an increased activation of the bilateral IFG. This location is where right-handers prefer to store objects they are planning to subsequently use (cf. Toney & Thomas, 2006; Rezaee, Shojaee, Ghasemi, Moghaddam, & Momeni, 2010; Bryden, Pryde, & Roy, 2000). This could be confirmed by my own pilot data (unpublished). Hence, the location of an object can be meaningful in building expectations concerning the unfolding of an observed action: objects that are needed in order to perform an action must be easily reached and grasped for subsequent efficient usage. In a similar vein, it was found that grip facilitation depends on an object’s availability for manipulation, which points to an interaction between the object’s pragmatic properties and the action goal implied by the contextual setting (Kalénine, Shapiro, Flumini, Borghi, & Buxbaum, 2014). Hence, I argue that highly congruent COs located close to and to the right of the actor are perceived not only in relation to the observed action but also to the performing actor, i.e., as a potential future target of the actor.

4.3.2 COs located in the (virtual) peripersonal space of an observer

In a behavioural variation of Study 3 (unpublished data), using the exact same locations for COs, we found the shortest RTs for the recognition of actions including highly congruent COs that were either close-right or far-left with respect to the actor. Note that far-left positions for the actor are equivalent to close-right for the observer (180 degrees rotated). Against this backdrop, I reasoned that depending on the location of the COs, action observers might also establish their own (virtual) action space by perceiving COs not only with reference to the observed action but also with reference to themselves. A subsequent behavioural study assessing reachability scores for the COs' locations from an observing perspective[19] revealed that action observers are able to represent a CO's location with respect to themselves and to the actor (Study 2). Using these reachability scores to model the BOLD response in a reanalysis of the fMRI dataset from Study 1, activity in the premotor, inferior parietal, cingulate, and insula cortices parametrically increased with the CO location's (virtual) reachability for the observer. I argue that action observers are indeed furnished with their own action possibilities when COs are presented proximally to them (see Study 2). Two concepts are of importance here: the concept of *peripersonal space* (PPS; Rizozolatti et al., 1997; Bufacchi & Ianetti, 2018) and the concept of *affordance* (Gibson, 1977; 1979).

According to the concept of PPS (Rizzolatti et al., 1997; Bufacchi & Ianetti, 2018), action observers should represent the space proximal to the actor and the objects therein with reference to the actor. The same logic applies to objects proximal to the observer. If one considers that the action observers separated the table into an operational space for the observer and an operational space for the actor, an explanation for the missing main effect of location ergonomics in Study 1 arises: both COs located proximally to the observer and COs located proximally to the actor have the same proximity in their respective operational space. The effects of COs located proximally to the observer might thus have levelled out effects of COs located proximally to the actor. In a similar vein, the *shared action space* — an area reachable by both (the actor and the observer) — did not reveal any significant activation pattern compared with the action spaces of the actor and the observer. This suggests that this space is

---

[19] To assess these scores, action observers were presented with the same videos as were used in the fMRI study.

not actually represented as a *shared action space* but as a space that instead remains unused by both the actor and the observer, compared with their respective allocated action zones.

In addition, the spatial relation between objects and agents is meaningful as it was found to impact upon object *affordances*. Macaque and human studies show that visually presented objects can automatically trigger the representation of an action, provided that they are located within reaching space, i.e., PPS (cf. Maranesi et al., 2014). This should apply to objects outside the PPS (and near the main acting hand) of the observer but inside of that of the actor (Brozzoli et al., 2013, 2014; see also Costantini et al., 2011) and vice versa. In that sense, COs located within the PPS of action observers should trigger additional CO-related action representations, drawing on the IPL and PMv (cf. Schubotz et al., 2014). This can be confirmed by the findings from Study 2.[20] In support of this argument, Brozzoli et al. (2013) have conducted an interesting fMRI experiment, revealing that the PMv responds to both objects close to one's own hand and objects close to another person's hand. On a more general level, these findings are very plausible, since action observation and execution (and imitation) are sustained by overlapping brain regions (Caspers et al., 2010).

It remains unclear whether action observers exploit their potential action possibilities with regard to the observed action or even with regard to the actor (i.e., with regard to a potential *interaction*). It cannot be ruled out that results can be solely traced back to the convenience of a CO's location, irrespective of the observed action. However, as

i) it was the participants' task to watch and understand the observed action, and
ii) no a priori attention was drawn towards the COs,

---

[20] Note that the elicited brain activations for the observer-referenced processing of COs partly corresponded to a network of brain regions found to relate to processing sensory events within the PPS, including the supplementary motor area (SMA; bordering the midcingulate cortex), the anterior insula, the PPC, premotor areas and the cerebellum (cf. Grivaz, Blanke, & Serino, 2017).

I speculate that the observed action would be considered by action observers when exploiting their potential action possibilities. In this case, several scenarios are conceivable when action observers encounter a CO in their PPS:

i) They perceive the CO as part of an overarching action goal of the actor. Thus, their brain engages into a potential reach, grasp and handover movement to “help” the actor.

ii) They perceive the CO as part of a *joint* overarching action goal — a complementary task performance with a joint action outcome (cf. Newman-Norlund, Noordzij, Meulenbroek, & Bekkering, 2007).

iii) The afforded action provided by the CO in the observer’s PPS is independent of the observed action’s potential outcome.

While i) could be the case irrespective of the COs goal affinity, ii) and iii) should depend on it. Here, ii) would be triggered for either highly congruent or lowly congruent COs, whereas iii) would rather be triggered for lowly congruent or incongruent COs.[21] Nevertheless these are speculations, which is why further studies are required to explicitly investigate this issue in depth. Since the IFG was co-activated with the COs’ increased convenience for the observer, goal representations may have played a central role.

To sum up, the location of the CO is meaningful as it can provide either the actor or the observer with possible actions. In the case of the observer, these potential actions might — but do not have to — relate to an overarching action goal of the observed action. In the case of the actor, COs are suggested to be processed with respect to the observed action’s anticipated outcome. In the next section, I will outline the view that the processing of CO information is driven by its informative value in relation to the observed action scene.

---

[21] If i) is dependent on the COs’ goal affinity, it should primarily be triggered for highly congruent COs.

4.4 Using contextual information: a matter of relevance?

In the previous sections, I demonstrated that CO information impacts action observation under several circumstances: either when COs strongly match or strongly contradict the observed action, or when their location additionally implies subsequent usage by the actor or the observer. Under all these circumstances, contextual information provides relevant cues for action prediction and hence goal anticipation. Is the informative value of contextual information (i.e. its relevance) of an observed action the critical factor that triggers its use in such predictive processes?

There is ample evidence from studies in children and adults that action observers do not only interpret actions post hoc, but actively predict how they will continue (e.g., Flanagan & Johansson, 2003; Falck-Ytter, Gredebäck, & von Hofsten 2006; Uithol & Paulus, 2014; Hrkać et al., 2014; Wurm et al., 2014). Support for action prediction also comes from EEG studies and the N400 amplitude, which was found to be elicited by meaningful actions (i.e., when people intend to employ objects in a meaningful manner) but not by meaningless actions. Moreover, this effect is driven by the unexpectedness of information, rather than by simple congruence or incongruence to the target stimulus (cf. Amoruso et al., 2013), underlining that what is meaningful can change, depending on the context of a given situation.

Using fMRI, it has been shown that areas that are traditionally assigned to the AON were also found to contribute to the prediction of abstract stimuli (Schubotz & Cramon, 2004; Schubotz, 2007), indicating the AON's general role in predictive tasks. I propose that CO information is selectively used in order to predict an action's unfolding.

In the following sections, I will discuss the results of my studies in light of a predictive coding framework for action prediction (cf. Kilner et al., 2007; Kilner, 2011). I will lay out how the integration of relevant contextual information fits with current theories of predictive processing. I will further put a special emphasis on the IFG as an area that, during action observation, selects *relevant* contextual information to update or revise currently operating predictive models of action outcomes.

4.4.1 Using context in light of action prediction

In this section, I will discuss how context can support the prediction of actions and thereby potentially enable the specification or revision of generative models of action prediction. It has been shown that perception is facilitated by top-down influences, i.e., predictions (cf. Trapp & Bar, 2015). These predictions are significantly influenced by contextual information (cf. Pezzulo & Cisek, 2016). In real life, input to the visual system is often noisy: that is, objects and actions appear with occlusion, interruption, sub-optimal illumination and so forth. In these cases, (compatible) contextual information has been shown to provide action observers with top-down cues that facilitate the recognition of objects (cf. Bar, 2004) or actions (cf. Wurm et al., 2017), by constraining what to expect given a specific contextual setting. Initially, the enormous amount of sensory information already creates competition vis-à-vis where to look in a given scene. Selective attention has proven an elegant human mechanism for selecting the relevant information from the perceptual environment in order to accomplish a task (Desimone and Duncan, 1995). However, the competition does not end with this initial selection of what to attend to (e.g., the action). Instead, in the case of objects, one particular object might map onto different neuronal patterns stored in memory, i.e., different generative models (cf. Clark, 2013). In a similar vein, the *affordance competition hypothesis* (Cisek, 2007) postulates that several actions are specified in parallel by the brain, based on currently available information from the environment (i.e., affordances). In the meantime, further information is collected to bias the perception of these competing actions in favour of selecting a single representation or action outcome (see also Pezzulo & Cisek, 2016). Again, the information needed to resolve this competition is typically provided by the environment of the observed action, i.e., it is context-sensitive. Pezzulo & Cisek (2016) have proposed an elegant way of conceptualizing action prediction by extending the affordance competition hypothesis. In their model, termed the *hierarchical affordance competition hypothesis*, they suggest that brains are continuously engaged in generating predictions (e.g., about future opportunities) rather than just reacting to already available affordances. Thereby, the brain is able to link different levels of abstraction and to bias the selection of immediate actions (e.g., cutting a tomato) according to the predicted long-term opportunities they make possible (e.g., making a salad). It is suggested that these processes follow a nested cascade of expectations that prescribe the next affordance to be exploited (CO information) without necessarily specifying how exactly the lower-level motor acts should be executed. This autonomy leads to continuous competition at the lower levels of the hierarchy between alternative ways to specify the demands of the higher levels. Not only can (higher-level) decisions about action plans bias immediate action selection, but the results of the competition between affordances at any low level of the hierarchy (e.g., a highlighted scientific paper) can influence choices at higher hierarchical levels (e.g., cutting a tomato) by creating a residual prediction error that cannot be minimized without changing a long-term plan (e.g., preparing a snack for learning). Ultimately, the selected action can be executed through continuous feedback control.[22]

Visual context provides the necessary broadness and degrees of abstraction to serve as such a top-down signal, able to favour a large set of ongoing perceptual processing steps towards a single interpretation of the perceived visual input, i.e., a long-term action goal. In this sense, context provides two functions: initially, it helps the observer selectively attend to what is relevant, while in a second step, it helps resolve the competition between various similar

---

[22] Although primarily related to action planning and execution, the hierarchical affordance competition hypothesis it is also applicable to action prediction during the observation of an action.

activated representations (Trapp & Bar, 2015) by providing a (temporal) more abstract or distant long-term goal. Thus, contextual information can indeed impact action observation at different levels of the action hierarchy, from immediate consequences up to higher-level action goals.

We have limited knowledge about how these different aspects relating to action prediction operate on a temporal scale, i.e., at which points they may impact each other. With regard to my studies, the action, the CO and the room category were all presented simultaneously and were likely processed largely in parallel. Representations of these information types are cross-linked in the brain and most probably thus impact upon each other. In this context, one can only speculate about several possible options regarding the temporal aspects of prediction formation:

When observing somebody cutting a tomato, one is presented with different types of information simultaneously, which lead to the selection of a specific generative model (here, of a short-term goal, e.g., a cut tomato) in a bottom-up manner. Thus, the generative model for "cutting a tomato" is initially activated by the presentation of core information, the internal goal of the actor (study-task), and the gist of the observed scene. It is conceivable that this generative model in turn biases (to some extent) a set of potential generative models at higher levels of the hierarchy, i.e., potential long-term goals.

In this scenario an additional highly congruent or incongruent CO provides a highly informative contextual bottom-up cue for what to expect next, i.e., a predictive cue for the unfolding of the action. The moment in which this CO is incorporated into the observed action could thus be regarded as a *decision point* (cf. Botvinick & Bylsma, 2005). The selection of the correct action at such a decision point should be benefited by screening the immediate environment of an observed action. Hence further affordances of the environment can be exploited, i.e., CO information. In the case of a highly congruent CO, the currently operating predictive model can thus be extended (or specified) with respect to a specific overarching action goal, e.g., preparing a salad. In the case of an incongruent CO, a more distal goal has to be anticipated that links “a cut tomato” with the incongruent CO information, leading to a revision of the predicted model. Lowly congruent COs do not provide any valuable cues for what to expect next, as they are only processed as part of the room category (i.e., they are processed as part of the gist of the scene). Hence, no higher-level generative model is activated, i.e., no increased activation pattern is observed.

Another possibility could be that the action and the CO both simultaneously trigger generative models (at the same level of the action hierarchy). In the case of incongruent COs, these will be opposing models, and the consequent conflict can be resolved by retrieving a new overarching action goal, leading to the activation of a higher-level generative model. In the case of highly congruent COs, both action and CO would concurrently activate related generative models (e.g., cutting a tomato, cutting a cucumber) that amplify each other and directly lead to the activation of a higher-level generative model (e.g., preparing a salad) that specifies and predicts upcoming action steps.

In support of such speculations, I found enhanced activation for highly congruent and incongruent COs in several AON nodes, potentially reflecting the triggering of representations of the action steps necessary to achieve a (specific) overarching action goal. As these predicted action steps are never perceived by the action observers, the engagement of the corresponding brain areas should indeed be high (cf. Waszak et al., 2012) and could be particularly demonstrated by increased activation of the IFG in several circumstances: activity in the bilateral IFG (BA 44, 45, 47) significantly increased for both incongruent and highly congruent COs, compared to lowly congruent COs (Study 3). The left IFG was significantly engaged when COs were located proximally to action observers (Study 2) and so was the bilateral IFG, when highly congruent COs were presented in the preferred action zone of an observed actor (Study

1). I will discuss these results in detail in the next section, when proposing that the IFG does not simply reflect integration attempts of more or less congruent contextual information. Instead I suggest that the IFG reflects the informative impact of COs on observed actions, by using contextual information to supply and update currently operating predictive models. I argue that previous findings concerning the IFG's function can be reconciled with this proposition.

4.4.2 The role of the IFG during action observation reconsidered

I will start by discussing issues of lateralization in the IFG, followed by a recapitulation and discussion of the present findings in light of previous findings concerning the IFG. I will close this section by proposing a revised role for the IFG.

The IFG has been associated with effortful contextual integration across different domains (Poldrack et al., 1999; Smirnov et al., 2014; van Schie, Toni, & Bekkering, 2006). Most studies emphasize or only discuss the left IFG (referred to as *Broca's area*), even when their results reveal bilateral activation (e.g., Kröger et al., 2012). While the left IFG has consistently been associated with the processing of semantics, many studies reporting on the right IFG have investigated response inhibition, e.g., go/no-go tasks (Levy & Wagner, 2011). However, inhibition is not in conflict with selection. In order to select among competing representations — a function that has been proposed for the left IFG (Badre & Wagner, 2007) — one has to select in favour of one representation while at the same time inhibiting others (cf. Levy & Wagner, 2011). For this mechanism to be effective, it is crucial to be able to consider the possible responses concurrently: an ability also referred to as *contextual control*. Contextual control was found to activate the bilateral IFG (Koechlin & Summerfield, 2007). The lateralization of IFG activation might depend on the type of stimulus, with language being processed in the left hemisphere and visuo-spatial stimuli in the right hemisphere (e.g., Golby et al., 2001; Walter, Bretschneider, Grön & Zurowski, 2003). In support of this proposition, it is worth noting that paradigms investigating motor inhibition mostly make use of visuo-spatial stimuli, whereby the verbal aspects of linguistics are less relevant (cf. Levy and Wagner, 2011), e.g., the location of letters matters but not their identity. In contrast, the processing of linguistic stimuli activates the left IFG (Badre & Wagner, 2007). The stimuli used in this thesis and in previous action observation paradigms (Wurm & Schubotz, 2012; Hrkać et al., 2014; Hrkać et al., 2015; Wurm et al., 2014) are non-verbal, but relatively easy to verbalize.[23] Such stimuli have often been found to lead to bilateral activation in the PFC (e.g., Golby et al., 2001; Hrkać et al., 2014). Thus, it is conceivable that the left and the right IFG reflect similar processes that are lateralized depending on the type of stimulus. Therefore, I do not differentiate between left and right IFG activations.

In my studies, activation of the IFG was observed when contextual information had the potential to enrich the interpretation of an observed action. In Study 1, I found increased activation of the bilateral IFG (BAs 44 and 45) when highly matching COs were presented close to and to the right of the actor. As such, CO information could be used in order to anticipate the observed action's outcome. In Study 2, left IFG (BA 44) activation increased parametrically with an increase in the convenience of a CO's location for the action observer, irrespective of the goal affinity. The location of the COs was thus meaningful as it provided action observers with "their own" action possibilities. When the significance threshold was lowered, also activation of the right IFG (BA 44) was revealed. In Study 3, I observed a significant increase of activity for both highly congruent and incongruent COs in all IFG compartments (BAs 44, 45 and 47). However, incongruent COs imposed an additional demand on BA 47 when compared to highly congruent COs. I argue that these results imply that IFG activation reflects the informative impact CO information can have on observed actions.

---

[23] Note that in the present studies, it was the participants' task to respond to verbal descriptions of the actions.

Previous accounts of the IFG's function (Badre & Wagner, 2007; Koechlin & Summerfield, 2007; Uddén & Bahlmann, 2012; Petrides, 2005) have proposed that it is hierarchically organized according to a functional stepwise gradient along the rostro-caudal axis, whereby top-down control is exerted from anterior to posterior regions (Koechlin et al., 2003). The essence here is that the more posterior the area (BA 44), the more it contributes to constrain the immediate action requirements or options. More anterior sites (BAs 45 and 47), however, are more and more content-independent and associated with high-level goals (Badre, Poldrack, Pare-Blagoev, Insler, & Wagner, 2005; Badre & D'Esposito, 2009; cf. Buckner, 2003). This hierarchical organization situates the IFG as a good candidate for contributing to the prediction of actions at different levels of the action hierarchy.

Correspondingly, the IFG has been associated with the processing of action goals in many action observation paradigms. Schubotz and Cramon (2009) found the (left) IFG to be engaged whenever the goals of an observed action changed. In a similar vein, IFG activation has been found to increase whenever an actor performs unrelated action steps in succession, implying missing goal coherence (Hrkać et al., 2014). Attenuation of the IFG has been observed when goals become predictable, i.e., when observing goal-coherent episodes of action steps (Wurm et al., 2014). Note that in all these circumstances, the processing of the action goals was not task-relevant. Rather, the IFG engagement seems to reflect an autonomic process.

These results already imply that the IFG does not simply reflect demands to integrate more or less compatible contextual information in the observed action, but that it is goal-sensitive. For example, in Hrkać et al. (2014), in which every action step was performed by a new actor — producing a situation in which goal coherence was present but the actor was incoherent — no increased activation in the IFG could be observed. The same pattern was revealed for lowly congruent COs (Studies 1 and 3). While the situation in the former study could be resolved by the participants' potential assumption that they were observing a joint-action performance, lowly congruent COs could be interpreted as part of the room category. Despite seemingly conflicting actor information or action–object associations, IFG activity did not increase. Moreover, my studies did not reveal a parametric increase in IFG activation with increasing incongruence, i.e., the lowest IFG engagement for highly congruent COs. Instead, the presence of highly congruent COs increased activity in all IFG compartments, as did incongruent COs. Hence, the IFG appears to take account of contextual information that specifies and/or enriches the interpretation of an observed action.

Based on previous findings, one might have expected activation of the IFG to reflect (only) missing goal coherence. In line with this proposition, the IFG was engaged when an event in a story could not be predicted because previous predictive information was missing (Virtue, Haberman, Clancy, Parrish, & Beeman, 2006). Again, the results of the present studies imply an extended role for the IFG: while missing goal coherence is triggered by incongruent COs (Study 3), this might apply to COs that are located proximally to action observers (Study 2) but does not apply to highly congruent COs (Study 1 and 3). All of the above-mentioned findings can however be reconciled when considering that the IFG reflects the informational impact of contextual information on observed actions. According to this suggestion, in cases of missing goal coherence, the IFG would be engaged, reflecting the effort to retrieve an overarching action goal when presented with conflicting action-related information. This would draw especially on BA 47, due to its ascribed function of controlled retrieval (Badre & Wagner, 2007), i.e., a top-down process activating goal-relevant knowledge, especially in the face of contradicting representations. Studies relating to missing goal coherence indeed involved the selective engagement of BA 47 (Wurm & Schubotz, 2012; Hrkać et al., 2014). The same pattern of activation was found for incongruent COs in Study 3. Although all IFG compartments were recruited for both highly congruent and incongruent COs, activation of BA 47 selectively increased for the presence of incongruent as opposed to highly congruent COs. Thus, *increased* *activation* of BA 47 in response to incongruent vs highly congruent COs can be explained by the *increased demand* to retrieve an action outcome when confronted with strongly opposing action-related information. More specifically, incorporating both incongruent COs and the observed action under an overarching goal requires a much higher level of abstraction, evoked by the strength of the CO's association with an incompatible room category and hence with actions associated with this room category.

In cases of highly congruent information, the IFG potentially reflects a process of specifying the interpretation of an observed action by predicting an overarching action goal that incorporates a continuation of the currently observed object manipulation. In both cases, the interpretation of an observed action is enriched and the currently operating predictive models can be updated and specified.[24] The same logic applies to COs in observer-convenient locations. These COs provide relevant information about what might happen next and thereby support the prediction of upcoming events. Finally, as the prediction of others' behaviour is of vital importance (Prinz, 2006), I propose that based on the novel insights revealed by the results of the present studies, the engagement of the IFG might indeed reflect a selective use of contextual information that supports the prediction of action outcomes.

4.5 Future directions

In the following, a number of ideas for further studies that would strengthen the findings and interpretations of the current studies are presented.

---

[24] In that sense, it is conceivable that the anterior IFG (BA 45/47) selectively uses relevant contextual information to form an overarching action goal, in concordance with temporal regions (stored learned target-context association), while the posterior IFG (BA 44/45) generates a sequence of action steps to achieve this goal, which is in turn anticipated by the adjacent PMv as well as the IPL This proposal is roughly coinciding to a model proposed by Amoruso et al. (2013).

With regard to Study 3, it cannot be completely ruled out that the results for incongruent COs could be traced back to an object–room conflict. Hence, the possibility remains that the presentation of the action itself was not necessary to elicit the currently observed pattern of activation. I suggest a modified version of Study 3 in which participants are also presented with videos containing the exact same room information and CO information, but without the actor performing an action. Brain activations for trials containing the actor could be compared with trials that do not, thereby accounting for possible activations that were due solely to object–room associations. This would also account for related issues in Study 2, whereby COs located proximally to the action observers might suffice in order to elicit the currently observed pattern of activation. Regarding Study 2, it would also be interesting to investigate how the goal affinity of the conveniently located COs (vis-à-vis the observer) might contribute to the processing of action goals. By investigating the differential engagement of the three IFG compartments in response to varying the goal affinity of these COs, one might be able to draw conclusions about the quality of action goals (as proposed in Section 4.3.2).

Finally, I propose a study that puts the main emphasis on investigating the revised role of the IFG, i.e., selectively using contextual information to predict action goals at different levels. Using eye tracking and fMRI simultaneously, action observers would initially be presented with action videos (3s) that depict actions involving highly congruent COs, similar to Studies 1 and 3. However, comparable to the manipulation of Hrkać et al. (2014) and Wurm et al. (2014), these videos would belong to a sequence of action steps (i.e., an episode), whereby each action step would be depicted in a short video (3s). Similar to Hrkać and colleagues (2014), these episodes would either be goal coherent or not. The most important factor is that after the presentation of the initial action step featuring a CO (that is not used), that CO would be incorporated by the actor into the action during the unfolding of the action episode (goal-coherent episode). Hence, the COs would receive a shift in their meaning, from contextual information to core information. If the IFG selectively uses context to predict action outcomes, an increase of activation in the IFG should be observed after the initial video. As soon as the action unfolds and the CO becomes integrated within the observed action-episode, IFG activity should however decrease as a function of predictability. On the other hand, where COs are not incorporated into the subsequent action steps of an observed episode — even though such an incorporation was implied by the presence of the CO in the first video — no attenuation should be observed. Using eye tracking could be an elegant way to account for the selective use of CO information. In that case, one would expect participants to make saccades and fixations away from the action and towards the CO information. Using EEG (and investigating the N400) during the above described study could also provide promising insights for two reasons: i) it potentially enables a deeper understanding of the temporal aspects of prediction formation and ii) it might corroborate the assumption that both highly congruent and incongruent COs provide (surprisingly) meaningful information for action prediction.

.

5. Conclusion

The successful prediction of others' actions is an example of how fascinating human cognitive abilities can be. Previous research has suggested that action observers consider multiple sources of information in an attempt to successfully predict action outcomes, even at the cost of implausible expectations.

This thesis built upon these findings by investigating under which circumstances contextual object information is used to constrain an observer's expectations, thereby providing novel insights into how contextual information is selectively used in order to enhance action prediction. In light of these results, the IFG's current role should be extended to acknowledge the area as one that operates between core and contextual information, and selects contextual information in a way that enriches the interpretation of observed actions. My studies indicate that the IFG is sensitive to the informative impact of contextual cues as it might selectively use context to infer action goals at higher levels of the action hierarchy. These processes were not task-relevant, underlining the inherited predictive nature of the human brain. In this sense, real-life predictions might indeed be governed by both expectation-constraining and expectation-generating computations.

Finally, this thesis provides benchmarks for the study of impairments related to autism spectrum disorder, since a range of these impairments have been associated with a lack of contextual sensitivity and a bias towards local processing, i.e., the failure to use context to guide behaviour.

References

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature neuroscience*, *11*(9), 1109.

Albright, T. D., & Stoner, G. R. (2002). Contextual influences on visual processing. *Annual review of neuroscience*, *25*(1), 339-379.

Amoruso, L., Couto, J. B., & Ibanez, A. (2011). Beyond Extrastriate Body Area (EBA) and Fusiform Body Area (FBA): context integration in the meaning of actions. *Frontiers in Human Neuroscience*, *5*, 124.

Amoruso, L., Finisguerra, A., & Urgesi, C. (2018). Contextualizing action observation in the predictive brain: Causal contributions of prefrontal and middle temporal areas. *Neuroimage*, *177*, 68-78.

Amoruso, L., Gelormini, C., Aboitiz, F. A., González, A., Manes, F., Cardona, J., & Ibanez, A. (2013). N400 ERPs for actions: building meaning in context. *Frontiers in human neuroscience*, *7*, 57.

Amoruso, L., Sedeño, L., Huepe, D., Tomio, A., Kamienkowski, J., Hurtado, E., ... & Manes, F. (2014). Time to tango: expertise and contextual anticipation during action observation. *Neuroimage*, *98*, 366-385.

Amoruso, L., & Urgesi, C. (2016). Contextual modulation of motor resonance during the observation of everyday actions. *Neuroimage*, *134*, 74-84.

Amunts, K., & Zilles, K. (2012). Architecture and organizational principles of Broca's region. *Trends in cognitive sciences*, *16*(8), 418-426.

Bach, P., Nicholson, T., & Hudson, M. (2014). The affordance-matching hypothesis: how objects guide action understanding and prediction. *Frontiers in human neuroscience, 8*, 254.

Badre, D., & D'esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical?. *Nature Reviews Neuroscience*, *10*(9), 659.

Badre, D., Poldrack, R. A., Pare-Blagoev, E. J., Insler, R. Z., & Wagner, A. D. (2005). Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron, 47*, 907-918.

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, *45*(13), 2883-2901.

Balser, N., Lorey, B., Pilgramm, S., Stark, R., Bischoff, M., Zentgraf, K., ... & Munzert, J. (2014). Prediction of human actions: Expertise and task-related effects on neural activation of the action observation network. *Human brain mapping*, *35*(8), 4016-4034

Bar, M. (2004). Visual objects in context. *Nature Reviews Neuroscience*, *5*(8), 617.

Bar, M., Tootell, R. B. H., Schacter, D. L., Greve, D. N., Fischl, B., Mendola, J. D., Rosen, B. R., Dale, A. M. (2001). Cortical mechanisms specific to explicit visual object recognition. *Neuron*. 29(2): 529–35.

Bar, M., & Ullman, S. (1996). Spatial context in recognition. *Perception*, *25*(3), 343-352.

Barenholtz, E. (2014). Quantifying the role of context in visual object recognition. *Visual Cognition, 22*(1), 30-56.

Beauchamp, M. S., Lee, K. E., Haxby, J. V., & Martin, A. (2002). Parallel visual motion processing streams for manipulable objects and human movements. *Neuron*, *34*(1), 149-159.

Beauchamp, M. S., & Martin, A. (2007). Grounding object concepts in perception and action: evidence from fMRI studies of tools. *Cortex*, *43*(3), 461-468.

Bedny, M., & Caramazza, A. (2011). Perception, action, and word meanings in the human brain: the case from action verbs. *Annals of the New York Academy of Sciences*, *1224*(1), 81-95.

Bekkering, H., Wohlschlager, A., & Gattis, M. (2000). Imitation of gestures in children is goal-directed. *The Quarterly Journal of Experimental Psychology: Section A*, *53*(1), 153-164.

Biederman, I., Glass, A. L., & Stacy, E. W. (1973). Searching for objects in real-world scenes. *Journal of experimental psychology*, *97*(1), 22.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: Detecting and judging objects undergoing relational violations. *Cognitive psychology*, *14*(2), 143-177.

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? A critical review and meta-analysis of 120 functional neuroimaging studies. *Cerebral Cortex*, *19*(12), 2767-2796.

Bookheimer, S. (2002). Functional MRI of language: new approaches to understanding the cortical organization of semantic processing. *Annual review of neuroscience*, *25*(1), 151-188.

Borghi, A. M., Caramelli, N., & Setti, A. (2005). Conceptual information on objects' locations. *Brain and language*, *93*(2), 140-151.

Botvinick, M. M., & Bylsma, L. M. (2005). Distraction and action slips in an everyday task: Evidence for a dynamic representation of task context. *Psychonomic bulletin & review*, *12*(6), 1011-1017.

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 556.

Brozzoli, C., Gentile, G., Bergouignan, L., & Ehrsson, H. H. (2013). A shared representation of the space near oneself and others in the human premotor cortex. *Current Biology*, *23*(18), 1764-1768.

Brozzoli, C., Ehrsson, H. H., & Farnè, A. (2014). Multisensory representation of the space near the hand: from perception to action and interindividual interactions. *The Neuroscientist*, *20*(2), 122-135.

Bryden, P. J., Pryde, K. M., & Roy, E. A. (2000). A performance measure of the degree of hand preference. *Brain and cognition*, *44*(3), 402-414.

Buckner, R. L. (2003). Functional–anatomic correlates of control processes in memory. *Journal of Neuroscience*, *23*(10), 3999-4004.

Bufacchi, R. J., & Iannetti, G. D. (2018). An action field theory of peripersonal space. *Trends in cognitive sciences*.

Buxbaum, L. J., Johnson-Frey, S. H., & Bartlett-Williams, M. (2005). Deficient internal models for planning hand–object interactions in apraxia. *Neuropsychologia*, *43*(6), 917-929.

Buxbaum, L. J., & Kalénine, S. (2010). Action knowledge, visuomotor activation, and embodiment in the two action systems. *Annals of the New York Academy of Sciences*, *1191*(1), 201-218.

Buxbaum, L. J., Kyle, K. M., Tang, K., & Detre, J. A. (2006). Neural substrates of knowledge of hand postures for object grasping and functional object use: Evidence from fMRI. *Brain Research*, *1117*(1), 175-185.

Byrne, R. W., & Russon, A. E. (1998). Learning by imitation: A hierarchical approach. *Behavioral and brain sciences*, *21*(5), 667-684.

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2004). Action observation and acquired motor skills: an FMRI study with expert dancers. *Cerebral cortex*, *15*(8), 1243-1249.

Carr, L., Iacoboni, M., Dubeau, M. C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: a relay from neural systems for imitation to limbic areas. *Proceedings of the national Academy of Sciences*, *100*(9), 5497-5502.

Caspers, S., Zilles, K., Laird, A. R., & Eickhoff, S. B. (2010). ALE meta-analysis of action observation and imitation in the human brain. *Neuroimage*, *50*(3), 1148-1167.

Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the dorsal stream. *Neuroimage*, *12*(4), 478-484.

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1485), 1585-1599.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181-204.

Costantini, M., Ambrosini, E., Tieri, G., Sinigaglia, C., & Committeri, G. (2010). Where does an object trigger an action? An investigation about affordances in space. *Experimental brain research*, *207*(1-2), 95-103.

Costantini, M., Committeri, G., & Sinigaglia, C. (2011). Ready both to your and to my hands: mapping the action space of others. *PloS one*, *6*(4), e17923.

Craighero, L., Fadiga, L., Umiltà, C. A., & Rizzolatti, G. (1996). Evidence for visuomotor priming effect. *Neuroreport*, *8*(1), 347-349.

Craighero, L., Fadiga, L., Rizzolatti, G., & Umilta, C. (1998). Visuomotor priming. *Visual Cognition, 5*(1-2), 109–125.

Creem-Regehr, S. H. (2009). Sensory-motor and cognitive functions of the human posterior parietal cortex involved in manual actions. *Neurobiology of Learning and Memory*, *91*(2), 166-171.

Creem-Regehr, S. H., Gagnon, K. T., Geuss, M. N., & Stefanucci, J. K. (2013). Relating spatial perspective taking to the perception of other's affordances: Providing a foundation for predicting the future behavior of others. *Frontiers in Human Neuroscience*, *7*, 596.

Cross, E. S., Hamilton, A. F. D. C., Kraemer, D. J., Kelley, W. M., & Grafton, S. T. (2009). Dissociable substrates for body motion and physical experience in the human action observation network. *European Journal of Neuroscience*, *30*(7), 1383-1392.

Csibra, G., 2007. Action mirroring and action interpretation: An alternative account. In: P. Haggard, Y. Rosetti, & M. Kawato (Eds.), *Sensorimotor Foundations of Higher*

Cognition. Attention and Performance XXII (pp. 435-459). Oxford: Oxford University Press.

Csibra, G., & Gergely, G. (2007). ‘Obsessed with goals’: Functions and mechanisms of teleological interpretation of actions in humans. *Acta psychologica*, *124*(1), 60-78.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science*, *15*(8), 559-564.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, *18*(1), 193-222.

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: a neurophysiological study. *Experimental brain research*, *91*(1), 176-180.

di Pellegrino, G., & Làdavas, E. (2015). Peripersonal space in the brain. *Neuropsychologia*, *66*, 126-133.

Dilks, D. D., Julian, J. B., Paunov, A. M., & Kanwisher, N. (2013). The occipital place area is causally and selectively involved in scene perception. *Journal of Neuroscience*, *33*(4), 1331-1336

Downing, P. E., Chan, A. Y., Peelen, M. V., Dodds, C. M., & Kanwisher, N. (2005). Domain specificity in visual cortex. *Cerebral cortex*, *16*(10), 1453-1461.

Dum, R. P., & Strick, P. L. (2002). Motor areas in the frontal lobe of the primate. *Physiology & behavior*, *77*(4-5), 677-682.

Epstein, R. A., Bar, M., & Kveraga, K. (2014). Neural systems for visual scene recognition. *Scene vision*, 105-134.

Fagg, A. H., & Arbib, M. A. (1998). Modeling parietal–premotor interactions in primate control of grasping. *Neural Networks*, *11*(7-8), 1277-1303.

Falck-Ytter, T., Gredebäck, G., & von Hofsten, C. (2006). Infants predict other people's action goals. *Nature neuroscience*, *9*(7), 878.

Fiebach, C. J., & Schubotz, R. I. (2006). Dynamic anticipatory processing of hierarchical sequential events: a common role for Broca's area and ventral premotor cortex across domains?. *Cortex*, *42*(4), 499-502.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. Mcgraw-Hill Book Company.

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769.

Fogassi, L., Ferrari, P. F., Gesierich, B., Rozzi, S., Chersi, F., & Rizzolatti, G. (2005). Parietal lobe: from action organization to intention understanding. *Science*, *308*(5722), 662-667.

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews neuroscience*, *11*(2), 127.

Friston, K., Mattout, J., & Kilner, J. (2011). Action understanding and active inference. *Biological cybernetics*, *104*(1-2), 137-160.

Frith, C., & Frith, U. (2005). Theory of mind. *Current biology*, *15*(17), R644-R645.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531-534.

Frith, U., & Happé, F. (1994). Autism: Beyond “theory of mind”. *Cognition*, *50*(1-3), 115-132.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, *2*(12), 493-501.

Geyer, S., Matelli, M., Luppino, G., & Zilles, K. (2000). Functional neuroanatomy of the primate isocortical motor system. *Anatomy and embryology*, *202*(6), 443-474.

Gibson, J.J. (1977). The theory of affordances. In R.E. Shaw & J. Bransford (Eds.), Perceiving, Acting, and Knowing. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gibson, J. J. (1979). The theory of affordances. The ecological approach to visual perception.

Golby, A. J., Poldrack, R. A., Brewer, J. B., Spencer, D., Desmond, J. E., Aron, A. P., & Gabrieli, J. D. (2001). Material-specific lateralization in the medial temporal lobe and prefrontal cortex during memory encoding. *Brain*, *124*(9), 1841-1854.

Gold, B. T., Balota, D. A., Jones, S. J., Powell, D. K., Smith, C. D., & Andersen, A. H. (2006). Dissociation of automatic and strategic lexical-semantics: functional magnetic resonance imaging evidence for differing roles of multiple frontotemporal regions. *Journal of Neuroscience*, *26*(24), 6523-6532.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in neurosciences*, *15*(1), 20-25.

Gowen, E. (2012). Imitation in autism: why action kinematics matter. *Frontiers in integrative neuroscience*, *6*, 117.

Grafman, J., 2002. The structured event complex and the human prefrontal Cortex. In: Stuss DT. & Knight RT. Principles of frontal lobe function. Oxford University Press, New York.

Grafton, S. T., & Hamilton, A. F. D. C. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human movement science*, *26*(4), 590-616.

Graziano, M. (2006). The organization of behavioral repertoire in motor cortex. *Annu. Rev. Neurosci.*, *29*, 105-134.

Grefkes, C., & Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy*, 207(1), 3–17.

Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). The lateral occipital complex and its role in object recognition. *Vision research*, *41*(10-11), 1409-1422.

Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzchak, Y., & Malach, R. (1999). Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron*, *24*(1), 187-203.

Grivaz, P., Blanke, O., & Serino, A. (2017). Common and distinct brain regions processing multisensory bodily signals for peripersonal space and body ownership. *Neuroimage*, *147*, 602-618.

Grossman, E., Donnelly, M., Price, R., Pickens, D., Morgan, V., Neighbor, G., & Blake, R. (2000). Brain areas involved in perception of biological motion. *Journal of cognitive neuroscience*, *12*(5), 711-720.

Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and brain sciences*, *27*(3), 377-396.

Hamilton, A. F. D. C. (2009). Research review: Goals, intentions and mental states: Challenges for theories of autism. *Journal of Child Psychology and Psychiatry*, *50*(8), 881-892.

Hamilton, A. F. D. C., & Grafton, S. T. (2006). Goal representation in human anterior intraparietal sulcus. *Journal of Neuroscience*, *26*(4), 1133-1137.

Hamilton, A. F., & Grafton, S. T. (2007). The motor hierarchy: from kinematics to goals and intentions. *Sensorimotor foundations of higher cognition*, *22*, 381-408.

Hamilton, A. D. C., Joyce, D. W., Flanagan, J. R., Frith, C. D., & Wolpert, D. M. (2007). Kinematic cues in perceptual weight judgement and their origins in box lifting. *Psychological research*, *71*(1), 13-21.

Harel, A., Kravitz, D. J., & Baker, C. I. (2012). Deconstructing visual scenes in cortex: gradients of object and spatial layout information. *Cerebral Cortex*, *23*(4), 947-957.

Hayes, S. M., Nadel, L., & Ryan, L. (2007). The effect of scene context on episodic object recognition: parahippocampal cortex mediates memory encoding and retrieval success. *Hippocampus*, *17*(9), 873-889.

Hickok, G., & Hauser, M. (2010). (Mis) understanding mirror neurons. *Current Biology*, *20*(14), R593-R594.

Hrkać, M., Wurm, M. F., Kühn, A. B., & Schubotz, R. I. (2015). Objects mediate goal integration in ventrolateral prefrontal cortex during action observation. *PloS one*, *10*(7), e0134316.

Hrkać, M., Wurm, M. F., & Schubotz, R. I. (2014). Action observers implicitly expect actors to act goal-coherently, even if they do not: An fMRI study. *Human brain mapping*, *35*(5), 2178-2190.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS biology*, *3*(3), e79.

Ishida, H., Nakajima, K., Inase, M., & Murata, A. (2010). Shared mapping of own and others' bodies in visuotactile bimodal area of monkey parietal cortex. *Journal of Cognitive Neuroscience*, *22*(1), 83-96.

Jacob, P. (2008). What do mirror neurons contribute to human social cognition?. *Mind & Language*, *23*(2), 190-223.

Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain sciences*, *17*(2), 187-202.

Jeannerod, M., Arbib, M. A., Rizzolatti, G., & Sakata, H. (1995). Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends in neurosciences*, *18*(7), 314-320.

Johnson-Frey, S. H. (2004). The neural bases of complex tool use in humans. *Trends in cognitive sciences*, *8*(2), 71-78.

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision research*, *47*(26), 3286-3297.

Kalénine, S., Shapiro, A. D., Flumini, A., Borghi, A. M., & Buxbaum, L. J. (2014). Visual context modulates potentiation of grasp types during semantic object categorization. *Psychonomic bulletin & review*, *21*(3), 645-651.

Kamps, F. S., Julian, J. B., Kubilius, J., Kanwisher, N., & Dilks, D. D. (2016). The occipital place area represents the local elements of scenes. Neuroimage, 132, 417-424.

Kellenbach, M. L., Brett, M., & Patterson, K. (2003). Actions speak louder than functions: the importance of manipulability and action in tool representation. *Journal of cognitive neuroscience*, *15*(1), 30-46.

Keysers, C., & Perrett, D. I. (2004). Demystifying social cognition: a Hebbian perspective. *Trends in cognitive sciences*, *8*(11), 501-507.

Kilner, J. M. (2011). More than one pathway to action understanding. *Trends in cognitive sciences*, *15*(8), 352-357.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). Predictive coding: an account of the mirror neuron system. *Cognitive processing*, *8*(3), 159-166.

Kilner, J. M., & Frith, C. D. (2008). Action observation: inferring intentions without mirror neurons. *Current Biology*, *18*(1), R32-R33.

Koechlin, E., & Jubault, T. (2006). Broca's area and the hierarchical organization of human behavior. *Neuron*, *50*(6), 963-974.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science*, *302*(5648), 1181-1185.

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in cognitive sciences*, *11*(6), 229-235.

Kokinov, B. (1997, April). A dynamic theory of implicit context. In *Proceedings of the 2nd European conference on cognitive science* (pp. 9-11).

Kröger, S., Rutter, B., Stark, R., Windmann, S., Hermann, C., & Abraham, A. (2012). Using a shoe as a plant pot: neural correlates of passive conceptual expansion. *Brain research*, *1430*, 52-61.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in cognitive sciences*, *12*(2), 72-79.

Leshinskaya, A., & Caramazza, A. (2015). Abstract categories of functions in anterior parietal lobe. *Neuropsychologia*, *76*, 27-40.

Levy, B. J., & Wagner, A. D. (2011). Cognitive control and right ventrolateral prefrontal cortex: reflexive reorienting, motor inhibition, and action updating. *Annals of the New York Academy of Sciences*, *1224*(1), 40-62.

Liakakis, G., Nickel, J., & Seitz, R. J. (2011). Diversity of the inferior frontal gyrus—a meta-analysis of neuroimaging studies. *Behavioural brain research*, *225*(1), 341-347.

Mahon, B. Z., Milleville, S. C., Negri, G. A., Rumiati, R. I., Caramazza, A., & Martin, A. (2007). Action-related properties shape object representations in the ventral stream. *Neuron*, *55*(3), 507-520.

Malcolm, G. L., Groen, I. I., & Baker, C. I. (2016). Making sense of real-world scenes. *Trends in Cognitive Sciences*, *20*(11), 843-856.

Manera, V., Becchio, C., Cavallo, A., Sartori, L., & Castiello, U. (2011). Cooperation or competition? Discriminating between social intentions by observing prehensile movements. *Experimental Brain Research*, *211*(3-4), 547-556.

Maranesi, M., Bonini, L., & Fogassi, L. (2014). Cortical processing of object affordances for self and others' action. *Frontiers in psychology*, *5*, 538.

Milner, A. D., & Goodale (1995). The visual brain in action. Oxford University Press Oxford.

Moss, H. E., Abdallah, S., Fletcher, P., Bright, P., Pilgrim, L., Acres, K., & Tyler, L. K. (2005). Selecting among competing alternatives: selection and retrieval in the left inferior frontal gyrus. *Cerebral Cortex*, *15*(11), 1723-1735.

Murphy, G. L., & Wisniewski, E. J. (1989). Categorizing objects in isolation and in scenes: What a superordinate is good for. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(4), 572.

Newman-Norlund, R. D., Noordzij, M. L., Meulenbroek, R. G., & Bekkering, H. (2007). Exploring the brain basis of joint action: co-ordination of actions, goals and intentions. *Social Neuroscience*, *2*(1), 48-65.

Nicholson, T., Roser, M., & Bach, P. (2017). Understanding the goals of everyday instrumental actions is primarily linked to object, not motor-kinematic, information: evidence from fMRI. *PloS one*, *12*(1), e0169700.

Noppeney, U. T. A. (2008). The neural systems of tool and action semantics: a perspective from functional imaging. *Journal of Physiology-Paris*, *102*(1-3), 40-49.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in cognitive sciences*, *11*(12), 520-527.

Oosterhof, N. N., Tipper, S. P., & Downing, P. E. (2013). Crossmodal and action-specific: neuroimaging the human mirror neuron system. *Trends in cognitive sciences*, *17*(7), 311-318.

O'Reilly, R. C. (2010). The what and how of prefrontal cortical organization. *Trends in neurosciences*, *33*(8), 355-361.

Palmer, T. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*, 519-526.

Park, S., Brady, T. F., Greene, M. R., & Oliva, A. (2011). Disentangling scene content from spatial boundary: complementary roles for the parahippocampal place area and lateral occipital complex in representing real-world scenes. *Journal of Neuroscience*, *31*(4), 1333-1340.

Petrides, M. (2005). Lateral prefrontal cortex: architectonic and functional organization. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 781-795.

Pezzulo, G., & Cisek, P. (2016). Navigating the affordance landscape: feedback control as a process model of behavior and cognition. *Trends in cognitive sciences*, *20*(6), 414-424.

Poldrack, R. A., Wagner, A. D., Prull, M. W., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1999). Functional specialization for semantic and phonological processing in the left inferior prefrontal cortex. *Neuroimage*, *10*(1), 15-35.

Porter, R. (1985). The corticomotoneuronal component of the pyramidal tract: corticomotoneuronal connections and functions in primates. *Brain Research Reviews*, *10*(1), 1-26.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and brain sciences*, *1*(4), 515-526

Prinz, W. (2006). What re-enactment earns us. *Cortex*, *42*(4), 515-517.

Ramsey R., Cross, E. S., & Hamilton, A. F. D. C. (2011). Eye Can See What You Want: Posterior Intraparietal Sulcus Encodes the Object of an Actor’s Gaze. *Journal of Cognitive Neuroscience*, *23*(11), 3400-3409.

Rezaee, M., Shojaee, M., Ghasemi, A., Moghaddam, A., & Momeni, M. (2010). Which factors affect hand selection in adults? Combined effects of ocular dominance, task demand and object location. *Journal of Human Kinetics*, *26*, 39-44.

Riach, M., Holmes, P. S., Franklin, Z. C., & Wright, D. J. (2018). Observation of an action with a congruent contextual background facilitates corticospinal excitability: A combined TMS and eye-tracking experiment. *Neuropsychologia*, *119*, 157-164.

Rizzolatti, G., Cattaneo, L., Fabbri-Destro, M., & Rozzi, S. (2014). Cortical mechanisms underlying the organization of goal-directed actions and mirror neuron-based action understanding. *Physiological reviews*, *94*(2), 655-706.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annu. Rev. Neurosci.*, *27*, 169-192.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1997). The space around us. *Science*, *277*(5323), 190-191.

Rizzolatti, G., Fogassi, L., & Gallese, V. (2001). Neurophysiological mechanisms underlying the understanding and imitation of action. *Nature reviews neuroscience*, *2*(9), 661.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature reviews neuroscience*, *11*(4), 264-274.

Rumiati, R. I., Weiss, P. H., Shallice, T., Ottoboni, G., Noth, J., Zilles, K., & Fink, G. R. (2004). Neural basis of pantomiming the use of visually presented objects. *Neuroimage*, *21*(4), 1224-1231.

Saxe, R. (2005). Against simulation: the argument from error. *Trends in cognitive sciences*, *9*(4), 174-179.

Schiffer, A. M., Ahlheim, C., Ulrichs, K., & Schubotz, R. I. (2013). Neural changes when actions change: adaptation of strong and weak expectations. *Human brain mapping*, *34*(7), 1713-1727.

Schmahmann, J. D., & Pandya, D. N. (2006). Fiber Pathways of the Brain Oxford University Press. *New York*.

Schubotz, R. I. (2007). Prediction of external events with our motor system: towards a new framework. *Trends in cognitive sciences*, *11*(5), 211-218

Schubotz, R. I., & Fiebach, C. J. (2006). Integrative models of Broca's area and the ventral premotor cortex. *Cortex*, *42*(4), 461-463.

Schubotz, R. I., Sakreida, K., Tittgemeyer, M., & von Cramon, D. Y. (2004). Motor areas beyond motor performance: deficits in serial prediction following ventrolateral premotor lesions. *Neuropsychology*, *18*(4), 638-645.

Schubotz, R. I., & von Cramon, D. Y. (2009). The case of pretense: Observing actions and inferring goals. *Journal of Cognitive Neuroscience*, *21*(4), 642-653.

Schubotz, R. I., & von Cramon, D. Y. (2002). Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: an fMRI study. *Neuroimage*, *15*(4), 787-796.

Schubotz, R. I., & von Cramon, D. (2003). Functional-anatomical concepts of human premotor cortex: evidence from fMRI and PET studies. *Neuroimage*, *20*, S120.

Schubotz, R. I., & Von Cramon, D. Y. (2004). Sequences of abstract nonbiological stimuli share ventral premotor cortex with action observation and imagery. *Journal of Neuroscience*, *24*(24), 5467-5474.

Schubotz, R. I., Wurm, M. F., Wittmann, M. K., & von Cramon, D. Y. (2014). Objects tell us what action we can expect: dissociating brain areas for retrieval and exploitation of action knowledge during action observation in fMRI. *Frontiers in psychology*, *5*, 636.

Seitz, R. J., Schäfer, R., Scherfeld, D., Friederichs, S., Popp, K., Wittsack, H. J., ... & Franz, M. (2008). Valuating other people’s emotional face expression: a combined functional magnetic resonance imaging and electroencephalography study. *Neuroscience*, *152*(3), 713-722.

Shan, B., Zhang, W., Ma, L., Li, D., Cao, B., Tang, Y., ... & Tang, X. (2003). Chinese semantic processing cerebral areas. *Chinese Science Bulletin*, *48*(23), 2607-2610.

Silson, E. H., Chan, A. W. Y., Reynolds, R. C., Kravitz, D. J., & Baker, C. I. (2015). A retinotopic basis for the division of high-level scene processing between lateral and ventral human occipitotemporal cortex. *Journal of Neuroscience*, *35*(34), 11921-11935.

Silson, E. H., Groen, I. I., Kravitz, D. J., & Baker, C. I. (2016). Evaluating the correspondence between face-, scene-, and object-selectivity and retinotopic organization within lateral occipitotemporal cortex. *Journal of vision*, *16*(6), 1-21.

Silson, E. H., Steel, A. D., & Baker, C. I. (2016). Scene-selectivity and retinotopy in medial parietal cortex. *Frontiers in human neuroscience*, *10*, 412.

Simons, J. S., Koutstaal, W., Prince, S., Wagner, A. D., & Schacter, D. L. (2003). Neural mechanisms of visual object priming: evidence for perceptual and semantic distinctions in fusiform cortex. *Neuroimage*, *19*(3), 613-626.

Singh-Curry, V., & Husain M. (2009). The functional role of the inferior parietal lobe in the dorsal and ventral stream dichotomy. *Neuropsychologia*, *47*(6):1434-1448.

Smirnov, D., Glerean, E., Lahnakoski, J. M., Salmi, J., Jääskeläinen, I. P., Sams, M., & Nummenmaa, L. (2014). Fronto-parietal network supports context-dependent speech comprehension. *Neuropsychologia, 63*, 293-303.

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of comparative and physiological psychology*, *43*(6), 482-489.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in cognitive sciences*, *13*(9), 403-409.

Toney, A., & Thomas, B. H. (2006, January). Considering reach in tangible and table top design. In *First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP'06)* (pp. 2-pp). IEEE.

Trapp, S., & Bar, M. (2015). Prediction, context, and competition in visual recognition. *Annals of the New York Academy of Sciences*, *1339*(1), 190-198.

Tyler, L. K., Stamatakis, E. A., Bright, P., Acres, K., Abdallah, S., Rodd, J. M., & Moss, H. E. (2004). Processing objects at different levels of specificity. *Journal of cognitive neuroscience*, *16*(3), 351-362.

Uddén, J., & Bahlmann, J. (2012). A rostro-caudal gradient of structured sequence processing in the left inferior frontal gyrus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1598), 2023-2032.

Uithol, S., & Paulus, M. (2014). What do infants understand of others' action? A theoretical account of early social cognition. *Psychological Research*, *78*(5), 609-622.

Uithol, S., van Rooij, I., Bekkering, H., & Haselager, P. (2011). Understanding motor resonance. *Social neuroscience*, *6*(4), 388-397.

Ungerleider, L.G., Mishkin, M., 1982. Two cortical visual systems. Analysis of Visual Behavior D.J. Ingle, M.A. Goodale & R.J.W. Mansfield, Eds.: MIT Press. *Cambridge, MA*.

Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: a meta-analysis. *Neuroimage*, *48*(3), 564-584.

Van Schie, H. T., Toni, I., & Bekkering, H. (2006). Comparable mechanisms for action and language: neural systems behind intentions, goals, and means. *Cortex*, *42*(4), 495-498.

Vass, L. K., & Epstein, R. A. (2017). Common neural representations for visually guided reorientation and spatial imagery. *Cerebral cortex*, *27*(2), 1457-1471.

Vermeulen, P. (2015). Context blindness in autism spectrum disorder: Not using the forest to see the trees as trees. *Focus on autism and other developmental disabilities*, *30*(3), 182-192.

Virtue, S., Haberman, J., Clancy, Z., Parrish, T., & Beeman, M. J. (2006). Neural activity of inferences during story comprehension. *Brain research*, *1084*(1), 104-114.

Walter, H., Bretschneider, V., Grön, G., Zurowski, B., Wunderlich, A. P., Tomczak, R., & Spitzer, M. (2003). Evidence for quantitative domain dominance for verbal and spatial working memory in frontal and parietal cortex. *Cortex*, *39*(4-5), 897-911.

Waszak, F., Cardoso-Leite, P., & Hughes, G. (2012). Action effect anticipation: neurophysiological basis and functional consequences. *Neuroscience & Biobehavioral Reviews*, *36*(2), 943-959.

Wohlschläger, A., & Bekkering, H. (2002). Is human imitation based on a mirror-neurone system? Some behavioural evidence. *Experimental brain research*, *143*(3), 335-341.

Wokke, M. E., Knot, S. L., Fouad, A., & Ridderinkhof, K. R. (2016). Conflict in the kitchen: Contextual modulation of responsiveness to affordances. *Consciousness and cognition*, *40*, 141-146.

Wolpert, D. M. (1997). Computational approaches to motor control. *Trends in cognitive sciences*, *1*(6), 209-216.

Wolpert, D. M., Doya, K., & Kawato, M. (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *358*(1431), 593-602.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*(5232), 1880-1882.

Wurm, M. F., Ariani, G., Greenlee, M. W., & Lingnau, A. (2016). Decoding concrete and abstract action representations during explicit and implicit conceptual processing. *Cerebral Cortex*, *26*(8), 3390-3401.

Wurm, M. F., Artemenko, C., Giuliani, D., & Schubotz, R. I. (2017). Action at its place: Contextual settings enhance action recognition in 4-to 8-year-old children. *Developmental psychology*, *53*(4), 662.

Wurm, M. F., Cramon, D. Y., & Schubotz, R. I. (2012). The context–object–manipulation triad: Cross talk during action perception revealed by fMRI. *Journal of Cognitive Neuroscience*, *24*(7), 1548-1559.

Wurm, M. F., Hrkać, M., Morikawa, Y., & Schubotz, R. I. (2014). Predicting goals in action episodes attenuates BOLD response in inferior frontal and occipitotemporal cortex. *Behavioural brain research*, *274*, 108-117.

Wurm, M. F., & Lingnau, A. (2015). Decoding actions at different levels of abstraction. *Journal of Neuroscience*, *35*(20), 7727-7735.

Wurm, M. F., & Schubotz, R. I. (2012). Squeezing lemons in the bathroom: contextual information modulates action recognition. *Neuroimage*, *59*(2), 1551-1559.

Wurm, M. F., & Schubotz, R. I. (2017). What's she doing in the kitchen? Context helps when actions are hard to recognize. *Psychonomic bulletin & review*, *24*(2), 503-509.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological bulletin*, *127*(1), 3.

Zentgraf, K., Munzert, J., Bischoff, M., & Newman-Norlund, R. D. (2011). Simulation during observation of human actions–theories, empirical studies, applications. *Vision Research*, *51*(8), 827-835

Zhu, J., & Buckareff, A. A. (2006). Intentions are mental states. *Philosophical Explorations*, *9*(2), 235-242.

Zilles, K., & Amunts, K. (2010). Centenary of Brodmann's map—conception and fate. *Nature Reviews Neuroscience*, *11*(2), 139.

Zimmermann, E., Schnier, F., & Lappe, M. (2010). The contribution of scene context on change detection performance. *Vision research*, *50*(20), 2062-2068.

Abbreviations

1pp	first person perspective
3pp	third person perspective
ACC	anterior cingulate cortex
AG	angular gyrus
AON	action observation network
(a) IPS	(anterior) intraparietal sulcus
ANOVA	analysis of variance
BA	Brodmann area
BOLD	Blood oxygenation level-dependent CM context-manipulation
CO	contextual object
CL	close-left
CR	close-right
dlPFC	dorsolateral prefrontal cortex
EBA	extrastriate body area
EEG	electroencephalography
EPI	echo planar imaging
ER	error rate
FL	far-left
fMRI	function magnetic resonance imaging
FR	far-right
FWHM	full width half maximum
GA	goal affinity
GLM	general linear model
Hz	Hertz
IFG	inferior frontal gyrus
IPL	inferior parietal lobule
ITG	inferior temporal gyrus
L	left
LE	location ergonomics
lPFC	lateral prefrontal cortex
LOC	lateral occipital complex
M1	primary motor cortex
MCC	middle cingulate cortex
MFG	middle frontal gyrus
mPFC	medial prefrontal cortex
ms	millisecond
OPA	occipital place area
OTC	occipitotemporal cortex
PCT	predictive coding theory
PCC	posterior cingulate cortex
plPFC	posterior lateral prefrontal cortex
PMd	dorsal premotor cortex
pMTG	posterior middle temporal gyrus
PMv	ventral premotor cortex
PPA	parahippocampal place area
PPC	posterior parietal cortex
PPS	peripersonal space
pSTS	posterior superior temporal gyrus

R	right
ROI	region of interest
RSC	retrosplenial cortex
RT	reaction time
s	seconds
(pre) SMA	(pre) supplementary motor area
SMG	supramarginal gyrus
SPL	superior parietal lobule
SVC	small volume correction
TE	time echo
TFCE	threshold-free cluster enhancement
TMS	transcranial magnetic stimulation
TR	time of repetition
V1	primary visual cortex
V2-V6	extrastriate visual cortex V2-V6
vlPFC	ventrolateral prefrontal cortex

Curriculum Vitae

Name: Nadiya El-Sourani

Date of birth: 10/08/1988

Place of birth: Leer (Ostfriesland), Germany

Education

09/2013 – present	Ph.D. student (Dr. rer. nat.), University of Münster
09/2011 – 08/2013	M.Sc. studies in human factors and media-psychology, University of Twente, Enschede
09/2007 – 08/2010	B.Sc. studies in psychology, University of Twente, Enschede
07/2007	*Abitur*, Leer (Ostfriesland)

Research

09/2013 – present	Department of Psychology, University of Münster; Biological Psychology, Prof. Dr. Ricarda I. Schubotz

---

Employment

09/2013 – present	Department of Psychology, University of Münster // Research assistant
08/2012- 12/2012	User Experience Center, Deutsche Telekom AG // Internship
04/2011- 08/2011	Recruiting & Talent Service, Deutsche Telekom AG // Internship

Declarations

Declarations by the candidate documenting **open-science activities** and **on the consideration of ethical aspects** as part of the doctoral process and disclosure of **personal contribution** for manuscripts completed by two or more authors (cumulative dissertations)

Doctoral candidate: Nadiya El-Sourani

Title of dissertation: Neural Correlates of Exploiting Contextual Information During Action Observation: Evidence from Contextual Objects

1. Documentation of **open-science activities**

Manuscript 1

	Yes	No	If yes, please specify the source
Pre-registration		X	
Publication of data		X	
Publication of analysis scripts		X	
Publication of materials		X	
Open access publication		X	

Manuscript 2

	Yes	No	If yes, please specify the source
Pre-registration		X	
Publication of data		X	
Publication of analysis scripts		X	
Publication of materials		X	
Open access publication		X	

Manuscript 3

	Yes	No	If yes, please specify the source
Pre-registration		X	
Publication of data		X	
Publication of analysis scripts		X	
Publication of materials		X	
Open access publication		X	

2. Declaration on **the consideration of ethical aspects**

Study number	**Source (manuscript / chapter of dissertation):** e.g. study 1 in paper 2, study 1 described in chapter 4	**Was the study reviewed by an ethics commission?**	
		yes	**No**
1	Manuscript 1 & 2	X	
2	Manuscript 3	X	

3. Declaration of one's **personal contribution** to the submitted academic manuscripts by two or more authors

Manuscript 1

Title	Making sense of objects lying around: How contextual objects shape brain activity during action observation
Author(s)	Nadiya El-Sourani, Moritz F. Wurm, Ima Trempler, Gereon R. Fink, Ricarda I. Schubotz
Publication status:	not yet submitted ☐ (please mark with X) submitted ☐ in review ☐ in revision ☐ accepted ☐ published ☒
Journal	Neuroimage
Year of publication	2018
Description of your own contribution in the case of joint authorship:	- partly responsible for the study's conception and design - responsible for data collection - mainly responsible for processing, analysing and interpreting data - mainly responsible for drafting and revising the manuscript - marked as corresponding author

Manuscript 2

Title	Is it mine or is it yours? Proximal unused objects invite third-person perspective action observers to create their own action space		
Author(s)	Nadiya El-Sourani, Ima Trempler, Moritz F. Wurm, Gereon R. Fink, Ricarda I. Schubotz		
Publication status:	not yet submitted		(please mark with X)
	submitted		
	in review	X	
	in revision		
	accepted		
	published		
Journal	Neuroimage		
Year of publication			
Description of your own contribution in the case of joint authorship: - partly responsible for the study’s conception and design - responsible for data collection - mainly responsible for processing, analysing and interpreting data - mainly responsible for drafting and revising the manuscript - marked as corresponding author			

Manuscript 3

Title	Predictive impact of contextual objects during action observation: evidence from fMRI		
Author(s)	Nadiya El-Sourani, Ima Trempler, Moritz F. Wurm, Gereon R. Fink, Ricarda I. Schubotz		
Publication status:	not yet submitted		(please mark with X)
	submitted	x	
	in review		
	in revision		
	accepted		
	published		
Journal	Journal of Cognitive Neuroscience		
Year of publication			

Description of your own contribution in the case of joint authorship:

- partly responsible for the study's conception and design
- mainly responsible for data collection
- mainly responsible for processing, analysing and interpreting data
- mainly responsible for drafting the manuscript
- marked as corresponding author

Place, Date Signature of the doctoral candidate

Eidesstattliche Versicherungen

Hiermit versichere ich, Nadiya El-Sourani, dass ich

1. nicht wegen eines Verbrechens zu dem ich meine wissenschaftliche Qualifikation missbraucht habe, verurteilt worden bin,
2. keine früheren Promotionsversuche unternommen habe,
3. die Dissertation nicht bereits anderweitig als Prüfungsarbeit vorgelegt habe,
4. die vorgelegte Dissertation selbst und ohne unerlaubte Hilfe angefertigt habe, sowie alle in Anspruch genommenen Quellen und Hilfsmittel in der Dissertation angegeben habe,
5. die Beschreibung der experimentellen Arbeiten, wie in der vorgelegten Dissertation gekennzeichnet, auf drei folgenden wissenschaftlichen Abhandlungen basiere und ich in allen als verantwortlicher Autor ("corresponding author") gekennzeichnet bin.

Münster, den ____________________ ____
Unterschrift