

Contextual modulation of action perception

INAUGURAL – DISSERTATION

zur Erlangung des Doktorgrades
doctor rerum naturalium
Dr. rer. nat.
im Fachbereich Psychologie
der Westfälischen Wilhelms-Universität Münster

Vorgelegt von

Moritz Wurm

geboren am 11. November 1981 in Köln, Deutschland

Dekan: Prof. Markus Lappe

Erste Gutachterin: Prof. Ricarda I. Schubotz

Zweiter Gutachter: Prof. Markus Lappe

Datum der mündlichen Prüfung:

Datum der Promotion:

Acknowledgements

First and foremost, I would like to thank Ricarda I. Schubotz for her outstanding supervision. I am profoundly grateful to be inspired by her knowledge and insights, her enthusiasm, and her rigorous scientific integrity. During the years of my PhD preparation, she constantly supported me and gave me the opportunity to follow new ideas which emerged in that time.

I am very thankful to my second supervisor Prof. Markus Lappe for his interest in my work and his straightforwardness with regard to using the eyetracking facility of his lab.

I also want to thank Prof. D. Yves von Cramon for his valuable feedback regarding experimental data and manuscripts.

Furthermore, I am very grateful to the members of the Motor Cognition Group at the Max Planck Institute for Neurological Research: Katja Kornysheva for her exceptionally delightful company, scientific attitude, congenial sense of humour, and friendship; my office-mates Marike Schiffer and Christiane Ahlheim for fruitful discussions, their help with experimental and statistical questions, and their friendship; Anne Kühn for good company and cooperation with the organization of the "Experimentelles Forschungspraktikum" at the Psychological Institute in Münster; Anna Schönberger and Ellen Binder for their company and helpful and interesting discussions; Anika Sierk, Kim Krause, Klara Hagelweide, Erik Noorman and Silvan Siep for practical assistance; as well as Matthis Drolet for IT advice and company.

I would like to express special thanks to Christina Woitscheck, who wrote her bachelor thesis under my supervision, for her scientific rigour and excellent organization of

our developmental study and my interns Yuka Morikawa, Anna Schneider, Daniela Giuliani, Marcin Lipski, Constantin Erlenbeck, Lisbeth Cardenas-Morales, Sonja Krettek, Ilona Schneider, and Florian Bolenz for their help in experimental assistance.

Finally I want to thank my friends and my family for their encouragement and love.

Contents

0	Abstract	1
	Zusammenfassung	4
1	Introduction	8
1.1	Action structure	10
1.1.1	Action hierarchies	10
1.1.2	Goals	12
1.1.3	Intentions	13
1.2	Neural and functional models of action perception	15
1.2.1	Visual perception of manipulation and object perception	15
1.2.2	Involvement of the motor system in action perception	17
1.3	Mental state inference and ToM during action perception	22
2	Objectives of the thesis	27
3	General discussion	31
3.1	Evidence for involvement of ToM-related brain regions during action perception.....	31
3.1.1	The role of the temporoparietal junction (TPJ) in action perception from a third person perspective	32
3.1.2	Considering actors during action perception	33
3.2	The influence of contextual information on action perception	36

3.2.1	Neutral conditions and the differentiation of facilitation and interference	36
3.2.2	Context-manipulation (CM) vs. context-object (CO) associations	37
3.2.3	The context-object-manipulation (COM) triad	41
3.3	Future prospects	43
4	References	47
5	Appendix	60
5.1	List of Abbreviations	60
5.2	Figures	63
6	Reprints	65

0 Abstract

Perceiving the actions of others is a prerequisite for inference of mental states (e.g. intentions and desires) of the acting agent, prediction of future steps of the ongoing action, and eventual selection of appropriate re- and/or interactions. Action perception is therefore of special importance in a social environment.

During action perception, the cortical motor system, in particular the premotor cortex and corresponding parietal areas, is suggested to operate as an emulator to generate expectations about likely, forthcoming action steps. To this end, information about movements of body parts and involved objects is exploited. This directly relevant or primary action information is, however, in many cases insufficient for long-term goal and mental state inference. In our everyday life, actions are embedded in contextual situations, i.e. they are performed by agents with particular mental states and personality traits, and they take place at particular places and times. This additional or secondary information is likely to be integrated during action perception to enable a full understanding of an action. The present work aimed to investigate whether contextual information modulates action perception.

Exp. 1 was motivated by the simplistic view that action perception exclusively draws on the motor system, whereas regions involved in mental state inference (the so-called Theory-of-Mind (ToM) network) are demanded only when actions are implausible or new to the observer. To test whether the ToM network is also enrolled during perception of ordinary everyday actions, visual features that are characteristic for actions of others as opposed to own actions

were investigated. These features were a third (as opposed to a first) person perspective on the action, and the sight of the whole actor including his or her face. Functional Magnetic Resonance Imaging (fMRI) was used to record neural activity in the ToM-network, i.e. the temporoparietal junction (TPJ), the medial prefrontal cortex (mPFC), and the posterior cingulate cortex (PCC), during observation of everyday actions. Activity in regions of the ToM network was significantly enhanced when actions were observed from third person perspective, when the actors' faces were present, and when the actor switched from one trial to another. These results suggest that ToM regions are also engaged during perception of ordinary action, especially when visual cues imply that an action is performed by someone else.

Exp. 2 aimed to investigate the influence of contextual information provided by the scene an action is embedded in (contexts hereafter). Since most actions take place in particular domestic rooms, compatibility of action and context was speculated to modulate action perception. In a behavioral study (Exp. 2a), reaction times were recorded when subjects recognized actions in compatible, incompatible, or neutral contexts. We then used fMRI to identify regions that are modulated by compatibility of action and context (Exp. 2b). Actions were recognized slower in incompatible as compared to compatible or neutral contexts. In addition, neural activity was enhanced in the left ventrolateral prefrontal cortex (vlPFC) when subjects observed actions in incompatible as opposed to compatible or neutral contexts. These results provide evidence that contextual incompatibility affects a high processing level of action perception. Activation of the left vlPFC suggests that actions are embedded in superordinate action sequences, a process that is required for inference of long-term goals and intentions as well as prediction of forthcoming actions. This embedding of actions into action sequences appears to be particularly demanding when actions take place in incompatible contexts.

Exp. 2c examined the triadic relationship of context, object, and manipulation information. These three kinds of information are tightly coupled, since manipulations and objects usually co-occur in specific contexts. To disentangle the interaction between manipulations and context from interactions between objects and context, manipulation-object associations were decoupled by employing pantomime actions, i.e. actions on inappropriate objects. A 2x2 factorial design was realized where (1) neither manipulations nor objects, (2) only objects, (3) only manipulations, or (4) both manipulations and objects were compatible with the context. Using fMRI, it was found that retrieval of relevant information was hampered when invalid, to-be-suppressed object information and associated movement information were contextually related to the pantomimed action. Importantly, compatibility of context to either manipulation or objects triggered increased attention to either manipulation information (signified by activation of the motor system) or object information (signified by activation of object-processing regions, i.e. the lateral occipital complex (LOC)). These findings provide evidence that contextual information influences the neural processing of manipulation and object information during action perception. Further, findings corroborate that manipulations and objects of the same contextual category are stored in neighbouring clusters in semantic memory.

Together, the experimental data support the view that contextual information, provided by either person or contextual scene information, modulates action perception.

0 Zusammenfassung

Die Wahrnehmung von Handlungen anderer ist eine Grundvoraussetzung für die Erschließung mentaler Zustände (z.B. Intentionen, Wünsche und Überzeugungen einer handelnden Person), für die Vorhersage zukünftiger Handlungsschritte und letztendlich für die Auswahl angemessener Re- und/oder Interaktionen. Handlungswahrnehmung spielt daher eine tragende Rolle für das Leben in einer sozialen Umgebung. Während der Handlungswahrnehmung fungiert das kortikale motorische System (d.h. der prämotorische Cortex und korrespondierende parietale Areale) als ein Emulator, um Erwartungen über zukünftige Handlungsschritte zu generieren. Dabei wird Manipulations- und Objektinformation, d. h. Information über Bewegungen involvierter Körperteile und Objekte, genutzt. In den meisten Fällen ist allerdings die Analyse dieser primären (d. h. direkt relevanten) Information unzureichend, um die Langzeitziele und Beweggründe des Akteurs zu erfahren. Im alltäglichen Leben sind Handlungen in kontextuelle Situationen eingebunden, d. h. sie werden von Personen mit bestimmten Geisteszuständen und Charaktereigenschaften ausgeführt und finden an bestimmten Orten oder bestimmten Zeiten statt. Es ist daher wahrscheinlich, dass diese zusätzlichen, sekundären Informationen in den Prozess der Handlungswahrnehmung eingebunden werden, um ein umfassendes Verständnis einer Handlung zu ermöglichen. Die vorliegende Arbeit untersuchte, ob sekundäre, kontextuelle Information einen modulatorischen Einfluss auf Handlungswahrnehmung hat.

Exp. 1 wurde durch die simplistische Interpretation motiviert, dass Handlungswahrnehmung ausschließlich das motorische System involviert, während Regionen, die bei der Erschließung mentaler Zustände eine Rolle spielen (das sogenannte Theory-of-Mind(ToM)-Netzwerk), nur bei der Wahrnehmung neuer oder ungewohnter Handlungen zusätzlich aktiviert werden. Um zu zeigen, dass das ToM-Netzwerk auch bei der Wahrnehmung gewöhnlicher Alltagshandlungen eingebunden wird, wurden visuelle Handlungseigenschaften, die charakteristisch für die Handlungen anderer (im Gegensatz zu eigenen Handlungen) sind, näher untersucht. Untersuchte Charakteristika betrafen die Dritte (im Vergleich zur Erste)-Person-Perspektive sowie die Präsenz der Gesichter der Darsteller. Mit Hilfe der funktionellen Magnetresonanztomographie (fMRT) wurde die neuronale Aktivität im ToM-Netzwerk (d. h. im temporoparietalen Übergangsareal (TPJ), im medialen präfrontalen Cortex (mPFC) und im posterioren cingulären Cortex (PCC)) während der Betrachtung alltäglicher Handlungen gemessen. Aktivität im ToM-Netzwerk war signifikant stärker, wenn Handlungen aus einer Dritte-Person-Perspektive betrachtet wurden, wenn das Gesicht des Akteurs sichtbar war und wenn der Darsteller zwischen zwei gezeigten Handlungen wechselte. Diese Ergebnisse unterstützen die Annahme, dass das ToM-Netzwerk auch bei der Wahrnehmung von gewöhnlichen Alltagshandlungen involviert ist, insbesondere wenn visuelle Reize implizieren, dass eine Handlung von einer anderen Person ausgeführt wird.

Exp. 2 befasste sich mit dem Einfluss der Hintergrundszene (im Folgenden "Kontext"), in der eine Handlung stattfindet. Da die meisten Handlungen spezifisch für bestimmte Räume sind, wurde angenommen, dass die Kompatibilität von Handlung und Kontext einen modulatorischen Einfluss auf Handlungswahrnehmung hat. In einer Verhaltensstudie (Exp. 2a) wurden Reaktionszeiten gemessen, wenn Versuchspersonen Handlungen in kompatiblen, inkompatiblen

oder neutralen Kontexten erkennen sollten. Anschließend wurden mit Hilfe der fMRT Hirnregionen identifiziert, die durch Kompatibilität von Handlung und Kontext moduliert werden (Exp. 2b). Handlungen wurden langsamer erkannt, wenn sie in inkompatiblen – verglichen mit kompatiblen oder neutralen – Kontexten stattfanden. Außerdem war die neuronale Aktivität im linken ventrolateralen präfrontalen Cortex (vlPFC) höher, wenn Handlungen in inkompatiblen – verglichen mit kompatiblen oder neutralen – Kontexten betrachtet wurden. Die Ergebnisse lassen vermuten, dass kontextuelle Inkompatibilität einen Einfluss auf höhere Prozesse während der Handlungswahrnehmung hat. Die Aktivierung des linken vlPFC legt nahe, dass Handlungen automatisch in eine übergeordnete Handlungssequenz eingebunden werden. Erst dieser Prozess ermöglicht Rückschlüsse auf Langzeitziele und Intentionen einer Handlung sowie - damit verbunden - die Vorhersage zukünftiger Handlungen. Das Einbetten einer Handlung in Handlungssequenzen ist einleuchtenderweise anspruchsvoller, wenn eine Handlung in einer inkompatiblen Szene stattfindet.

In Exp. 2c wurde die triadische Beziehung zwischen Kontext, Objekt und Manipulation untersucht. Um Assoziationen zwischen Kontext und Manipulation sowie zwischen Kontext und Objekten getrennt voneinander analysieren zu können, wurden pantomimische Handlungen, die mit unpassenden Objekten ausgeführt wurden, verwendet. So konnte ein 2x2 faktorielles Design realisiert werden, in dem (1) weder Handlungen noch Objekte, (2) nur die Handlungen, (3) nur Objekte, oder (4) sowohl Handlungen als auch Objekte kompatibel mit der Szene waren. Mit Hilfe der fMRT konnte gezeigt werden, dass der Abruf relevanter Manipulationsinformation behindert wurde, wenn unpassende, zu unterdrückende Objektinformation und zugehörige Manipulationsinformation kontextuell verwandt mit der gesuchten pantomimischen Handlung waren. Insbesondere erhöhte Kontext-Kompatibilität zu entweder Handlungen oder Objekten die

Aufmerksamkeit auf entweder Manipulationsinformation (angedeutet durch verstärkte Aktivität im motorischen System) oder Objektinformation (angedeutet durch verstärkte Aktivierung in Objekt-verarbeitenden Regionen, d. h. im lateralen okzipitalen Komplex (LOC)). Die Ergebnisse weisen darauf hin, dass kontextuelle Information die neuronale Verarbeitung von Manipulation- und Objekt-Information beeinflusst und Objekte und Manipulationen innerhalb kontextueller Kategorien in Beziehung zueinander im semantischen Gedächtnis gespeichert sind.

Die vorgestellten Experimente unterstützen die Hypothese, dass kontextuelle Information, aktiviert durch visuelle Reize von Personen oder kontextuellen Szenen, einen Einfluss auf Handlungswahrnehmung hat.

1 Introduction

As social beings, understanding the actions of others is an important ability, as it enables to appropriately react to and interact with other agents. Beyond merely identifying an immediate action, a prerequisite for appropriate re- and interactions is the ability to capture underlying goals and intentions and to predict future steps of an unfolding action with ample precision.

Visually perceiving a typical action, e.g. the pouring of water from a bottle into a glass, draws on visual analysis of manipulation information, i.e. movement trajectories and posture of hands, and object information, i.e. pragmatic information about size and shape, and semantic information about associated affordances and outcomes. Both manipulation and object information are directly relevant for action recognition and are defined here as *primary action information*. On the basis of primary action information, an observer can infer possible goals and intentions, and can generate predictions about most likely future steps of the action. For example, the goal of the action may be to prepare a drink because the agent had the intention to quench his or her thirst. Hence, after filling the glass, the bottle might be placed on the table and the glass might be grasped and brought to the mouth of the agent. It may be, however, that the agent is wearing a waiter's dress and the action takes place in a restaurant. In this case, one would infer a different underlying intention, e.g. that the agent aims to serve the drink to a guest. Consequently, alternative predictions about the ongoing action become more likely, e.g. after filling the glass, it is passed to someone else.

The example demonstrates that information that is actually not relevant for the recognition of an ongoing action *step* is still often crucial for higher-level processes like inference of goals and intentions and prediction of forthcoming action steps or actions. In everyday life, we are exposed to a countless number of cues that may influence action perception beyond single action steps. Such *secondary* information that potentially modulates action perception will be referred to as *contextual information* hereafter.

Contextual information may be broadly subdivided into two classes. First, person-related contextual information is provided by the agent performing an action. This may be transient states like emotional expressions and appearance indicating mental or social states, and more enduring states like personality traits and idiosyncrasies. Second, contextual information is derived from the situation or scene in which an action takes place. This may also comprise temporal aspects like the time of day or season, and prototypical social or cultural events (e.g. birthday).

While the influence of contextual information on object information has extensively been investigated (Bar, 2004; Bar and Ullman, 1996; Biederman et al., 1973; Boyce et al., 1989; Chun and Jiang, 1998; Davenport and Potter, 2004; De Graef et al., 1990; Friedman, 1979; Ganis and Kutas, 2003; Gronau et al., 2008; Hayes et al., 2007; Hollingworth and Henderson, 1998; Jenkins et al., 2010; Joubert et al., 2008; Kirk, 2008; Oliva and Torralba, 2007; Palmer, 1975; Zimmermann et al., 2010), most studies on the contextual influence on action perception relate to primary action information (Jastorff et al., 2010; Liepelt et al., 2009; Liepelt et al., 2008; Marsh et al., 2011; but see Brass et al., 2007; Iacoboni et al., 2005; and Wyk et al., 2009 for exceptions). The present work therefore focused on the influence of secondary contextual information on action perception.

The following sections describe the structure of actions, neural and functional models of action perception, and possible points at which contextual information might affect action perception.

1.1 Action structure

Actions can be defined as (observable) change induced by intentional beings (Zacks and Tversky, 2001). As such, actions are a subclass of events, which may be defined as observable change irrespective whether induced by intentional beings or physical causes (Schubotz, 2007; Zacks and Tversky, 2001). Like events, actions can be temporally structured on multiple scales, depending on the definition of beginning and end of an action. For example, "making a sandwich" can be considered as a single action; however, on a more fine grained scale, it can be considered as a sequence of action steps, e.g. buttering bread, slicing cheese, and so on. On a larger, coarse-grained scale, it may be considered part of an overarching action, e.g. "preparing a meal" (Kurby and Zacks, 2008). This nested organization is suggested to be also reflected in the neural representation of action.

1.1.1 Action hierarchies

The hierarchical organization of action has been described in several models (Byrne and Russon, 1998; Csibra, 2007; Grafton and Hamilton, 2007; Hamilton and Grafton, 2006, 2007; Jeannerod, 1994; Kilner et al., 2007; Wolpert, 1997).

Kilner et al. differentiate four distinct levels (Kilner, 2011; Kilner et al., 2007): Beginning at the bottom of the hierarchy, the muscle level describes the pattern of muscle activity required to execute an action. Second, the kinematic level describes the movement kinematics of an effector in space and time. Third, the goal level refers to the desired outcome of an action. Finally, the intention level describes the motivation that drives the execution of a particular action.

With respect to the neuronal representations of the distinct levels, there are, however, some ambiguities. To make sense of the hierarchy, an action should be represented by a clearly defined set of neurons at each level. With respect to the muscle level this seems to be the case, since movements are eventually induced by the orchestrated firing pattern of lower motor neurons in the brainstem and the spinal chord at the end-plates of muscles (Porter, 1985). Likewise, the movement kinematics of an effector are determined by upper motor neurons in the motor and premotor cortex (Dum and Strick, 2002; Graziano, 2006). Kilner proposes that only a "visual representation of the kinematic level" is accessible to an observer's perception (Kilner et al., 2007). It is, however, questionable whether a visual representation of action kinematics is encoded by the same (sensori)*motor*¹ neurons required for execution of that action (see section 1.2.2). Although it is clear that at some point there is a matching of visual and motor responses (visuomotor neurons (Graziano et al., 2002)), it seems that the kinematic level, as defined by

¹ Motor neurons are commonly defined as neurons that respond to execution of a particular movement (usually detected via single cell recording, cf. e.g. Rizzolatti et al., 1987). However, this does not imply that such neurons are necessarily required for the execution of an action (e.g. postsynaptic neurons unidirectionally connected to neurons that trigger upper motor neurons). Another criterion for the definition of motor neurons could be the induction of movements during activation of a putative motor neuron (detected via intracortical microstimulation, e.g. Graziano et al., 2002). However, the induction of movements depends on duration and intensity of stimulation (cf. Schubotz, 2004). Thus, a motor neuron may be triggered (thereby inducing a movement) via transsynaptic spreading of activation during experimental stimulation of a presynaptic neuron that would not induce a movement under natural conditions. Hence, in both definitions there may be a *correlative*, but not necessarily a *causal* relationship of neural activity and induction of an overt movement.

Kilner, does not describe a single, but rather two representations of movement kinematics, i.e. an executional and an observational one.

Notably, there is a gradient of abstraction from the intention level (abstract) to the muscle level (concrete), as highlighted in the model of Wolpert (1997). This gradient is reflected by a reduction of degrees of freedom from the abstract intention level to the concrete muscle level. In other words, an intention can be realized by different actions, and an action can be realized by an infinite number of possible movement kinematics. During action execution, the most efficient means to achieve a particular outcome is selected on each level within the action hierarchy (Wolpert, 1997).

The hierarchical model of Hamilton & Grafton resembles the hierarchy proposed by Kilner (Hamilton and Grafton, 2007). In their model, however, no distinction between intentions and goals is made. Since psychologists and cognitive and computational neuroscientists often use the terms *goal* and *intention* in an ambiguous manner, the goal and the intention level deserve closer consideration in the following two sections.

1.1.2 Goals

The goal level can be sharply differentiated from the intention level under the premise that intention is defined as a mental state (see section 1.1.3). Nevertheless, it comprises a broad range of temporal scales, depending on what is considered as the outcome of an action (or how the beginning and the end of an action are defined, see first paragraph of section 1.1). For example, putting sugar in a cup of coffee has the outcome of sweetening the coffee. On higher levels, the outcome could also be described as "making coffee" or "preparing breakfast". In many

instances, it may hence be useful to differentiate overarching long-term goals from immediate goals or subgoals (e.g. in Hamilton and Grafton, 2006). The delineation between long-term goals and subgoals, however, remains a matter of opinion. The term *goal* is hence potentially problematic. This is reflected by its different usages in the literature. The term *goal* is used in (at least) three qualitatively different ways (Uithol et al., 2011): In its most minimalistic form, a goal is simply a target in space, e.g. an object to be grasped (Hamilton and Grafton, 2006), the end location of a reaching action (Bekkering et al., 2000), or a cross on a desk to be pointed at (Wohlschlagler and Bekkering, 2002). The according actions are often termed goal-directed or transitive actions as opposed to intransitive actions (Uithol et al., 2011). Second, many authors use the term *goal* synonymously with the term *action*. This usage is particularly frequent in early studies on mirror neurons (see below) in the macaque monkey (Rizzolatti and Craighero, 2004). The third usage of *goal* is in terms of a desired end-state or outcome of an action, e.g. a full cup of coffee (Csibra and Gergely, 2007). Given this wide range of definitions, it is not surprising that (functional imaging) studies reveal heterogeneous results for (the neural correlates of) 'goals' (Hamilton and Grafton, 2006, 2008; Hesse et al., 2009; Majdandzic et al., 2009; Majdandzic et al., 2007; Schubotz and von Cramon, 2009). In the following, the term is used in the latter way, i.e. a goal refers to an end-state of an action in the external world.

1.1.3 Intentions

Intentions are mental states that usually determine and therefore precede the execution of an action (Jeannerod, 2006; but see Desmuerget et al., 2009). Philosophers often differentiate between 'prior intentions' and 'intentions in action' (Pacherie, 2000; Searle, 1983). The intention

in action (or 'motor intention' (Jeannerod, 2006)) refers to the desired kinematic outcome (e.g. intending to raise the arm) whereas the prior intention refers to the desired goal (e.g. intending to wave goodbye). Thus, prior intentions, unlike motor intentions, have a conceptual content and precede motor intentions (Jacob and Jeannerod, 2005; Jeannerod, 2006).

In the action literature, the term *goal* or *long-term goal* is often used interchangeably with *intention* (Hamilton, 2009; Hamilton and Grafton, 2007; Iacoboni et al., 2005). However, given the definition in section 1.1.2, goals refer to forthcoming states in the external world. Conversely, intentions refer to internal states of the agent. Hence, intentions are not directly accessible to the perception of an observer (Premack and Woodruff, 1987), whereas goals should be perceivable at least at future states. To give an example, Iacoboni defines intention as "the 'why' of an action" (Iacoboni et al., 2005). The action "grasping" is suggested to have two possible intentions, "(grasping) to drink" and "(grasping) to clean". However, "drinking" and "cleaning" are themselves actions or, more precisely, overarching goals of the subgoal "grasping"), that either come true in the future or not. Instead, the overarching goal "drinking" may have the two possible intentions "(drinking) to quench thirst" and "(drinking) to engage in social activity" (Uithol et al., 2011). When asking subjects for "the 'why' of an action" performed by someone else, subjects may either try to infer goals or contemplate on the actor's mental states. However, mental state inference (Frith and Frith, 2005) likely differs from action goal inference (Frith and Frith, 2006; Saxe, 2005; Saxe et al., 2004; but see Gallese & Goldman, 1998). Thus, frivolous usage of the term *intention* entails the risk to confuse cognitive processes underlying mental state and goal inference (Cattaneo et al., 2007; Fogassi et al., 2005; Iacoboni et al., 2005; Ortigue et al., 2010; Ortigue et al., 2009).

1.2 Neural and functional models of action perception

Decoding the actions of others involves a range of cognitive processes, including the analysis of primary action information (e.g. movements of body parts and objects, knowledge about object use and functions) and integration of these different sources of information into a common reference frame. Contextual information, e.g. person-related and scene-related information, might modulate action perception at several levels of the action hierarchy. To better understand at which level contextual information might be integrated, neural and functional models of the perception of primary action information will be described in the following sections.

1.2.1 Visual perception of manipulation and object information

The initial requirement of visual action perception is the visual analysis of object and manipulation information. Visual information is propagated from the primary visual cortex V1 along two processing routes (Milner and Goodale, 1995; Ungerleider and Mishkin, 1982). The ventral, "semantic" route, also termed the "what" stream (Ungerleider and Mishkin, 1982), stretches from V1 via V2 and V4 to areas of the inferior temporal lobe and is generally concerned with visual representation and knowledge (Buxbaum and Kalenine, 2010; Jeannerod et al., 1995). The dorsal, "pragmatic" route, also termed "where" (Ungerleider and Mishkin, 1982) and "how" stream (Milner and Goodale, 1995), stretches from V1 to the parietal lobe. The exact function of the dorsal stream is less clear: Originally the dorsal stream was proposed to

mediate spatial perception (hence "where" stream) and visually guided action (hence "how" stream). This heterogeneity inveigled Rizzolatti and Matelli (2003) to further subdivide the dorsal stream into a dorso-dorsal (via V6 to the superior parietal lobe (SPL), processing visually guided action) and a ventro-dorsal stream (via V5 to the inferior temporal lobe (IPL), involved in space and action perception). However, this subdivision is speculative (Husain and Nachev, 2007) and is contradicted by a number of findings, e.g. both action execution and observation activate regions in both IPL and SPL to a similar degree (Grezes and Decety, 2001). It is more likely that the posterior parietal cortex is organized in a somatotopic manner (Buccino et al., 2001), paralleling the somatotopic organization of adjacent SI, M1 (Penfield and Rasmussen, 1950), and the premotor cortex (Schubotz and von Cramon, 2003). In line, SPL is connected to the dorsolateral prefrontal and dorsal premotor cortex, whereas the IPL is connected to ventrolateral prefrontal and ventral premotor cortex (Geyer et al., 2000; Schmahmann and Pandya, 2006). Against the traditional view (Ungerleider and Mishkin, 1982), ventral and dorsal streams seem to be strongly interconnected (Farivar, 2009).

Brain regions selective for visual processing of object information were found in the lateral occipitotemporal (OTC, part of the ventro-dorsal stream) and fusiform cortex (part of the ventral stream), together forming the lateral occipital complex (Grill-Spector et al., 2001; Grill-Spector et al., 1999). These regions are selectively activated when seeing objects vs. scrambled objects (Grill-Spector et al., 2001). A region in the posterior middle temporal gyrus (pMTG) has been found to be specific for the visual processing of tools and tool-specific movements (Beauchamp et al., 2002, 2003), hence termed MTG-tool area (MTG-TA; Downing et al., 2006).

Regions sensitive to the visual processing of body parts are the posterior superior temporal sulcus (pSTS) and the extrastriate and fusiform body areas (EBA and FBA,

respectively). These regions are activated during observation of body parts vs. non-body objects (Downing et al., 2001; Peelen and Downing, 2007). The pSTS has been shown to be particularly selective for the processing of biological motion (Allison et al., 2000; Peuskens et al., 2005). So-called snapshot neurons (coding e.g. the shape of a body posture) and optic flow pattern neurons (coding a particular movement pattern) in the pSTS are proposed to combine action-relevant form and motion (Oram and Perrett, 1996), enabling high-level biological motion processing (Giese and Poggio, 2003).

Several regions in the OTC show sensitivity to specific, visually perceived actions, as revealed in repetition suppression (or fMRI adaptation) experiments (Grill-Spector and Malach, 2001). For example, observation of the same vs. different actions or body movements in repeated trials attenuates the BOLD response in the OTC (Grafton and Hamilton, 2007; Jastorff et al., 2009; Jastorff and Orban, 2009; Noppeney et al., 2005), even when the repeated action is performed by another actor (Kable and Chatterjee, 2006; Wiggett and Downing, 2011). Together, these findings suggest that the specific kinematics of an action are encoded in the OTC. However, all of these studies also find activation in prefrontal regions (premotor cortex and/or ventrolateral prefrontal cortex (vlPFC)). Thus, it cannot be ruled out that attenuation of the OTC is the result of a top-down signal from frontal regions to perceptual regions (see section 1.2.2).

1.2.2 Involvement of the motor system in action perception

There is plenty of evidence that the cortical motor system (here referred to as premotor cortex and associated posterior parietal projection sites) is activated during action observation

(Caspers et al., 2010; Jeannerod, 2006). For example, the cortical excitability of arm and finger muscles is selectively facilitated during observation of actions involving either arm or finger movements (Strafella and Paus, 2000). According to the ideomotor principle (for a review see Shin et al., 2010; Stock and Stock, 2004), actions are represented by their sensory consequences (Carpenter, 1852; James, 1890). Hence, perceiving the consequences of an action excites the motor program required to execute that same action (Jeannerod, 1994; Prinz, 1997). It is not entirely clear to what extent activation of the motor system is necessary for action perception and, more specifically, action goal inference, or whether it simply reflects sensorimotor associations (Hickok, 2009; Saxe, 2005). Three major interpretations are currently discussed with respect to involvement of the motor system in action perception: *direct matching*, *action reconstruction*, and *predictive coding* (Hamilton, 2009; Zentgraf et al., 2011).

According to the *direct matching* account, also known as motor simulation theory, action goals are inferred by matching perceived action properties with the sensory consequences of self-performed actions (Rizzolatti and Craighero, 2004). Thus, "the motor system 'resonates' [with the observed action]" (Rizzolatti et al., 2001), thereby activating associated action goals and intentions (Gallese and Goldman, 1998). Activation of the motor system is hence *retrodictive*, i.e. visually perceived action kinematics are propagated bottom-up from the OTC to the motor system, thereby enabling access to knowledge about action goals. Support for direct matching came from the discovery of so-called mirror neurons in the macaque premotor area F5 (di Pellegrino et al., 1992; Gallese et al., 1996) and later in the parietal areas PFG and AIP (Fogassi et al., 2005; for a review see Rizzolatti and Sinigaglia, 2010). These neurons are active when the monkey executes an action as well as when it perceives the same or a related action performed

by someone else. In other words, these neurons are proposed to 'mirror' the observed action in the monkey's motor system and thereby enable action 'understanding'.

There are, however, a range of theoretical concerns with respect to goal inference via motor resonance/direct matching (Csibra, 2007; Hickok, 2009; Hickok and Hauser, 2010; Jacob, 2008; Uithol et al., 2011). Here, only two crucial objections are further discussed. First, it is not clear how action goals, that are not in the observer's own action repertoire, are inferred. For instance, a monkey is well able to recognize the goal of a human throwing action the monkey is physically unable to perform (Wood and Hauser, 2008). Similarly, innately paralyzed humans effortlessly infer the goals of others without having associated motor programs. Second, it is unclear at which level a matching of perceived properties with own motor programs can enable the inference of a single goal. Crucially, a particular action (kinematic) can have different goals (pouring water from a bottle to prepare a drink, to empty the bottle, to water flowers, or to tease a friend), likewise a particular goal can be achieved by many means (filling a glass with water by pouring a bottle, holding the glass under a tap, dipping the glass in a lake, etc). According to these opposing views, mirror neurons either reflect high-level, conceptual action information without necessarily being related to the own action repertoire (Jacob, 2009) or emulative processes to generate expectations about future stages of the action (see below).

Action reconstruction, also referred to as the *inferential model*, suggests that involvement of the motor system during action perception does not subserve action goal inference (Csibra, 2007). According to this view, action goals originate outside the motor system (e.g. in the vLPFC (Badre, 2008; Grafman, 2002; Petrides, 2005)), whereas the motor system is involved to compute the most appropriate means to attain a particular goal. Hence, goal inference is proposed to occur outside the motor system, e.g. by propagating relevant action information

from the OTC directly to prefrontal regions. Inferred goals are then used to 'reconstruct' the lower level representations of an action, e.g. its kinematics. Hence, goal inference is proposed to *precede* motor system involvement (Csibra, 2007).

Predictive coding (Kilner et al., 2007) proposes that the motor system is used to generate internal models of an observed action and to thereby predict an action's perceptual consequences (Friston et al., 2011; Kilner et al., 2007). Predictive coding is adapted from motor control theory, which is primarily concerned with the regulation of own movements (Sperry, 1950; Wolpert et al., 1995). According to the motor control theory, a goal is attained when the anticipated, desired end-state of the action overlaps with the actual, sensory results of the action. Therefore, an efference copy of the desired end-state is generated and compared with reafferent signals resulting from the action (i.e. extero- and interoceptive feedback). A discrepancy between predicted and actual sensory feedback (i.e. the prediction error) is used to adjust the movements and to finally attain the goal. Importantly, the prediction can be used to differentiate between self-produced and externally induced change (Jeannerod, 2006) and to attenuate sensory consequences of self-produced movements (Blakemore et al., 1998; Keysers and Perrett, 2004). For example, neurons in the upper bank of the monkey STS, which are selective for a large variety of visually perceived body movements, do not respond to the sight of the monkey's own movements (Hietanen and Perrett, 1993; Keysers and Perrett, 2004). More sophisticated models take into account that reafferent feedback from sensory organs is corrupted by noise and often too slow for efficient adjustments of ongoing movements. To this end, so-called forward models are implemented. Forward models are emulators of an ongoing action and are hence capable to estimate states of the action before actual feedback is available. When feedback finally arrives, it is combined with the estimates of the forward model. Actual and predicted feedback may differ

due to possible drifts of the motor commands, resulting in a predicting error. Forward models hence capture the causal relationship between motor commands and their sensory consequences (Wolpert and Ghahramani, 2000). Importantly, forward models might be run 'offline' and thereby contribute to a number of cognitive tasks, e.g. motor imagery, action planning, and anticipating the outcome of actions of others (Grush, 2004; Miall, 2003; Wolpert et al., 2003). Furthermore, forward models are capable of emulating the dynamics of non-animate objects and events (Davidson and Wolpert, 2005; Schubotz, 2007). Hence, internal emulation (as opposed to simulation, which by definition (Grush, 2004) should draw on the very same (motor) neurons required for execution of a respective action) is capable to account for the ability to estimate the states of actions we cannot produce ourselves. In this case, only exteroceptive (e.g. visual and auditory) knowledge is incorporated in the model (Schubotz, 2007). Studies have shown that the premotor cortex is involved in prediction of both actions and events (Schubotz et al., 2004; Schubotz and von Cramon, 2002, 2003, 2004), making this region a likely candidate for generating forward models.

The predictive coding framework suggests that forward models of sensory states are inverted to infer their most likely causes (Kilner et al., 2007). In the same way, the forward model of an action 'learns' to predict the sensory (e.g. visual) feedback of an action, the forward model can use sensory input, after prior rule-based learning, to infer forthcoming goals and underlying intentions of an action. This inference process is formally equivalent to empirical Bayesian inference (Friston et al., 2011; Kilner et al., 2007). During inference, prior knowledge (*priors*) is re-enacted by sensory input at each level of the action hierarchy. By minimizing the prediction error at all levels, the most likely cause of perceivable consequences is inferred. Importantly, sensory input from secondary cues, i.e. contextual information, is integrated into the

model (Kilner et al., 2007). Thereby, the most likely intention of an action can be inferred even when the kinematics are associated with two equivalently weighted prior intentions (e.g. drinking to quench thirst in a sports situation vs. drinking to engage in social activity in a party situation).

There are many parallels between the concept of action reconstruction and that of predictive coding (Zentgraf et al., 2011): Similar to forward models, action reconstruction is suggested to serve the prediction of forthcoming steps of an unfolding action and to attenuate predictable sensory input. Furthermore, emulation is used to verify or reject goals attributed to observed actions. The framework of action reconstruction also emphasizes the importance of empirical Bayes during goal inference and the possibility to incorporate contextual information in the inference process (Csibra, 2007). However, in the predictive coding framework, Bayesian inference (i.e. the process enabling goal inference) occurs within the motor system (Friston et al., 2011), whereas in the action reconstruction framework, Bayesian inference seems to occur outside the motor system (Csibra, 2007).

1.3 Mental state inference and ToM during action perception

As elaborated in section 1.1.3, intention inference likely draws on cognitive processes that are distinct from processes enabling goal inference. Although simulation theorists proposed a model of how mirror neurons might enable mental state inference (the so-called simulation theory of mind-reading; Gallese and Goldman, 1998), empirical findings do not support this theory (Saxe, 2005).

Processes subserving mental state inference can be subdivided into inference of transitory states (e.g. intentions, emotions, desires) and inference of enduring characteristics (e.g. beliefs, personality traits, social scripts, idiosyncrasies). These inferential processes are thought to build on each other (Malle, 1999; Van Overwalle, 2009): First, transitory states like intentions and desires are identified and categorized. Then, these transitory states shape the corresponding trait that is attributed to the agent. During both inferential processes, several brain regions are consistently activated, i.e. the temporoparietal junction (TPJ), the medial prefrontal cortex (mPFC), and the posterior cingulate cortex (PCC) (Carrington and Bailey, 2009; Gallagher and Frith, 2003; Van Overwalle, 2009). However, a recent meta-analysis found that attributing temporary states engages primarily the TPJ, whereas contemplating about more enduring traits draws mainly on the mPFC (Van Overwalle, 2009).

Attributing states and behaviour to other agents requires a clear distinction of own and others' mental states. The TPJ may have a key role in this differentiation between one's own and another agent's perspective (Brass et al., 2009; Decety and Grezes, 2006; Ruby and Decety, 2001, 2003). This interpretation is corroborated by the finding that transcranial magnetic stimulation(TMS)-induced virtual lesions of the right TPJ evoke out-of-body-experiences and autoscopic hallucinations (Blanke and Arzy, 2005). Moreover, patients with first rank symptoms of schizophrenia (i.e. they do not experience their actions as their own) have a reduction of cerebral blood flow in TPJ as revealed by positron emission tomography (PET; Farrer et al., 2004). Related to these findings, TPJ is activated during body-related perspective taking (Vogeley and Fink, 2003) as well as mental perspective taking in social judgments (Abraham et al., 2008; Ruby and Decety, 2003; Saxe and Kanwisher, 2003; Saxe and Powell, 2006; Young et al., 2010). Finally, the TPJ has been associated with visuospatial orienting (Corbetta et al., 2008).

Considering an integrative interpretation of this functional spectrum, Corbetta and coworkers suggested that "attention signals in the TPJ are important to switch between internal, bodily, or self-perspective and external, environmental, or other's viewpoint" (Corbetta et al., 2008). This interpretation, however, does not account for the likely involvement of TPJ in spatial *transformation* of visual information. Initially, visual information is represented retinotopically in the occipital lobe. Neuropsychological evidence from neglect, arising from damage in either the temporal (Medina et al., 2009) or inferior parietal lobe (Verdon et al., 2010), suggests that at higher processing levels visual information is represented allocentrically in the temporal lobe and egocentrically in the IPL (Kravitz et al., 2011). The TPJ lies at the intersection of the occipital, temporal, and parietal lobe. Hence, from an anatomical point of view, the TPJ is well suited for visuospatial transformation of (visual) information. From there, it is not far to ToM-related functions like mental perspective taking.

Reasoning about other actors' more enduring traits and idiosyncrasies requires the integration of different sources of information in episodic and semantic memory. The mPFC as a highly supramodal region is likely involved in integrational processes that subserve ToM (Amodio and Frith, 2006; Ramnani and Owen, 2004). Ramnani and Owen (2004) proposed that "anterior PFC integrates the outcomes of two or more separate cognitive operations in the pursuit of a higher behavioral goal" (Ramnani and Owen, 2004). Notably, mPFC is a collective term for medial BAs 9, 10, 11 and the anterior cingulate and paracingulate cortex. The functional profiles of these mPFC subregions are heterogeneous and cannot be discussed here in detail. However, with regard to ToM, mPFC most often refers to BA 10, which is activated during false belief tasks (Van Overwalle and Baetens, 2009b). In these tasks, subjects need to relate the beliefs of

others to their own beliefs. The function of BA 10 in ToM may hence be the relation of "several (and, if applied concurrently, conflicting) realities or propositions" (Schubotz, 2011).

Knowledge about personality traits and social habits as well as desires and intentions should improve the predictability of actions. It seems therefore plausible that mentalizing takes part in action perception. Intuitively, both action perception and mentalizing influence each other interdependently: An observed action may activate associated mental states; conversely, knowledge about someone else's mental states influences action perception in a constraining way. Strikingly, however, ToM involvement during action perception is rarely observed in imaging studies. In a recent meta-analysis among 151 studies on action perception, only 13 reported activation of ToM regions. ToM regions were involved during observation of unusual (German et al., 2004), implausible (Brass et al., 2007), or unintended (Buccino et al., 2007) actions, or when subjects were explicitly instructed to contemplate about the actor's intentions (Buccino et al., 2007; de Lange et al., 2008; Iacoboni et al., 2005; Spunt et al., 2010). Some authors concluded that in case of unusual actions "there is no template in one's behavioral repertoire to match them with. This renders the mirror system inadequate and requires extensive thought about the "why" of the action. As a consequence, the mentalizing system takes precedence." (Van Overwalle and Baetens, 2009a; see also Kilner & Frith, 2008). Likewise, Brass et al. concluded that "the mirror network seems to play a role only in situations in which no active inferential processing is required to identify the goal of the observed behaviour because both the action and its stereotypic context are highly familiar and map onto corresponding motor schemes already represented in the observer's action repertoire." These statements imply that during observation of ordinary everyday actions agents do not automatically consider mental states of actors. A compelling explanation for the lack of ToM activity during action observation

has, however, been neglected: In the remaining studies on action perception, putative ToM activity could be equally high in two experimental conditions and is therefore cancelled out in the imaging contrast. This may particularly be the case when two action conditions to be contrasted contain the same characteristic features of someone else's actions, e.g. a third person perspective on the action and the presence of the agent's body including his or her face. It is hence plausible that mental state inference takes part in action perception but its involvement has so far been overseen.

2 Objectives of the thesis

The present work aimed at investigating contextual information that modulates action perception. Behavioral and fMRI measurements were applied. The basic approach was to instruct subjects to attend to video clips of everyday actions and to infer their goals. In the following, the term *action* is used synonymously with *object manipulation*². Primary action information was kept constant and only secondary information (i.e. the independent variable) was manipulated in the experimental conditions. Two studies (Experiment 1 and Experiments 2a and 2b) were carried out to answer the following questions:

- (I) Does (1) a third person perspective (3pp) as opposed to a first person perspective (1pp) and (2) actor information, provided by sight of the actors including their faces, modulate the activity of ToM regions during observation of ordinary everyday action? (Wurm et al., 2011)

- (II) Does contextual information provided by the scene in which an action takes place modulate (1) the speed of action recognition, and (2) the activity of brain

² Actions are sometimes subdivided in the categories hand actions, which are mostly transitive (i.e. they involve objects), and whole body movements (e.g. running, climbing), which are mostly intransitive (e.g. Noppeney, 2008). Object manipulations are defined here as transitive hand actions that induce a change of configuration or state of involved objects.

regions, particularly those involved in action perception and/or processing of contextual scenes? (Wurm and Schubotz, 2011; Wurm et al., in preparation-b)

Exp. 1 investigated the involvement of the ToM network during action perception (Wurm et al., 2011). Previous studies on the role of ToM in action perception concluded that ToM is only involved when actions are implausible, unintended, or new to the observer (Kilner and Frith, 2008; Van Overwalle and Baetens, 2009a). It was hypothesized that the ToM network is also involved in perception of ordinary actions; however, in most imaging studies respective activation might have been cancelled out when contrasting two action conditions that both contain the same visual features characteristic for someone else's actions. Candidate features were a 3pp (as opposed to a 1pp), visual presence of the actors' faces, and episodic person knowledge (e.g. idiosyncrasies and former activities). Subjects were scanned with fMRI while observing actions filmed from a 1pp, a 3pp showing hands only, and a 3pp showing the upper body including the actors' faces (3pp+ hereafter). A subset of 3pp+ trials was preceded by 3pp+ trials involving the same actor (actor repetition) while another subset of 3pp+ trials was preceded by 3pp+ trials involving a different actor (actor switch). Exploiting the mechanisms of repetition suppression (Grill-Spector and Malach, 2001; Hamilton and Grafton, 2006; Kourtzi and Kanwisher, 2000; Thompson-Schill et al., 1999), contrasting actor switch vs. actor repetition trials were expected to capture brain activity associated with the processing of actor-related information. Importantly, subjects were not instructed to mind the actors. Instead, their attention was focussed on the actions by informing them that some trials were followed by verbal action descriptions, matching the preceding action or not, that had to be confirmed or rejected. Whole brain images of the following contrasts were computed: 3pp vs. 1pp, 3pp+ vs. 3pp, and actor

switch vs. actor repetition. In addition, mean percent signal changes were recorded in four independently defined regions of interests (ROIs; left and right TPJ, mPFC, PCC)³.

Exp. 2 aimed at investigating the effect of contextual compatibility of action and the context in which an action takes place (Wurm and Schubotz, 2011; Wurm et al., in preparation-b). To this end, a behavioral (Exp. 2a) and an fMRI experiment (Exp. 2b) were carried out. Both experiments employed actions that were filmed in compatible, incompatible, and neutral contextual scenes (contexts hereafter), i.e. domestic rooms in case of the compatible and incompatible condition and a uniform, white background in case of the neutral condition. The background in the neutral trials hence did not provide explicit cues for attribution to a specific contextual scene or situation. The neutral condition was employed to disentangle facilitatory effects of the compatible and interference effects of the incompatible condition. Facilitatory effects were defined as significant *decrease* in reaction times (RTs) or BOLD responses between compatible and both incompatible and neutral contexts; likewise, interference effects were defined as significant *increase* of RTs or BOLD responses between incompatible and both compatible and neutral contexts. In Exp. 2a, subjects were instructed to observe the actions and press a button as soon as they recognized the presented action. In Exp. 2b, subjects were presented with the same set of actions, plus additional conditions employing pantomime actions (Wurm et al., in preparation-b; see below). Similar to Exp. 1, subjects were instructed to infer the goals of the actions in order to answer verbal action descriptions with respect to preceding actions. The aim of the fMRI experiment was to determine which level of action perception

³ Coordinate peaks of ToM regions were extracted from 14 studies using ToM belief tasks reported in Van Overwalle & Baetens (2009).

(visual analysis of action kinematics or inference of overarching goals/prediction of subsequent action steps) is affected by contextual scene information. Modulation of the kinematic level was assumed to be reflected by changes in neural activity in the OTC (Beauchamp and Martin, 2007; Hamilton and Grafton, 2007; Jastorff et al., 2009; Jastorff and Orban, 2009; Kable and Chatterjee, 2006; Noppeney et al., 2005), whereas modulation of goal inference/action prediction was assumed to be reflected by changes in neural activity in the premotor cortex and the vIPFC (Iacoboni et al., 2005; Johnson-Frey et al., 2003; Schubotz, 2007; Schubotz and von Cramon, 2004, 2009; Van Overwalle and Baetens, 2009a; van Schie et al., 2006).

Pantomime actions, which were measured interleaved with Exp. 2b in a random trial design, were employed to investigate the triadic relationship between context, manipulation, and objects during action perception (Wurm et al., in preparation-b). It was of particular interest to disentangle context-manipulation (CM) and context-object (CO) compatibility effects, since manipulations and their associated objects usually co-occur in particular contexts. To investigate CM and CO (in)compatibility effects independently, actions were uncoupled from their associated objects. To this end, pantomime actions were employed (i.e. pretend actions performed with inappropriate objects; see Schubotz and von Cramon, 2009). In a 2x2 factorial design, (1) actions were either compatible or incompatible with the context, and (2) involved objects were either compatible or incompatible with the scene. Main effects of CM and CO (in)compatibility as well as the interaction of CM and CO associations was analyzed using whole-brain contrast images.

3 General discussion

The present work aimed to investigate contextual modulation of action perception. To this end, two sources of contextual information were experimentally manipulated: Exp. 1 tested the influence of actor-related information, whereas Exp. 2 tested the influence of contextual scene information on action perception. In the following sections, results and limitations of the experiments are summarized and discussed.

3.1 Evidence for involvement of ToM-related brain regions during action perception

Results of Exp. 1 revealed that during action observation (1) a third person perspective, or (2) sight of an actor's face, or (3) seeing a new vs. seeing the same actor are potential triggers of neural activity in ToM regions: A ROI analysis in independently determined ToM regions revealed that observing an action from a third person as compared to a first person perspective increased the BOLD response in the left TPJ. Sight of the whole actor vs. seeing only hands increased activity in bilateral TPJ, mPFC, and PCC. The mPFC and PCC were also enhanced when observing a different actor as in the preceding trial. These findings were further substantiated by significant activation of these regions in whole brain contrast images.

3.1.1 The role of the temporoparietal junction (TPJ) in action perception from a third person perspective

Action observation from a third person as compared to a first person perspective enhanced activity in TPJ bilaterally, although activation was relatively weak and did not pass corrections for multiple comparisons even when using liberal thresholds. However, in the same fMRI session pantomime actions employing inappropriate objects were also tested⁴. Contrasting 3pp vs. 1pp pantomime actions revealed similar activation in bilateral TPJ. It is unlikely that TPJ activation reflect false positive signals in two independent contrasts. As outlined in section 1.3, TPJ is likely involved in visuospatial transformation as well as in mental perspective taking, i.e. perspective taking not necessarily in a bodily sense (Ruby and Decety, 2003). There are hence different interpretations that, although not necessarily mutually exclusive, emphasize different functional involvements in the experiment.

On the one hand, TPJ could be involved in *bodily* perspective taking, i.e. the visuospatial transformation of the action. Adopting the perspective of the actor would optimize the match with the actor's reference frame and thereby facilitate the recognition of the action. This may be of particular importance during observation of complex object manipulation as opposed to pointing or grasping actions that can be performed equally well with both hands (Hesse et al., 2009; Shmuelof and Zohary, 2008). A mapping would not be necessarily required when the actions principally could be inferred by object information. However, the actions were presented intermixed with pantomime actions on inappropriate objects. To determine whether an object is manipulated in an appropriate way or not, subjects had to match object-implied manipulation

⁴ Pantomime actions were implemented to test an additional hypothesis: It was speculated that 3pp features interact with pantomime; however, since no interaction was observed, the factor pantomime is not discussed explicitly in this section. Pantomime actions are discussed in section 3.2.3 in more detail.

information with the observed manipulation. Hence, both real and pantomime actions probably benefitted from perspective taking.

On the other hand, TPJ activity could signify an increased engagement in *mental* perspective taking, i.e. considering that the action is caused by someone else without actually transforming the action into a first person frame of reference. In this case, TPJ activation would accompany action perception from a third person perspective, but would not be beneficial for identification of the action. An argument against this interpretation may be that other regions involved in mental state inference were not enhanced when actions were perceived from a third as compared to a first person perspective. However, taking the actors' perspective may be a prerequisite for other "subsequent" ToM functions, like considering the actors' mental states, that may draw on other regions.

Finally, it was claimed that TPJ might be involved in mechanisms "to keep apart self- and other-related motor representations" (Brass et al., 2009). Hence, the third person perspective might provide a 'not me' signal to tag an observed action as caused by someone else. However, if this were the case, stronger self-other conflict should emerge when an action is observed from a first person perspective, i.e. the action looks like it is performed by myself although it is not. An increased activation would then be expected for the reverse contrast, i.e. when observing someone else's action from a first as compared to a third person perspective.

3.1.2 Considering actors during action perception

Seeing an actor including his or her face compared to seeing only the hands of an actor triggered activation in all considered ToM ROIs, i.e. in PCC, mPFC, and bilateral TPJ. Because

the two conditions differed in the amount of body- and face-related visual information, the observed activations should be treated with caution, particularly activation in the TPJ. The TPJ is located adjacent to the pSTS, which is associated with the recognition of gaze direction (Allison et al., 2000) and biological motion (Puce and Perrett, 2003). Hence, activation in the TPJ may be due to spreading activation from the pSTS. There is evidence, however, that activation of the PCC and the mPFC are not merely deducible to increased demands in the processing of the additional *perceptual* body and face information because these regions also showed increased activation when the actor switched with respect to the preceding action. Here, the amount of visual information is identical in both conditions. The only difference was that in the preceding trial either the same or a different actor was shown. Activation was found in the BA 10 and anterior pregenual cingulate cortex and the posterior precuneus. These regions are activated in a number of studies on face familiarity (Gobbini and Haxby, 2007; Gobbini et al., 2004; Leibenluft et al., 2004; Trinkler et al., 2009). The mPFC is related to the formation of person-related knowledge whereas the precuneus and the PCC are associated with the retrieval of person-related knowledge from episodic memory (Gobbini et al., 2004; Todorov et al., 2007). In actor switch trials, the formation and recollection of actor-related knowledge was probably more demanding than in actor repetition trials, increasing the activation in the mPFC and the precuneus/PCC.

Two major kinds of information could be associated with a particular actor:

First, subjects could acquire knowledge about the agent's physical appearance, his or her face, and emotional expressions. With particular respect to face-related information, it is notable that there was no repetition suppression in the fusiform face area (FFA), an area that is consistently activated during the processing of face information (Kanwisher et al., 1997). However, a lack of FFA activity modulation cannot be treated as evidence that formation and

retrieval of face-related information did not play a role in Experiment 1. Interestingly, two other studies employing actor switches vs. repetitions did not observe attenuation in the FFA either (Kable and Chatterjee, 2006; Wiggett and Downing, 2011).

Second, it is plausible that subjects acquired knowledge about the different actions associated with a particular person, because the task drew the subjects' attention to the actions. In the course of the experiment, each actor appeared ten times; however, the actions performed by one actor were not related to each other. It may be an automatic tendency to retrieve previous actions when seeing the same actor again in order to relate these actions to each other and to the current one, in order to infer possible long-term goals.

On the basis of Experiment 1, it is not possible to disentangle these two possibilities. However, a follow-up fMRI experiment aimed at differentiating actor-related and action-related coherence building during action perception (Hrkac et al., in preparation; Wurm et al., in preparation-a). Subjects were presented with actions that (1) belonged to an overarching long-term goal or not and (2) were performed by the same or different actors. Irrespective of whether performed by the same or different actors, activation was found in the precuneus for coherent actions. Conversely, no increased neural responses were found for actor coherence when considered independently of the action's coherence⁵. This result suggests that precuneus activation in Exp. 1 is due to a retrieval of previous actions of the same actor.

⁵ Instead, a *decrease* of activation was found in bilateral FFA and ventromedial PFC.

3.2 The influence of contextual information on action perception

In Exp. 2 both behavioral and fMRI findings revealed significant effects of contextual scene information on action perception. Reaction times were slower and BOLD responses were increased in the left vlPFC when actions took place in incompatible as compared to compatible or to neutral contexts. As both experiments revealed no significant differences between compatible and neutral contexts, results point to interfering rather than facilitatory influences of contextual information on action perception.

Activation of the left vlPFC suggests that contextual modulation affected a high processing level of action analysis. The most posterior part of vlPFC, BA 44, is associated with the structuring of event sequences (Grafman, 2002; Schubotz and Fiebach, 2006). More anterior parts, i.e. BA 45 and BA 47, are involved in retrieval (BA 47) and selection (BA 45) of semantic memory (Badre and Wagner, 2007). Hence, the functional profiles of vlPFC subregions suggest that actions were embedded in overarching action sequences that resolve the conflict between observed action and incompatible, context-related actions in the best way possible. This requires the retrieval and selection of action memories.

3.2.1 Neutral conditions and the differentiation of facilitation and interference

Both behavioral results of Exp. 2a and fMRI results of Exp. 2b provide converging evidence for interference effects when actions were embedded in contexts normally not associated with the action. Conversely, there was no evidence for facilitatory effects when actions took place in their regular context. However, the neutral condition, which was employed

to differentiate interference and facilitation effects, needs to be interpreted with caution: As the background of the neutral condition was uniformly white and devoid of objects, it differed perceptually from the compatible and incompatible conditions on several levels of visual processing (e.g. figure-ground segmentation, surface extraction, detection of contours and shapes, object recognition). Hence, it cannot be entirely ruled out that putative facilitatory effects were compensated by lower perceptual demands during observation of the neutral condition trials.

It is surprisingly hard to tackle the problem of implementing a neutral condition to disentangle facilitation and interference effects. In studies on the influence of contextual compatibility of scenes on object recognition, a neutral condition is often omitted; however, observed effects are interpreted as facilitatory effects although no control condition was implemented to differentiate facilitation and interference (for a review see Bar, 2004). Studies employing a neutral condition often control for low-level visual features, e.g. by using a scrambled background (Boyce et al., 1989; Ganis and Kutas, 2003). However, a scrambled background may be considered unnatural and thereby provoke interference effects. Hence, adjusting perceptual properties of baseline and test conditions still entails the danger to misinterpret reaction time differences between compatible and neutral conditions as being facilitatory.

3.2.2 Context-manipulation (CM) vs. context-object (CO) associations

The employed actions always involved objects. Hence, objects were compatible or incompatible with the contexts as well. Importantly, action compatibility was always coupled,

and thus confounded, with object compatibility. Hence, observed effects could be due to context-object rather than context-action compatibility.

Two scenarios may be considered: First, context-object incompatibility effects lead to interference irrespective of the task: Actions were recognized equally well in both compatible and incompatible contexts, but only the objects, which of course were also perceived during the presentation of the actions, evoked interference effects. Second, as objects are recognized faster when they are embedded in compatible as compared to incompatible contextual scenes (Biederman et al., 1982; Ganis and Kutas, 2003) and manipulation as well as object information was likely exploited to identify the actions, contextually affected object recognition might in turn have influenced action recognition. However, objects can be recognized within 150 ms (Thorpe et al., 1996) and responses to object recognition occur 600 ms after stimulus onset (Ganis and Kutas, 2003). In both Exp. 2a and 2b actions started around 700 ms after video onset, i.e. well after objects have been recognized. Thus, one might argue that CO compatibility effects should not substantially affect recognition of actions, that unfolded much later in time. This argument, however, can account only for those objects that can be manipulated in different ways and therefore result in different actions. In these cases, subjects had to wait at least 700 ms until they were able to identify one of the possible actions associated with the objects. Since responses occurred relatively late, i.e. around 1400 ms after video onset, it seems that subjects indeed waited until actions started to unfold. However, in order to unambiguously show that context affects action recognition in the absence of context-object effects, experiments are needed where (1) no object information is provided or (2) context-object (CO) compatibility is uncoupled from context-manipulation (CM) compatibility during action recognition.

With regard to the first option, two yet unpublished behavioral studies found significant context compatibility effects when (1) actions were pixelated so that objects were unidentifiable and (2) pantomime actions without objects were employed. The first study used exactly the same experimental design as Exp. 2a. The only differences were that videos were in black-and-white and the section of the videos where hands and objects occurred was pixelated so that objects were unidentifiable in single static video frames. Behavioral results revealed that actions were recognized more easily when they took place in compatible as compared to incompatible or neutral contexts. Furthermore, actions were recognized faster in compatible as compared to incompatible or neutral trials. Reaction time differences, however, were not significant (see Appendix Figure A1). Second, in a developmental study, children at the ages of four to eight were presented video clips of pantomime actions that did not involve objects. Pantomime actions took place in kitchen contexts, children's room contexts, or neutral contexts as described above. The children watched the videos in a quiet room at the kindergarten or at school and were instructed to name the presented pantomime actions. In each age cohort, children successfully identified actions more often when they took place in compatible as compared to incompatible contexts (see Appendix Figure A2).

Hence, both follow-up experiments provide evidence that contextual information modulates action recognition in the absence of confounding CO compatibility. Importantly, this does not mean that CO compatibility effects can be excluded in Exp. 2a, but that contextual information is indeed capable to affect action recognition, which has not been shown so far.

Furthermore, both experiments provide converging evidence for facilitation instead of interference effects. With respect to the problems of the neutral condition discussed in section 3.2.1, it is unlikely that putative lower visual processing demands in the uniform neutral

background substantially affected the recognisability of the actions, because subjects had three seconds to identify the action. Importantly, if perceptual differences between neutral and (in)compatible conditions affected the recognisability, one should rather expect decreased error rates in the neutral relative to the incompatible condition, because the contours of pixelated actions or pantomimes should stand out from the white background of the neutral context more clearly. Indeed studies on the influence of contextual scene information on object recognition found even lower error rates for recognition of objects in neutral as compared to compatible context (Davenport and Potter, 2004; Murphy and Wisniewski, 1989).

The results of the follow-up experiment of Exp 2a and the developmental study do not contradict the results of Exp. 2a: First, the former experiment collected reaction times and the latter experiments collected error rates. These measures must not necessarily provide similar results. Notably, interference effects can only be detected in the reaction time measures. This is because interference can only arise after successful recognition of the action, i.e. in the correctly identified actions⁶. Most importantly, in both the follow-up experiment of Exp. 2a and the developmental study, actions were much harder to identify as evident from the high error rates. The results corroborate the intuition that contextual information is particularly helpful when primarily relevant information is not or only partially available. In sum, there is evidence for contextual compatibility effects on action recognition in the absence of confounding object information.

With respect to the second option, in Exp. 2b CO compatibility was uncoupled from CM compatibility by employing pantomime actions involving inappropriate objects; e.g. cracking an

⁶ Incompatibility effects in error rates rather reflect "misguidance" effects, i.e. when subjects gave incorrect or no responses because the incompatible context impeded identification of the action or subjects confused the observed action with another action compatible to the context.

egg was mimed with an orange and an orange squeezer (Wurm and Schubotz, 2011). Thereby, it was possible to manipulate CO compatibility irrespectively of CM compatibility, yielding a 2x2 factorial design: First, with regard to the factor CM, manipulations were either compatible or incompatible with the context. Second, with regard to the factor CO, pantomime actions involved objects that were either compatible or incompatible with the context. If vIPFC activation observed in Exp. 2b is due to CO incompatibility, a similar activation should be observed when contrasting pantomime involving context-incompatible vs. context-compatible objects. This was not the case (Wurm and Schubotz, 2011). It is hence unlikely that activation of the vIPFC in Exp. 2b is due to mere CO incompatibility irrespectively of the action. However, negative results do not prove the absence of effects. Hence, absence of increased BOLD responses in the pantomime control cannot conclusively prove the absence of CO compatibility effects in Exp. 2b.

3.2.3 The context-object-manipulation (COM) triad

The employment of pantomime actions allowed analyzing the interdependencies in the triadic relationship of context, object, and manipulation information. Activation of (1) regions belonging to the cortical motor system (left PMv, the left aIPS, and bilateral SMA) for CM compatibility and (2) the LOC for CO compatibility provide evidence that contextual information affects the neural signatures of manipulation and object recognition (Wurm et al., in preparation-b).

Importantly, increased activation was found for *compatibility* as compared to incompatibility of context and either manipulation or objects, whereas during recognition of

normal actions increased activation was found for *incompatibility* as compared to compatibility of context and actions. This putative discrepancy can be resolved when the mechanism of action inference from pantomime is considered in more detail: Inferring actions from pantomime requires focusing attention on manipulation information, because only the posture and movements of the hands give indication about mimed actions. During analysis of manipulation information, object information that matches hand posture and movements is activated. Conversely, perceived, invalid object information and associated manipulation information has to be suppressed. During recognition of pantomime there is hence conflicting activation of valid object and manipulation information and invalid object and manipulation information. In the special case that both objects (e.g. orange squeezer and orange) and manipulation (cracking an egg) are compatible with the context (kitchen), the conflict is strongest, because valid and invalid object and manipulation information belong to the same contextual category. This constellation may be compared with either the condition where only the objects (e.g. pencil and pencil-sharpener) or only the manipulation (e.g. sharpening a pencil) are incompatible with the context. In both cases, the conflict is less strong because valid and invalid object and manipulation information belong to a different contextual category.

If this mechanism exists, activation due to the same vs. different contextual categories of valid and invalid object and manipulation information should also be observed when no context information is presented in the videos. Indeed, very similar activation of the bilateral OTC and the dorsal attention network was found when contrasting pantomime actions in neutral settings that were performed with objects of the same or a different contextual category as the manipulation (Wurm et al., in preparation-b). This finding provides strong support for the proposed mechanism.

The overall results have some important implications for the associations between context, object, and manipulation information in semantic memory: First, object and manipulation information seem to be encoded in clusters in semantic memory, with stronger associations within than between contextual categories. These categories likely reflect the frequency of co-occurrence in a particular context. Second, these networks can be activated by the presence of contextual information: Contextual activation of perceived, invalid object information increases the demands to retrieve object information that matches the observed manipulation. Hence, in comparison with the constellation where both retrieval of valid object and manipulation information is very demanding, object information is hard to retrieve whereas manipulation information is more easily retrieved. Therefore, activation of the motor system surfaces to a stronger degree. Conversely, if perceived invalid object information is not also activated by the context, retrieval of valid object information is less demanding. In comparison to the constellation where both retrieval of valid object and manipulation information is strongly demanding, regions involved in object processing surfaces to a stronger degree.

3.3 Future prospects

With regard to Exp. 1, results do not provide clear evidence for a particular function of TPJ in perception of actions from a third person perspective (see section 3.1.1). The first hypothesis proposed that TPJ is involved in *bodily* perspective taking, i.e. to adopt the actor's perspective on the action. The second hypothesis proposed that TPJ activation reflects *mental* perspective taking, i.e. the third person perspective triggers mentalizing processes to a stronger degree than a first person perspective. Both hypotheses could be tested in an fMRI experiment

where the action is rotated parametrically from the first person perspective (0°) to the third person perspective (180°). If TPJ activation increases in correlation with the degree of rotation, bodily perspective taking would be more likely, because the difficulty of bodily perspective taking should increase parametrically with rotation of the action⁷. Conversely, mental perspective taking should be triggered as soon as the 'like me' impression of the first person perspective is disrupted (e.g. when the action is rotated 90°). Hence, no parametric effect should be observed. A more rigorous test, however, would be to disrupt TPJ in a TMS-experiment. TMS should interact with the perspective from which an action is observed, decreasing the speed of action recognition particularly for the third person perspective. If TPJ activation reflects increased mentalizing, i.e. a process unlikely to be required or beneficial for action recognition, TMS should not affect the speed of action recognition. Problematically, both experiments would *positively* test the bodily perspective taking hypothesis, but only *negatively* test the mental perspective taking hypothesis. It is hard to design an experiment that positively tests the mental perspective taking hypothesis because mentalizing is unlikely to be required or beneficial for identification of an action. A hypothetical experiment would have to measure the degree to which mentalizing accompanies perception of actions from a third person perspective, e.g. by employing an adequate questionnaire.

The observed effect for actor switches in Exp. 1 suggests that actions performed by the same actor were automatically related to one another (Wurm et al., 2011). Future experiments should investigate relevant factors that serve as cues for coherence building in action perception. As mentioned in section 3.1.2, a current project investigates the factors constancy of the acting

⁷ In a comparable fMRI experiment, subjects had to judge the laterality of parametrically rotated hands. Reaction times and activity in several parietal regions, including right angular gyrus partially overlapping with activation found in Exp. 1, increased parametrically with the degree of hand rotation (de Lange et al., 2005).

agent and constancy of the overarching goal; however, other factors, like constancy of the context and constancy of involved objects, should also be considered.

With regard to Exp. 2, it remains to be solved what detail of the provided context, i.e. a rich domestic setting background entailing a number of context-specific objects, evoked interference in the case of incompatibility. Was it the global context information or rather particular objects in the background? An argument for the former option is that contextual information is extracted without making saccades to particular context-specifying objects (Metzger and Antes, 1983). However, if observed actions are automatically embedded into sequences of actions serving a superordinate goal, single objects in the background should provide important cues about possible overarching goals. Hence, one idea could be to construct overarching goal contexts with objects that are compatible (e.g. utensils for making pancake) or incompatible (e.g. utensils for making coffee) with the action (cracking an egg). If objects are recognized, a similar interference effect should be observed when the objects are incompatible with the long-term goal, irrespective of general contextual affiliation. Whether objects in the background of the domestic settings are explicitly focused during action recognition could easily be tested by using eye tracking.

Finally, a crucial question concerns the temporal dynamics of the context, object, and manipulation perception. There is not yet consensus whether context recognition precedes and therefore is involved in object recognition. Likewise, although it seems intuitively plausible that recognition of manipulation information is by nature slower than object recognition, there is not yet experimental data with respect to the temporal dynamics of object recognition in action perception and vice versa. Hence, it should be tested, by employing behavioral and/or

electrophysiological measures, whether there is a cascade in the perception of context, object, and manipulation information.

4 References

- Abraham, A., Werning, M., Rakoczy, H., von Cramon, D.Y., Schubotz, R.I., 2008. Minds, persons, and space: an fMRI investigation into the relational complexity of higher-order intentionality. *Conscious Cogn* 17, 438-450.
- Allison, T., Puce, A., McCarthy, G., 2000. Social perception from visual cues: role of the STS region. *Trends Cogn Sci* 4, 267-278.
- Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. *Nat Rev Neurosci* 7, 268-277.
- Badre, D., 2008. Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends Cogn Sci* 12, 193-200.
- Badre, D., Wagner, A.D., 2007. Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia* 45, 2883-2901.
- Bar, M., 2004. Visual objects in context. *Nat Rev Neurosci* 5, 617-629.
- Bar, M., Ullman, S., 1996. Spatial context in recognition. *Perception* 25, 343-352.
- 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, 149-159.
- Beauchamp, M.S., Lee, K.E., Haxby, J.V., Martin, A., 2003. fMRI responses to video and point-light displays of moving humans and manipulable objects. *J Cogn Neurosci* 15, 991-1001.
- Beauchamp, M.S., Martin, A., 2007. Grounding object concepts in perception and action: evidence from fMRI studies of tools. *Cortex* 43, 461-468.
- Bekkering, H., Wohlschlagel, A., Gattis, M., 2000. Imitation of gestures in children is goal-directed. *Q J Exp Psychol A* 53, 153-164.
- Biederman, I., Glass, A.L., Stacy, E.W., Jr., 1973. Searching for objects in real-world scenes. *J Exp Psychol* 97, 22-27.
- Biederman, I., Mezzanotte, R.J., Rabinowitz, J.C., 1982. Scene perception: detecting and judging objects undergoing relational violations. *Cogn Psychol* 14, 143-177.
- Blakemore, S.J., Wolpert, D.M., Frith, C.D., 1998. Central cancellation of self-produced tickle sensation. *Nat Neurosci* 1, 635-640.

-
- Blanke, O., Arzy, S., 2005. The out-of-body experience: disturbed self-processing at the temporo-parietal junction. *Neuroscientist* 11, 16-24.
- Boyce, S.J., Pollatsek, A., Rayner, K., 1989. Effect of background information on object identification. *J Exp Psychol Hum Percept Perform* 15, 556-566.
- Brass, M., Ruby, P., Spengler, S., 2009. Inhibition of imitative behaviour and social cognition. *Philos Trans R Soc Lond B Biol Sci* 364, 2359-2367.
- Brass, M., Schmitt, R.M., Spengler, S., Gergely, G., 2007. Investigating action understanding: inferential processes versus action simulation. *Curr Biol* 17, 2117-2121.
- Buccino, G., Baumgaertner, A., Colle, L., Buechel, C., Rizzolatti, G., Binkofski, F., 2007. The neural basis for understanding non-intended actions. *Neuroimage* 36 Suppl 2, T119-127.
- Buccino, G., Binkofski, F., Fink, G.R., Fadiga, L., Fogassi, L., Gallese, V., Seitz, R.J., Zilles, K., Rizzolatti, G., Freund, H.J., 2001. Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. *Eur J Neurosci* 13, 400-404.
- Buxbaum, L.J., Kalenine, S., 2010. Action knowledge, visuomotor activation, and embodiment in the two action systems. *Ann N Y Acad Sci* 1191, 201-218.
- Byrne, R.W., Russon, A.E., 1998. Learning by imitation: a hierarchical approach. *Behav Brain Sci* 21, 667-684; discussion 684-721.
- Carpenter, W.B., 1852. On the influence of suggestion in modifying and directing muscular movement, independently of volition. *Proceedings of the Royal Institution*, 147-154.
- Carrington, S.J., Bailey, A.J., 2009. Are there theory of mind regions in the brain? A review of the neuroimaging literature. *Hum Brain Mapp* 30, 2313-2335.
- 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, 1148-1167.
- Cattaneo, L., Fabbri-Destro, M., Boria, S., Pieraccini, C., Monti, A., Cossu, G., Rizzolatti, G., 2007. Impairment of actions chains in autism and its possible role in intention understanding. *Proc Natl Acad Sci U S A* 104, 17825-17830.
- Chun, M.M., Jiang, Y., 1998. Contextual cueing: implicit learning and memory of visual context guides spatial attention. *Cogn Psychol* 36, 28-71.
- Corbetta, M., Patel, G., Shulman, G.L., 2008. The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58, 306-324.
- 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 Psychol (Amst)* 124, 60-78.
- Davenport, J.L., Potter, M.C., 2004. Scene consistency in object and background perception. *Psychol Sci* 15, 559-564.
- Davidson, P.R., Wolpert, D.M., 2005. Widespread access to predictive models in the motor system: a short review. *J Neural Eng* 2, S313-319.
- De Graef, P., Christiaens, D., d'Ydewalle, G., 1990. Perceptual effects of scene context on object identification. *Psychol Res* 52, 317-329.
- de Lange, F.P., Hagoort, P., Toni, I., 2005. Neural topography and content of movement representations. *J Cogn Neurosci* 17, 97-112.
- de Lange, F.P., Spronk, M., Willems, R.M., Toni, I., Bekkering, H., 2008. Complementary systems for understanding action intentions. *Curr Biol* 18, 454-457.
- Decety, J., Grezes, J., 2006. The power of simulation: imagining one's own and other's behavior. *Brain Res* 1079, 4-14.
- Desmurget, M., Reilly, K.T., Richard, N., Szathmari, A., Mottolese, C., Sirigu, A., 2009. Movement intention after parietal cortex stimulation in humans. *Science* 324, 811-813.
- di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., Rizzolatti, G., 1992. Understanding motor events: a neurophysiological study. *Exp Brain Res* 91, 176-180.
- Downing, P.E., Chan, A.W., Peelen, M.V., Dodds, C.M., Kanwisher, N., 2006. Domain specificity in visual cortex. *Cereb Cortex* 16, 1453-1461.
- Downing, P.E., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. *Science* 293, 2470-2473.
- Dum, R.P., Strick, P.L., 2002. Motor areas in the frontal lobe of the primate. *Physiol Behav* 77, 677-682.
- Farivar, R., 2009. Dorsal-ventral integration in object recognition. *Brain Res Rev* 61, 144-153.
- Farrer, C., Franck, N., Frith, C.D., Decety, J., Georgieff, N., d'Amato, T., Jeannerod, M., 2004. Neural correlates of action attribution in schizophrenia. *Psychiatry Res* 131, 31-44.
- Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., Rizzolatti, G., 2005. Parietal lobe: from action organization to intention understanding. *Science* 308, 662-667.
- Friedman, A., 1979. Framing pictures: the role of knowledge in automatized encoding and memory for gist. *J Exp Psychol Gen* 108, 316-355.

- Friston, K., Mattout, J., Kilner, J., 2011. Action understanding and active inference. *Biol Cybern* 104, 137-160.
- Frith, C., Frith, U., 2005. Theory of mind. *Curr Biol* 15, R644-646.
- Frith, C., Frith, U., 2006. The neural basis of mentalizing. *Neuron* 50, 531-534.
- Gallagher, H.L., Frith, C.D., 2003. Functional imaging of 'theory of mind'. *Trends Cogn Sci* 7, 77-83.
- Gallese, V., Fadiga, L., Fogassi, L., Rizzolatti, G., 1996. Action recognition in the premotor cortex. *Brain* 119 (Pt 2), 593-609.
- Gallese, V., Goldman, A., 1998. Mirror neurons and the simulation theory of mind-reading. *Trends Cogn Sci* 2, 493-501.
- Ganis, G., Kutas, M., 2003. An electrophysiological study of scene effects on object identification. *Brain Res Cogn Brain Res* 16, 123-144.
- German, T.P., Niehaus, J.L., Roarty, M.P., Giesbrecht, B., Miller, M.B., 2004. Neural correlates of detecting pretense: automatic engagement of the intentional stance under covert conditions. *J Cogn Neurosci* 16, 1805-1817.
- Geyer, S., Matelli, M., Luppino, G., Zilles, K., 2000. Functional neuroanatomy of the primate isocortical motor system. *Anat Embryol (Berl)* 202, 443-474.
- Giese, M.A., Poggio, T., 2003. Neural mechanisms for the recognition of biological movements. *Nat Rev Neurosci* 4, 179-192.
- Gobbini, M.I., Haxby, J.V., 2007. Neural systems for recognition of familiar faces. *Neuropsychologia* 45, 32-41.
- Gobbini, M.I., Leibenluft, E., Santiago, N., Haxby, J.V., 2004. Social and emotional attachment in the neural representation of faces. *Neuroimage* 22, 1628-1635.
- 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., 2007. Evidence for a distributed hierarchy of action representation in the brain. *Hum Mov Sci* 26, 590-616.
- Graziano, M., 2006. The organization of behavioral repertoire in motor cortex. *Annu Rev Neurosci* 29, 105-134.
- Graziano, M.S., Taylor, C.S., Moore, T., 2002. Complex movements evoked by microstimulation of precentral cortex. *Neuron* 34, 841-851.

-
- Grezes, J., Decety, J., 2001. Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Hum Brain Mapp* 12, 1-19.
- Grill-Spector, K., Kourtzi, Z., Kanwisher, N., 2001. The lateral occipital complex and its role in object recognition. *Vision Res* 41, 1409-1422.
- Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzhak, Y., Malach, R., 1999. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24, 187-203.
- Grill-Spector, K., Malach, R., 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol (Amst)* 107, 293-321.
- Gronau, N., Neta, M., Bar, M., 2008. Integrated contextual representation for objects' identities and their locations. *J Cogn Neurosci* 20, 371-388.
- Grush, R., 2004. The emulation theory of representation: motor control, imagery, and perception. *Behav Brain Sci* 27, 377-396; discussion 396-442.
- Hamilton, A.F., 2009. Goals, intentions and mental states: challenges for theories of autism. *J Child Psychol Psychiatry* 50, 881-892.
- Hamilton, A.F., Grafton, S.T., 2006. Goal representation in human anterior intraparietal sulcus. *J Neurosci* 26, 1133-1137.
- Hamilton, A.F., Grafton, S.T., 2007. The motor hierarchy: From kinematics to goals and intentions. In: Haggard P, Rossetti Y, Kawato M, editors. *Sensorimotor Foundations of Higher Cognition*. Oxford: Oxford University Press., pp 381-408.
- Hamilton, A.F., Grafton, S.T., 2008. Action outcomes are represented in human inferior frontoparietal cortex. *Cereb Cortex* 18, 1160-1168.
- 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, 873-889.
- Hesse, M.D., Sparing, R., Fink, G.R., 2009. End or means--the "what" and "how" of observed intentional actions. *J Cogn Neurosci* 21, 776-790.
- Hickok, G., 2009. Eight problems for the mirror neuron theory of action understanding in monkeys and humans. *J Cogn Neurosci* 21, 1229-1243.
- Hickok, G., Hauser, M., 2010. (Mis)understanding mirror neurons. *Curr Biol* 20, R593-594.
- Hietanen, J.K., Perrett, D.I., 1993. Motion sensitive cells in the macaque superior temporal polysensory area. I. Lack of response to the sight of the animal's own limb movement. *Exp Brain Res* 93, 117-128.

-
- Hollingworth, A., Henderson, J.M., 1998. Does consistent scene context facilitate object perception? *J Exp Psychol Gen* 127, 398-415.
- Hrkac, M., Wurm, M.F., Morikawa, Y., Schubotz, R.I., in preparation.
- Husain, M., Nachev, P., 2007. Space and the parietal cortex. *Trends Cogn Sci* 11, 30-36.
- 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, e79.
- Jacob, P., 2008. What Do Mirror Neurons Contribute to Human Social Cognition? *Mind and Language* 32, 91-223.
- Jacob, P., 2009. The tuning-fork model of human social cognition: a critique. *Conscious Cogn* 18, 229-243.
- Jacob, P., Jeannerod, M., 2005. The motor theory of social cognition: a critique. *Trends Cogn Sci* 9, 21-25.
- James, W., 1890. *Principles of Psychology*. Holt, New York.
- Jastorff, J., Clavagner, S., Gergely, G., Orban, G.A., 2010. Neural mechanisms of understanding rational actions: middle temporal gyrus activation by contextual violation. *Cereb Cortex* 21, 318-329.
- Jastorff, J., Kourtzi, Z., Giese, M.A., 2009. Visual learning shapes the processing of complex movement stimuli in the human brain. *J Neurosci* 29, 14026-14038.
- Jastorff, J., Orban, G.A., 2009. Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. *J Neurosci* 29, 7315-7329.
- Jeannerod, M., 1994. The representing brain: Neural correlates of motor intention and imagery. *Behav Brain Sci* 17, 187-245.
- Jeannerod, M., 2006. *Motor Cognition: What Action Tells the Self*. Oxford University Press, New York.
- Jeannerod, M., Arbib, M.A., Rizzolatti, G., Sakata, H., 1995. Grasping objects: the cortical mechanisms of visuomotor transformation. *Trends Neurosci* 18, 314-320.
- Jenkins, L.J., Yang, Y.J., Goh, J., Hong, Y.Y., Park, D.C., 2010. Cultural differences in the lateral occipital complex while viewing incongruent scenes. *Soc Cogn Affect Neurosci* 5, 236-241.
- Johnson-Frey, S.H., Maloof, F.R., Newman-Norlund, R., Farrer, C., Inati, S., Grafton, S.T., 2003. Actions or hand-object interactions? Human inferior frontal cortex and action observation. *Neuron* 39, 1053-1058.

-
- Joubert, O.R., Fize, D., Rousselet, G.A., Fabre-Thorpe, M., 2008. Early interference of context congruence on object processing in rapid visual categorization of natural scenes. *J Vis* 8, 11-18.
- Kable, J.W., Chatterjee, A., 2006. Specificity of action representations in the lateral occipitotemporal cortex. *J Cogn Neurosci* 18, 1498-1517.
- Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17, 4302-4311.
- Keysers, C., Perrett, D.I., 2004. Demystifying social cognition: a Hebbian perspective. *Trends Cogn Sci* 8, 501-507.
- Kilner, J.M., 2011. More than one pathway to action understanding. *Trends Cogn Sci* 15, 352-357.
- Kilner, J.M., Friston, K.J., Frith, C.D., 2007. Predictive coding: an account of the mirror neuron system. *Cogn Process* 8, 159-166.
- Kilner, J.M., Frith, C.D., 2008. Action observation: inferring intentions without mirror neurons. *Curr Biol* 18, R32-33.
- Kirk, U., 2008. The neural basis of object-context relationships on aesthetic judgment. *PLoS One* 3, e3754.
- Kourtzi, Z., Kanwisher, N., 2000. Cortical regions involved in perceiving object shape. *J Neurosci* 20, 3310-3318.
- Kravitz, D.J., Saleem, K.S., Baker, C.I., Mishkin, M., 2011. A new neural framework for visuospatial processing. *Nat Rev Neurosci* 12, 217-230.
- Kurby, C.A., Zacks, J.M., 2008. Segmentation in the perception and memory of events. *Trends Cogn Sci* 12, 72-79.
- Leibenluft, E., Gobbini, M.I., Harrison, T., Haxby, J.V., 2004. Mothers' neural activation in response to pictures of their children and other children. *Biol Psychiatry* 56, 225-232.
- Liepelt, R., Ullsperger, M., Obst, K., Spengler, S., von Cramon, D.Y., Brass, M., 2009. Contextual movement constraints of others modulate motor preparation in the observer. *Neuropsychologia* 47, 268-275.
- Liepelt, R., Von Cramon, D.Y., Brass, M., 2008. How do we infer others' goals from non-stereotypic actions? The outcome of context-sensitive inferential processing in right inferior parietal and posterior temporal cortex. *Neuroimage* 43, 784-792.

-
- Majdandzic, J., Bekkering, H., van Schie, H.T., Toni, I., 2009. Movement-Specific Repetition Suppression in Ventral and Dorsal Premotor Cortex during Action Observation. *Cereb Cortex* 19, 2736-2745.
- Majdandzic, J., Grol, M.J., van Schie, H.T., Verhagen, L., Toni, I., Bekkering, H., 2007. The role of immediate and final goals in action planning: an fMRI study. *Neuroimage* 37, 589-598.
- Malle, B.F., 1999. How people explain behavior: a new theoretical framework. *Pers Soc Psychol Rev* 3, 23-48.
- Marsh, A.A., Kozak, M.N., Wegner, D.M., Reid, M.E., Yu, H.H., Blair, R.J., 2011. The neural substrates of action identification. *Soc Cogn Affect Neurosci* 5, 392-403.
- Medina, J., Kannan, V., Pawlak, M.A., Kleinman, J.T., Newhart, M., Davis, C., Heidler-Gary, J.E., Herskovits, E.H., Hillis, A.E., 2009. Neural substrates of visuospatial processing in distinct reference frames: evidence from unilateral spatial neglect. *J Cogn Neurosci* 21, 2073-2084.
- Metzger, R.L., Antes, J.R., 1983. The nature of processing early in picture perception. *Psychol Res* 45, 267-274.
- Miall, R.C., 2003. Connecting mirror neurons and forward models. *Neuroreport* 14, 2135-2137.
- Milner, A.D., Goodale, M.A., 1995. *The Visual Brain in Action*. Oxford University Press, Oxford.
- Murphy, G.L., Wisniewski, E.J., 1989. Categorizing objects in isolation and in scenes: what a superordinate is good for. *J Exp Psychol Learn Mem Cogn* 15, 572-586.
- Noppeney, U., Josephs, O., Kiebel, S., Friston, K.J., Price, C.J., 2005. Action selectivity in parietal and temporal cortex. *Brain Res Cogn Brain Res* 25, 641-649.
- Oliva, A., Torralba, A., 2007. The role of context in object recognition. *Trends Cogn Sci* 11, 520-527.
- Oram, M.W., Perrett, D.I., 1996. Integration of form and motion in the anterior superior temporal polysensory area (STPa) of the macaque monkey. *J Neurophysiol* 76, 109-129.
- Ortigue, S., Sinigaglia, C., Rizzolatti, G., Grafton, S.T., 2010. Understanding actions of others: the electrodynamics of the left and right hemispheres. A high-density EEG neuroimaging study. *PLoS One* 5, e12160.
- Ortigue, S., Thompson, J.C., Parasuraman, R., Grafton, S.T., 2009. Spatio-temporal dynamics of human intention understanding in temporo-parietal cortex: a combined EEG/fMRI repetition suppression paradigm. *PLoS One* 4, e6962.
- Pacherie, E., 2000. The content of intentions. *Mind and Language* 15, 400-432.

- Palmer, S.E., 1975. The effects of contextual scenes on the identification of objects. *Mem Cognit* 3, 519-526.
- Peelen, M.V., Downing, P.E., 2007. The neural basis of visual body perception. *Nat Rev Neurosci* 8, 636-648.
- Penfield, W., Rasmussen, T., 1950. *The Cerebral Cortex of Man. A Clinical Study of Localization of Function.* . The Macmillan Comp., New York.
- Petrides, M., 2005. Lateral prefrontal cortex: architectonic and functional organization. *Philos Trans R Soc Lond B Biol Sci* 360, 781-795.
- Peuskens, H., Vanrie, J., Verfaillie, K., Orban, G.A., 2005. Specificity of regions processing biological motion. *Eur J Neurosci* 21, 2864-2875.
- Porter, R., 1985. The corticomotoneuronal component of the pyramidal tract: corticomotoneuronal connections and functions in primates. *Brain Res* 357, 1-26.
- Premack, D., Woodruff, G., 1987. Does the chimpanzee have a theory of mind? *Behav Brain Sci* 1, 515-526.
- Prinz, W., 1997. Perception and Action Planning. *European Journal of Cognitive Psychology* 9, 129-154.
- Puce, A., Perrett, D., 2003. Electrophysiology and brain imaging of biological motion. *Philos Trans R Soc Lond B Biol Sci* 358, 435-445.
- Ramnani, N., Owen, A.M., 2004. Anterior prefrontal cortex: insights into function from anatomy and neuroimaging. *Nat Rev Neurosci* 5, 184-194.
- Rizzolatti, G., Craighero, L., 2004. The mirror-neuron system. *Annu Rev Neurosci* 27, 169-192.
- Rizzolatti, G., Fogassi, L., Gallese, V., 2001. Neurophysiological mechanisms underlying the understanding and imitation of action. *Nat Rev Neurosci* 2, 661-670.
- Rizzolatti, G., Gentilucci, M., Fogassi, L., Luppino, G., Matelli, M., Ponzoni-Maggi, S., 1987. Neurons related to goal-directed motor acts in inferior area 6 of the macaque monkey. *Exp Brain Res* 67, 220-224.
- Rizzolatti, G., Matelli, M., 2003. Two different streams form the dorsal visual system: anatomy and functions. *Exp Brain Res* 153, 146-157.
- Rizzolatti, G., Sinigaglia, C., 2010. The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nat Rev Neurosci* 11, 264-274.
- Ruby, P., Decety, J., 2001. Effect of subjective perspective taking during simulation of action: a PET investigation of agency. *Nat Neurosci* 4, 546-550.

-
- Ruby, P., Decety, J., 2003. What you believe versus what you think they believe: a neuroimaging study of conceptual perspective-taking. *Eur J Neurosci* 17, 2475-2480.
- Saxe, R., 2005. Against simulation: the argument from error. *Trends Cogn Sci* 9, 174-179.
- Saxe, R., Carey, S., Kanwisher, N., 2004. Understanding other minds: linking developmental psychology and functional neuroimaging. *Annu Rev Psychol* 55, 87-124.
- Saxe, R., Kanwisher, N., 2003. People thinking about thinking people. The role of the temporo-parietal junction in "theory of mind". *Neuroimage* 19, 1835-1842.
- Saxe, R., Powell, L.J., 2006. It's the thought that counts: specific brain regions for one component of theory of mind. *Psychol Sci* 17, 692-699.
- Schmahmann, J.D., Pandya, D.N., 2006. *Fiber Pathways of the brain*. Oxford University Press, New York.
- Schubotz, R.I., 2004. Human premotor cortex: beyond motor performance. In *MPI Series in Human Cognitive and Brain Sciences (Vol. 50)*, Max Planck Institute of Human Cognitive and Brain Sciences.
- Schubotz, R.I., 2007. Prediction of external events with our motor system: towards a new framework. *Trends Cogn Sci* 11, 211-218.
- Schubotz, R.I., 2011. Long-term planning and prediction: visiting a construction site in the human brain. In *Welsch, W., Singer, W., Wunder, A. (Eds.), Interdisciplinary Anthropology: Continuing Evolution of Man*. Springer Berlin Heidelberg.
- Schubotz, R.I., Fiebach, C.J., 2006. Integrative Models of Broca's Area and the Ventral Premotor Cortex. *Cortex* 42, 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, 638-645.
- Schubotz, R.I., von Cramon, D.Y., 2002. Predicting perceptual events activates corresponding motor schemes in lateral premotor cortex: an fMRI study. *Neuroimage* 15, 787-796.
- Schubotz, R.I., von Cramon, D.Y., 2003. Functional-anatomical concepts of human premotor cortex: evidence from fMRI and PET studies. *Neuroimage* 20 Suppl 1, S120-131.
- Schubotz, R.I., von Cramon, D.Y., 2004. Sequences of abstract nonbiological stimuli share ventral premotor cortex with action observation and imagery. *J Neurosci* 24, 5467-5474.
- Schubotz, R.I., von Cramon, D.Y., 2009. The case of pretense: observing actions and inferring goals. *J Cogn Neurosci* 21, 642-653.

-
- Searle, J.R., 1983. *Intentionality. An essay in the philosophy of mind.* Cambridge university Press, Cambridge.
- Shin, Y.K., Proctor, R.W., Capaldi, E.J., 2010. A review of contemporary ideomotor theory. *Psychol Bull* 136, 943-974.
- Shmuelof, L., Zohary, E., 2008. Mirror-image representation of action in the anterior parietal cortex. *Nat Neurosci* 11, 1267-1269.
- Sperry, R.W., 1950. Neural basis of the spontaneous optokinetic response produced by visual inversion. *J Comp Physiol Psychol* 43, 482-489.
- Spunt, R.P., Satpute, A.B., Lieberman, M.D., 2010. Identifying the what, why, and how of an observed action: an fMRI study of mentalizing and mechanizing during action observation. *J Cogn Neurosci* 23, 63-74.
- Stock, A., Stock, C., 2004. A short history of ideo-motor action. *Psychol Res* 68, 176-188.
- Strafella, A.P., Paus, T., 2000. Modulation of cortical excitability during action observation: a transcranial magnetic stimulation study. *Neuroreport* 11, 2289-2292.
- Thompson-Schill, S.L., D'Esposito, M., Kan, I.P., 1999. Effects of repetition and competition on activity in left prefrontal cortex during word generation. *Neuron* 23, 513-522.
- Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. *Nature* 381, 520-522.
- Todorov, A., Gobbini, M.I., Evans, K.K., Haxby, J.V., 2007. Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia* 45, 163-173.
- Trinkler, I., King, J.A., Doeller, C.F., Rugg, M.D., Burgess, N., 2009. Neural bases of autobiographical support for episodic recollection of faces. *Hippocampus* 19, 718-730.
- Uithol, S., van Rooij, I., Bekkering, H., Haselager, P., 2011. Understanding motor resonance. *Soc Neurosci*, 1-10.
- 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., 2009. Social cognition and the brain: a meta-analysis. *Hum Brain Mapp* 30, 829-858.
- Van Overwalle, F., Baetens, K., 2009b. 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, 495-498.

-
- Verdon, V., Schwartz, S., Lovblad, K.O., Hauert, C.A., Vuilleumier, P., 2010. Neuroanatomy of hemispatial neglect and its functional components: a study using voxel-based lesion-symptom mapping. *Brain* 133, 880-894.
- Vogeley, K., Fink, G.R., 2003. Neural correlates of the first-person-perspective. *Trends Cogn Sci* 7, 38-42.
- Wiggett, A.J., Downing, P.E., 2011. Representation of action in occipito-temporal cortex. *J Cogn Neurosci* 23, 1765-1780.
- Wohlschläger, A., Bekkering, H., 2002. Is human imitation based on a mirror-neurone system? Some behavioural evidence. *Exp Brain Res* 143, 335-341.
- Wolpert, D.M., 1997. Computational approaches to motor control. *Trends Cogn Sci* 1, 209-216.
- Wolpert, D.M., Doya, K., Kawato, M., 2003. A unifying computational framework for motor control and social interaction. *Philos Trans R Soc Lond B Biol Sci* 358, 593-602.
- Wolpert, D.M., Ghahramani, Z., 2000. Computational principles of movement neuroscience. *Nat Neurosci* 3 Suppl, 1212-1217.
- Wolpert, D.M., Ghahramani, Z., Jordan, M.I., 1995. An internal model for sensorimotor integration. *Science* 269, 1880-1882.
- Wood, J.N., Hauser, M.D., 2008. Action comprehension in non-human primates: motor simulation or inferential reasoning? *Trends Cogn Sci* 12, 461-465.
- Wurm, M.F., Hrkac, M., Morikawa, Y., Schubotz, R.I., in preparation-a.
- Wurm, M.F., Schubotz, R.I., 2011. Squeezing lemons in the bathroom: contextual information modulates action recognition. *Neuroimage* doi: 10.1016/j.neuroimage.2011.08.038 [Epub ahead of print].
- Wurm, M.F., von Cramon, D.Y., Schubotz, R.I., 2011. Do we mind other minds when we mind other minds' actions? an fMRI study. *Human Brain Mapping* doi: 10.1002/hbm.21176. [Epub ahead of print].
- Wurm, M.F., von Cramon, D.Y., Schubotz, R.I., in preparation-b. The context-object-manipulation (COM) triad: cross talk during action perception revealed by fMRI.
- Wyk, B.C., Hudac, C.M., Carter, E.J., Sobel, D.M., Pelphrey, K.A., 2009. Action understanding in the superior temporal sulcus region. *Psychol Sci* 20, 771-777.
- Young, L., Camprodon, J.A., Hauser, M., Pascual-Leone, A., Saxe, R., 2010. Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role of beliefs in moral judgments. *Proc Natl Acad Sci U S A* 107, 6753-6758.

Zacks, J.M., Tversky, B., 2001. Event structure in perception and conception. *Psychol Bull* 127, 3-21.

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

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

5 Appendix

5.1 List of abbreviations

1pp	first person perspective
3pp	third person perspective
ACC	anterior cingulate cortex
AG	angular gyrus
aIPS	anterior intraparietal sulcus
ALE	activation likelihood estimation
ANOVA	analysis of variance
BA	Brodmann area
BOLD	Blood oxygenation level-dependent
CM	context-manipulation
CO	context-object
EBA	extrastriate body area
EPI	echo planar imaging
ER	error rate
FBA	fusiform body area
FFA	fusiform face area
fMRI	function magnetic resonance imaging
FWHM	full width half maximum
FOV	field of view
GLM	general linear model
Hz	Hertz
HRF	hemodynamic response function
hMT	human motion area

IFG	inferior frontal gyrus
IFS	inferior frontal sulcus
IPL	inferior parietal lobe
L	left
LOC	lateral occipital complex
M1	primary motor cortex
MDEFT	modified driven equilibrium Fourier transform
MTG-TA	middle temporal gyrus tool area
mPFC	medial prefrontal cortex
ms	millisecond
OTC	occipitotemporal cortex
PCC	posterior cingulate cortex
PET	positron emission tomography
PMd	dorsal premotor cortex
pMTG	posterior middle temporal gyrus
PMv	ventral premotor cortex
PPA	parahippocampal place area
pSTS	posterior superior temporal gyrus
R	right
ROI	region of interest
RSC	retrosplenial cortex
RT	reaction time
s	seconds
SFS	superior frontal sulcus
SI	primary somatosensory cortex
SEM	standard error of mean
SMA	supplementary motor area
SMG	supramarginal gyrus
SPL	superior parietal lobe
T	Tesla
TE	time echo

TMS	transcranial magnetic stimulation
ToM	Theory of Mind
TOS	transverse occipital sulcus
TR	time of repetition
TPJ	temporoparietal junction
V1	primary visual cortex
V2-V6	extrastriate visual cortex V2-V6
vIPFC	ventrolateral prefrontal cortex
vmPFC	ventromedial prefrontal cortex

5.2 Figures

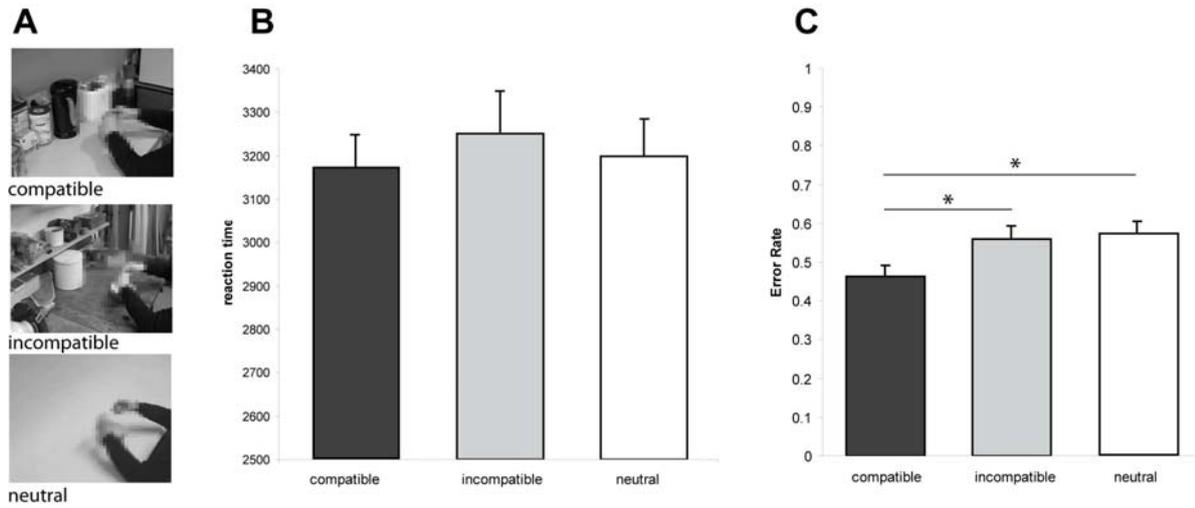


Figure A1: Behavioral follow-up experiment of Exp. 2a. Design was equivalent to Exp. 2a, but videos were black and white and actions were pixelated. Example frames of stimuli (A), reaction times (B), and error rates (C). Error bars indicate SEM.

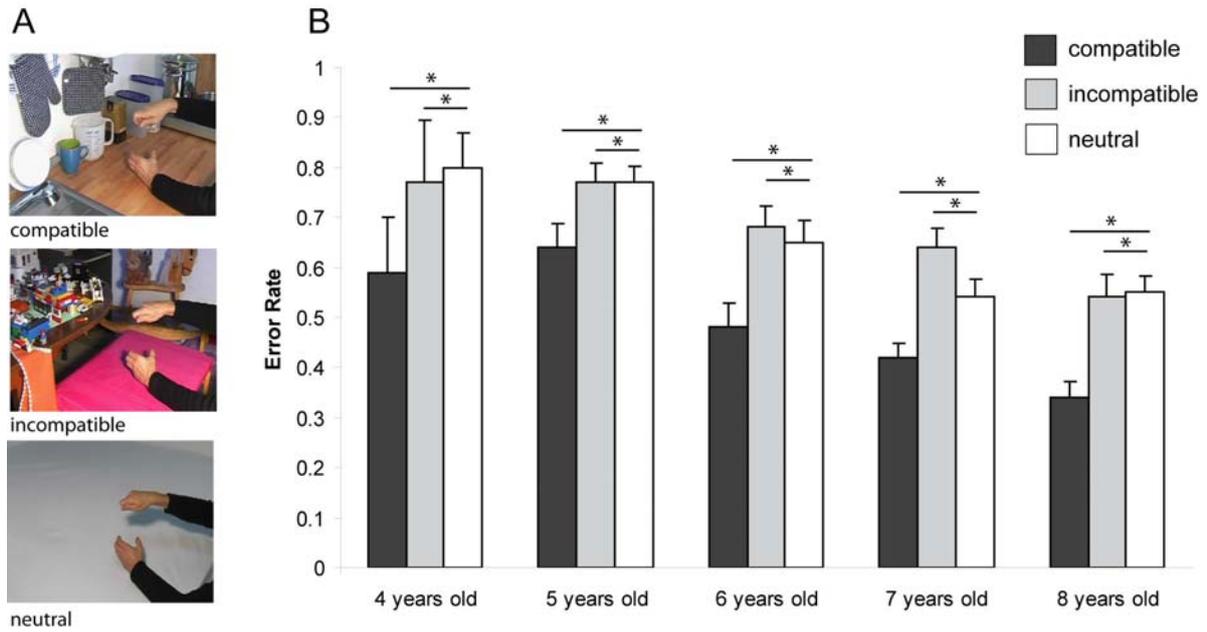


Figure A2: *Developmental study. Children aged four to eight years were presented with pantomime actions involving no objects. 10 kitchen actions and 10 children's room actions were filmed in compatible, incompatible or neutral contexts (A). Children were asked to guess the action that was pantomimed in the videos (B). Error bars indicate SEM.*

6 Reprints

- (I) Wurm, M.F., von Cramon, D.Y., Schubotz, R.I. (2011). Do we mind other minds when we mind other minds' actions? an fMRI study. *Human Brain Mapping*, doi: 10.1002/hbm.21176. [Epub ahead of print].
- (II) Wurm, M.F., Schubotz, R.I. (2011). Squeezing lemons in the bathroom: contextual information modulates action recognition. *Neuroimage*, doi: 10.1016/j.neuroimage.-2011.08.038 [Epub ahead of print].
- (III) Wurm, M.F., von Cramon, D.Y., Schubotz, R.I. (in preparation). The context-object-manipulation (COM) triad: cross talk during action perception revealed by fMRI.

Do We Mind Other Minds When We Mind Other Minds' Actions? A Functional Magnetic Resonance Imaging Study

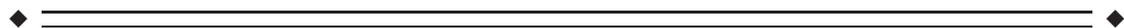
Moritz F. Wurm*, D. Yves von Cramon, and Ricarda I. Schubotz

Max Planck Institute for Neurological Research, Cologne, Germany



Abstract: Action observation engages higher motor areas, possibly reflecting an internal simulation. However, actions considered odd or unusual were found to trigger additional activity in the so-called theory of mind (ToM) network, pointing to deliberations on the actor's mental states. In this functional magnetic resonance imaging study, the hypothesis was tested that an allocentric perspective on a normal action, and even more so the sight of the actor's face, suffices to evoke ToM activity. Subjects observed short videos of object manipulation filmed from either the egocentric or the allocentric perspective, the latter including the actor's face in half of the trials. On the basis of a regions of interest analysis using ToM coordinates, we found increased neural activity in several regions of the ToM network. First, perceiving actions from an allocentric compared with the egocentric perspective enhanced activity in the left temporoparietal junction (TPJ). Second, the presence of the actor's face enhanced activation in the TPJ bilaterally, the medial prefrontal cortex (mPFC) and posterior cingulate cortex (PCC). Finally, the mPFC and PCC showed increased responses when the actor changed with respect to the preceding trial. These findings were further corroborated by zmap findings for the latter two contrasts. Together, findings indicate that observation of normal everyday actions can engage ToM areas and that an allocentric perspective, seeing the actor's face and seeing a face switch, are effective triggers. *Hum Brain Mapp* 00:000–000, 2011. © 2011 Wiley-Liss, Inc.

Key words: action observation; object manipulation; perspective; egocentric; allocentric; theory of mind; face perception; motor system



INTRODUCTION

Action observation usually engages the higher motor system, including particularly premotor and parietal areas. These are often referred to as “mirror neuron system”, in reference to macaque mirror neurons [Rizzolatti and Craighero, 2004]. In contrast, unexpected or contextually inconsistent actions were also reported to activate the so-called “Theory

of Mind (ToM) network”, suggesting considerations about the actor's mental states [Brass et al., 2007; German et al., 2004]. However, actions are always performed by persons to whom we also spontaneously attribute mental states. Hence, the exact conditions for the recruitment of the ToM network during action observation remain unclear.

Among 151 studies on action observation [Van Overwalle and Baetens, 2009], only 13 report ToM activity. All of these 13 studies show videos of unusual or implausible actions, but many of them also use an allocentric perspective, i.e., actions are shown from the third person perspective (3pp). Importantly, 3pp implies that the action is caused by another agent. In contrast to the first person perspective (1pp), the 3pp may be particularly conducive to this impression as it first entails that right and left hand are flipped by a 180° rotation to the opposite perspective, and second, that we usually see the actor's face. One therefore may hypothesize that a 3pp suffices to engage ToM during observation even of normal action.

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Moritz F. Wurm, Max Planck Institute for Neurological Research, Motor Cognition Group, Gleueler Street 50, 50931 Cologne, Germany. E-mail: mwurm@nf.mpg.de

Received for publication 4 January 2010; Revised 18 June 2010; Accepted 30 August 2010

DOI: 10.1002/hbm.21176

View this article online at wileyonlinelibrary.com.

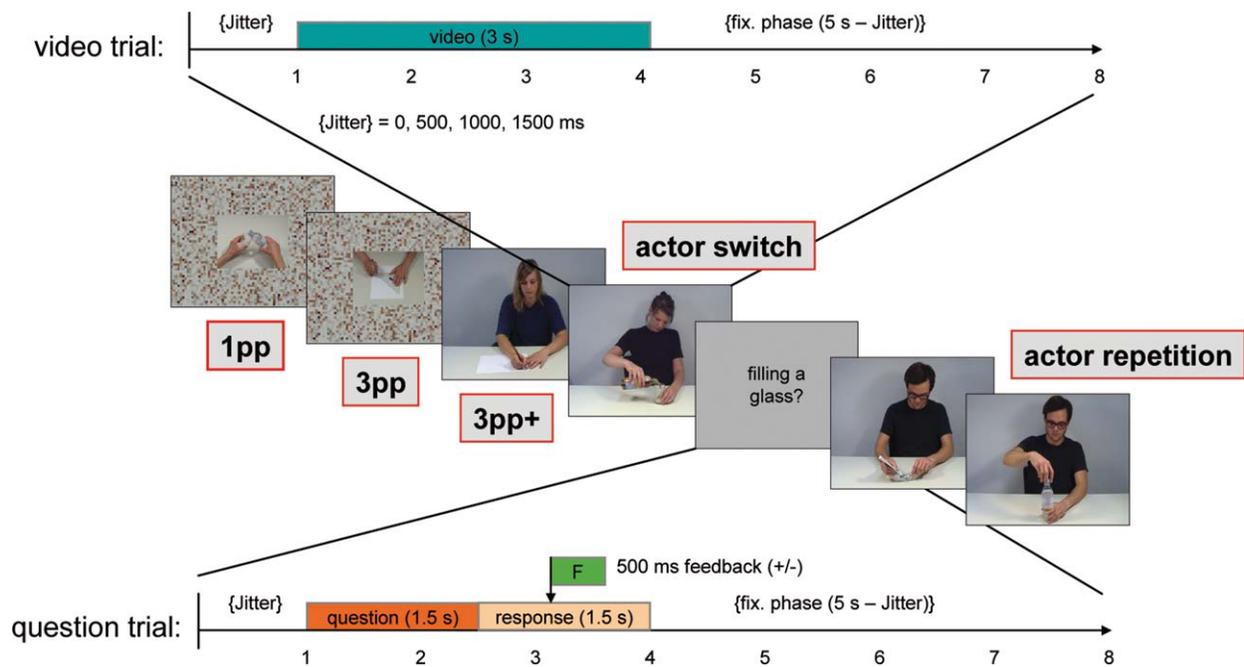


Figure 1.

Stimuli and experimental design. Video trials and question trials were interleaved in an event-related design. Video trials were composed of 3 s long video clips of everyday life actions filmed from an egocentric perspective (1pp) and two allocentric perspectives showing either hands only (3pp) or the whole actor including the actor's face (3pp+). 3pp+ trials entered a repeti-

tion suppression analysis when they were preceded by another 3pp+ trial showing either the same actor (actor repetition) or a different actor (actor switch). 20% of video trials were followed by question trials requiring participants to confirm or reject verbal action descriptions with respect to the preceding trial.

Interestingly enough, we are ignorant about the neural effects of an observer's perspective on everyday object manipulation. So far, grasping [Shmuelof and Zohary, 2008], placing [Hesse et al., 2009], and intransitive movements [Jackson et al., 2006], as well as static pictures of body parts [Chan et al., 2004; Saxe et al., 2006a] have been investigated with respect to the influence of the observer's perspective. These studies consistently report increased activation in premotor-parietal areas of the contralateral hemisphere for egocentric compared with allocentric perspective, whereas the reverse contrasts revealed activation in either occipital areas, including the lingual gyrus [Hesse et al., 2009; Jackson et al., 2006] and/or ipsilateral premotor-parietal regions [Hesse et al., 2009; Shmuelof and Zohary, 2008]. Notably, in all of these studies, the action goal was (i) known to the observer beforehand (e.g., grasping); (ii) largely invariant across the entire study (e.g., grasping whatever object is presented); and (iii) of low complexity (e.g., only grasp, then stop without further manipulation). Although these features were suitable for the purpose of the cited studies, they do not necessarily coincide with action interpretation and recognition in everyday life, where goals are often unknown, variant, and complex.

This functional magnetic resonance imaging (fMRI) study aimed to investigate whether action observation from an allocentric perspective recruits ToM, even when mentaliz-

ing is not explicitly required. To systematically disentangle the two characteristics featuring allocentric observation, i.e., perceiving body movements from a perspective other than my own, and seeing the actor's face, we conducted two analyses. First, everyday actions filmed from an allocentric perspective (3pp) were contrasted with those filmed from an egocentric perspective (1pp) to tap effects of the perspective. Second, to investigate effects due to the sight of the actor's face, we compared actions filmed from an allocentric perspective including the actor's face (3pp+) with the 3pp actions, which showed only the actor's hands. Finally, the effect of switching actors between successive movies was analyzed. To this end, 3pp+ actions performed by an actor who was new with respect to the preceding trial were compared with those performed by the same actor again.

METHODS

Participants

Twenty-one right-handed, healthy volunteers (16 female) participated in the study (age range = 22–27 years; mean age = 24.0 years). After being informed about potential risks and being screened by a physician of the institution,

subjects gave written informed consent before the fMRI measurements. The study was performed according to the Declaration of Helsinki. Data were handled anonymously.

Stimuli and Tasks

After a training phase of nine trials (six video trials and three question trials that were not used in the analysis), subjects were presented with movies showing actions (action trials) and with written action descriptions referring to these actions (question trials, Fig. 1). Each action trial (8 s) started with a movie (3 s) followed by a fixation phase. To enhance the temporal resolution of the BOLD response, a variable jitter (0, 500, 1000, or 1500 ms) was inserted before the movie. Fifty percent of the actions was performed on appropriate objects (e.g., pouring water from a bottle into a glass, altogether 60 everyday life actions) and 50% on inappropriate objects (e.g., making the same movement with a pen and a candle [cf. Schubotz and von Cramon, 2009]). However, in this study, we limit our analysis to trials performed with appropriate objects.

Subjects were instructed to attend to the presented movies. They were informed that some of the movies would be followed by an action description that either matched or did not match the content of the preceding movie and that they were to indicate whether the description matched the movie (accept) or not (reject). When a question trial was presented, subjects immediately delivered their responses on a two-button response box, using their index finger to accept and their middle finger to reject. 50% of the action descriptions matched and 50% did not. Great care was taken that the action descriptions used in the question trials referred merely to the low-level goals of actions (e.g., filling a glass), which were directly derivable from the object manipulations. Phrasings like “preparing a drink” were avoided not to refer to the intention of the actor ensuring that the task is implicit with respect to ToM.

Three experimental conditions were implemented by showing the actor’s hands from an egocentric perspective (1pp), the actor’s hands from an allocentric perspective (3pp), and the whole actor including the actor’s face from an allocentric perspective (3pp+). In addition, the trial succession was balanced such that each of the nine possible transitions occurred an equal number of times. 3pp+ trials were further analyzed by using a switching protocol: trials that were preceded by 3pp+ trials showing a different actor (actor switch) were compared with trials that were preceded by trials showing the same actor again (actor repetition).

All in all, four actresses and four actors performed a different set of actions, equally balanced with respect to the experimental conditions. 1pp videos were filmed over the shoulder of the actor resulting in an egocentric “like-me” perspective, whereas 3pp videos were filmed from the same distance and angle, but the actor sat on the other side of the table to show exactly the same detail of the hands differing solely in perspective. 3pp+ videos were filmed face to face with the actor

using a different angle and a different camera zoom factor. Actresses and actors were instructed not to show any facial expression and to focus their gaze on the objects to be manipulated. Moreover, the viewpoint of the camera did not allow for unambiguous detection of actors’ gaze direction.

The size of the hands and manipulated objects were kept constant across conditions by scaling down 1pp and 3pp videos to the same hand size as the 3pp+ videos to avoid confounding visibility effects. In addition, resized videos were placed on a scrambled background to provide an identical amount of visual information (Fig. 1, Supporting Information).

Twenty percent of the analyzed movies (i.e., 24 of 120 actions) were followed by a question trial that had the length of a regular trial (1.5 s description, 1.5 response phase, and a 5-s fixation phase). Accordingly, 120 trials (40 1pp trials, 40 3pp trials, and 40 3pp+ trials) entered the analyses contrasting 3pp > 1pp and 3pp+ > 3pp, and 11 of the 40 3pp+ trials were used for the contrast actor switch > actor repetition. Finally, 12 empty trials (fixation baseline) were presented after each second question trial.

In a post-fMRI session survey, subjects were presented with a questionnaire to measure the subjects’ ability to recognize the actors’ faces. To this end, subjects first guessed how many different actors had occurred in the video clips and subsequently were presented with 16 pictures of faces, eight of which were faces of the actresses and actors and eight were unrelated (new) faces. Subjects rated on a scale from 1 to 6 whether or not the faces occurred in the video clips. Behavioral performance in face recognition was assessed by a modified version of the Discrimination index $P(r)$, which is the difference between hit rate and false alarm rate [Snodgrass and Corwin, 1988]. The hit rate was defined as the sum of ratings of correctly recognized faces relative to the sum of the maximal rating score of all faces shown in the videos, and the false alarm rate as the sum of ratings to falsely indicated unrelated faces relative to the sum of the maximal rating score of all unrelated faces.

MRI Data Acquisition

Imaging was performed on a 3-T Siemens (München, Germany) Trio system equipped with a standard birdcage head coil. Participants were placed on the scanner bed in supine position with their right index and middle fingers positioned on the appropriate response buttons of a response box. Form-fitting cushions were used to prevent head, arm, and hand movements. Participants were provided earplugs to attenuate scanner noise. Twenty-six axial slices (192 mm field of view; 64 × 64 pixel matrix; 4 mm thickness; 1 mm spacing; in-plane resolution of 3 × 3 mm) covering the whole brain were acquired using a single-shot gradient EPI sequence (2000 ms repetition time; 30 ms echo time; 90° flip angle; 116 kHz acquisition bandwidth) sensitive to BOLD contrast. Before functional imaging, 26 anatomical T1-weighted MDEFT images [Norris, 2000; Ugurbil et al., 1993] were acquired. In a separate

session, high-resolution whole brain images were acquired from each subject to improve the localization of activation foci using a T1-weighted 3D-segmented MDEFT sequence.

MRI Data Analysis

After motion correction using rigid-body registration to the central volume, fMRI data were processed using the software package LIPSIA [Lohmann et al., 2001]. To correct for the temporal offset between the slices acquired in one image, a cubic-spline interpolation was used. Low-frequency signal changes and baseline drifts were removed using a temporal high-pass filter with a cutoff frequency of 1/99 Hz. Spatial smoothing was performed with a Gaussian filter of 5.65 mm FWHM. To align the functional data slices with a 3D stereotactic coordinate reference system, a rigid linear registration with six degrees of freedom (three rotational, three translational) was performed. The rotational and translational parameters were acquired on the basis of the MDEFT and the EPI-T1 slices to achieve an optimal match between these slices and the individual 3D reference dataset. The MDEFT volume dataset with 160 slices and 1-mm slice thickness was standardized to the Talairach stereotactic space [Talairach and Tournoux, 1988]. The rotational and translational parameters were subsequently transformed by linear scaling to a standard size. The resulting parameters were then used to transform the functional slices using trilinear interpolation, so that the resulting functional slices were aligned with the stereotactic coordinate system, thus generating output data with a spatial resolution of $3 \times 3 \times 3$ mm (27 mm^3). The statistical evaluation was based on a least-squares estimation using the general linear model for serially autocorrelated observations [Friston et al., 1995; Worsley and Friston, 1995]. The design matrix was generated with a gamma function, convolved with the hemodynamic response function and its first derivative. Brain activations were analyzed time-locked to onset of the movies, and the analyzed epoch comprised the full duration (3 s) of the presented movies, the duration of the null events (8 s), and the reaction time in action description trials (max. 3 s). The model equation, including the observation data, the design matrix, and the error term, was convolved with a Gaussian kernel of dispersion of 4 s FWHM to account for the temporal autocorrelation (Worsley and Friston, 1995). In the following, contrast images, i.e., beta value estimates of the raw-score differences between specified conditions, were generated for each participant. As all individual functional datasets were aligned to the same stereotactic reference space, the single-subject contrast images were entered into a second-level random effects analysis for each of the contrasts.

One-sample t tests were used for the group analyses across the contrast images of all subjects that indicated whether observed differences between conditions were significantly distinct from zero. The t values were subsequently transformed into Z scores.

To correct for false-positive results, in a first step, an initial voxel-wise z -threshold was set to $z = 2.576$ ($P = 0.005$). In a second step, the results were corrected for multiple comparisons using cluster-size and cluster-value thresholds obtained by Monte Carlo simulations at a significance level of $P = 0.05$, i.e., the reported activations are significantly activated at $P < 0.05$, corrected for multiple comparisons at the cluster level.

RESULTS

Behavioral Results

Performance was assessed by error rates and reaction times. Repeated-measures ANOVAs were performed for each of these measures with the levels 1pp, 3pp, and 3pp+.

Regarding reaction times, a main effect was found [$F_{(2,40)} = 10.293$, $P < 0.001$]. Paired samples t tests reflected that responses to the action descriptions were significantly slower for allocentric perspectives (mean \pm standard error; 3pp: 1102 ± 60 ms, 3pp+: 1062 ± 49 ms) compared with trials shown from an egocentric perspective [975 ± 48 ms; t_{20} (1pp-3pp) = -4.136 , $P < 0.001$; t_{20} (1pp-3pp+) = -2.984 , $P = 0.007$]. Regarding error rates, there was no significant effect [1pp: $7.2 \pm 2.7\%$, 3pp: $9.3 \pm 2.0\%$, 3pp+: $8.8 \pm 1.6\%$; $F_{(2,40)} = 26.578$, $P = 0.66$].

Face recognition was assessed by a postsession recognition test, i.e., subjects guessed the number of actors and actresses appearing in the experiment and performed a face recognition test by discriminating faces belonging to the actors and actresses, and faces of unfamiliar persons. On average, subjects spontaneously estimated that 4.7 ± 1 different actors and actresses were shown in the experiment. Subjects correctly recognized, on average, $60.2 \pm 4.9\%$ of the actors and actresses (hit rate = 0.60) and correctly rejected $86.1 \pm 5.2\%$ of the unrelated new faces (false alarm rate = 0.24). The average discrimination index (hits minus false alarms) was 0.46 ± 0.5 . A paired-samples t test showed that the discrimination index was significantly different from the chance level of 0 ($t_{20} = 7.561$, $P < 0.001$). These results indicate that, although not required for solving the task, subjects noticed that actions were performed by different actors and that they were able to remember them after the experiment.

FMRI Results

Perspective

To investigate whether allocentrically perceived actions recruit ToM in addition to the premotor-parietal network, we analyzed 3pp and 1pp trials. Compared with baseline, both perspectives revealed an extensive bilateral activation pattern of occipital, premotor-parietal, and temporal regions. However, the direct contrast 3pp > 1pp revealed an increase of neural activity in the lingual gyrus and ToM regions,

TABLE I. Anatomical specification, Brodmann area, hemisphere (R, right; L, left), Talairach coordinates (x , y , and z), and maximal Z scores (Z) of activations in perspective conditions ($1pp > 3pp$, and vice versa; for $1pp > 3pp$ corrected cluster threshold, $P < 0.05$)

Area	Brodmann area	Hemisphere	Talairach coordinates			Z
			x	y	z	
3pp > 1pp						
TPJ	39	L	-38	-48	24	2.95
		L	-47	-60	27	2.53
		R	49	-60	30	2.70
mPFC	32	R	13	36	0	2.59
PCC	31	R	1	-54	30	2.47
Lingual gyrus	17	R	1	-84	0	3.44
1pp > 3pp						
PMd	6	L	-28	-12	57	3.64
aIPS	7	L	-31	-42	54	5.26
SMG	40	L	-50	-27	36	4.02
pIPS	7/19	L	-22	-81	36	4.27
EBA/hMT	19/37	L	-43	-60	9	4.69
Cuneus	18	L	-10	-99	15	4.11
		R	17	-93	18	4.89
Cerebellum		R	16	-57	-36	3.75
		R	35	-39	-24	3.67

including posterior cingulate cortex (PCC), medial prefrontal cortex (mPFC, BA 32), and bilateral temporoparietal junction (TPJ). Because these activations did not survive the correction for multiple comparisons, we further analyzed activation sites via regions of interest (ROI) analyses using averaged coordinates of ToM belief studies (see below).

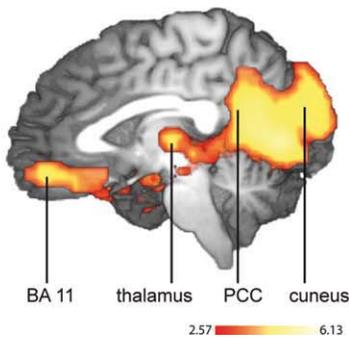
Regarding the opposite contrast, $1pp > 3pp$ revealed enhanced activity in the left dorsal premotor cortex (PMd), left supramarginal gyrus (SMG), and adjacent intraparietal sulcus (IPS), a left temporo-occipital region, bilateral

cuneus, and right cerebellum. According to the literature (Downing et al., 2007), we suggest that activation in the temporo-occipital region comprised the (hardly separable) extrastriate body area (EBA) and human motion area (hMT; Table I).

Sight of the Actor

Allocentric observation of actions, including faces compared with those restricted to the hands ($3pp+ > 3pp$),

A 3pp+ > 3pp



B actor switch > actor repetition

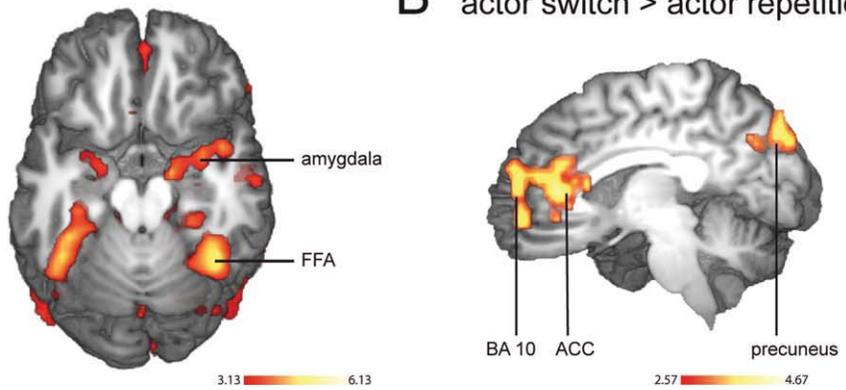


Figure 2.

Effects of the actor's face and their switches. **A**, areas activated for actions showing the upper body including the face ($3pp+$) compared with allocentrically perceived hands ($3pp$, corrected cluster threshold $P < 0.05$). For the axial view, a higher threshold ($z = 3.13$) was chosen to accentuate activation peaks. **B**, areas activated for actor switch compared with actor repetition trials (corrected cluster threshold $P < 0.05$).

TABLE II. Anatomical area, Brodmann area, hemisphere, Talairach coordinates (x, y, and z), and maximal Z scores of significant activations in face conditions (3pp+ > 3pp, actor switch > actor repetition; corrected cluster threshold $P < 0.05$)

Area	Brodmann area	Hemisphere	Talairach coordinates			Z
			x	y	z	
3pp+ > 3pp						
mPFC	11	R	2	48	-12	3.72
PCC	31	R	1	-56	33	3.33
Temporal pole	38	R	40	15	-21	4.10
aSTS	22	R	47	-12	-6	4.33
pSTS	39	L	-49	-66	18	5.24
		R	40	-63	18	5.64
IFG	45	R	47	30	3	4.96
FFA	37	L	-37	-48	-12	4.80
		R	34	-48	-9	5.44
Cuneus	31	L	-7	-69	6	6.13
Amygdala		L	-19	-6	-15	3.64
		R	14	-9	-12	4.46
Thalamus, Pulvinar		L	-19	-27	3	5.22
		R	20	-27	0	4.97
Actor switch > actor repetition						
Precuneus	7		2	-78	39	4.67
ACC	33		8	33	12	4.36
mPFC	10		-5	54	-3	3.33
			5	51	18	3.50
Actor repetition > actor switch						
pIPS	7	L	-28	-54	51	4.43
		R	14	-57	60	4.80
aIPS	7	L	-53	-33	48	4.63
		R	50	-30	48	4.61
Postcentral gyrus	40	L	-58	-21	27	3.21
EBA/hMT	19	L	-43	-66	-3	4.07
		R	44	-51	-9	4.25

yielded extended bilateral activation in the fusiform and parahippocampal gyrus, inferior temporal gyrus, right anterior and posterior superior temporal sulcus (STS), right temporal pole, EBA/hMT, bilateral amygdala, pulvinar nucleus of thalamus (PLV), mPFC (BA 11), cuneus extending into PCC and retrosplenial cortex, and right inferior frontal gyrus (IFG, BA 45) (Fig. 2A; Table II). Maximum activation within the region of the fusiform gyrus amounted to the coordinates reported for the fusiform face area (FFA) [Kanwisher et al., 1997; Spiridon et al., 2006].

Actor Switch versus Actor Repetition

Observing a different actor than in the preceding trial (actor switch) yielded enhanced activation in medial frontal areas (BA 10 and pregenual as well as subgenual ACC) and posterior precuneus. In contrast, repetition of the same actor increased activity in bilateral inferior postcentral gyrus and anterior IPS as well as in EBA/hMT (Fig. 2B; Table II).

ROI Analysis

To discuss the differential activation of the ToM network, we averaged the coordinates of 14 studies reported in Van Overwalle and Baetens [2009] using ToM belief tasks for the definition of ToM-ROIs [Abraham et al., 2008; Ferstl and von Cramon, 2002; Gallagher et al., 2000; Gobbin et al., 2007; Hynes et al., 2006; Kobayashi et al., 2007; Mitchell, 2008; Perner et al., 2006; Saxe and Kanwisher, 2003; Saxe and Powell, 2006; Saxe et al., 2006b; Sommer et al., 2007; Vogeley et al., 2001; Wakusawa et al., 2007]: left TPJ: -51, -59, 26; right TPJ: 54, -49, 22; PCC: -1, -55, 33; mPFC: -3, 50, 20. Mean beta values were extracted from the averaged coordinate voxel plus six adjacent voxels of the following conditions versus rest: 1pp, 3pp, 3pp+, actor switch, and actor repetition. Mean beta values are shown in Figure 3. With regard to the effect of perspective (3pp > 1pp), the activation of the left TPJ was significantly higher in 3pp-R compared with 1pp-R ($F_{(1,20)} = -2.072$, $P = 0.026$; paired samples t test). The presence of the actor's face (3pp+ > 3pp) revealed significant

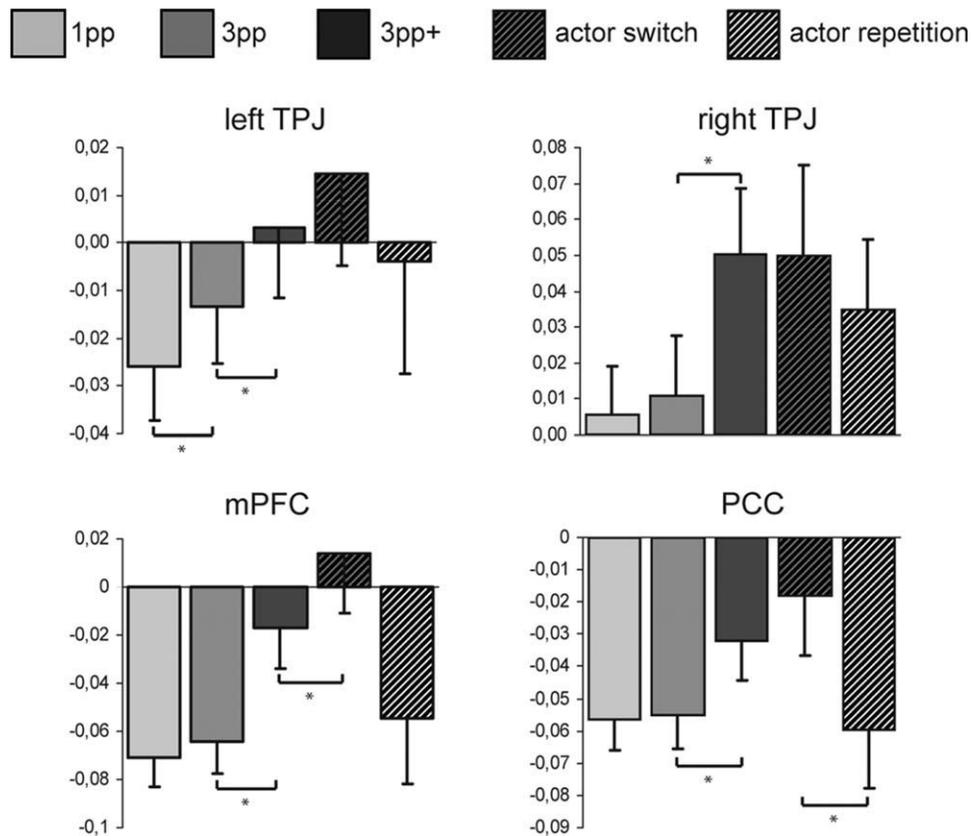


Figure 3.

ROI analysis in ToM regions defined by averaging coordinates of 14 ToM belief studies reported in Van Overwalle and Baetens (2009). Mean beta values were extracted from the contrasts 1pp > rest, 3pp > rest, 3pp+ > rest, actor switch > rest, and actor repetition > rest. Error bars indicate standard error of mean.

enhancement in all ToM regions (left TPJ [$F_{(1,20)} = -2.046$, $P = 0.028$], right TPJ [$F_{(1,20)} = -5.567$, $P < 0.001$], PCC [$F_{(1,20)} = -4.023$, $P < 0.001$], mPFC [$F_{(1,20)} = -2.227$, $P = 0.019$]; paired-samples t tests). Finally, the activation of ToM regions was enhanced for actor switches compared with actor repetitions, yielding significant effects in the PCC ($F_{(1,20)} = 2.446$, $P = 0.012$; paired-samples t test), and the mPFC ($F_{(1,20)} = 2.286$, $P = 0.017$; paired samples t test).

DISCUSSION

This fMRI study investigated whether perspective on the action and the visibility on the actor's face modulates an observer's mentalizing (ToM) network. An independent ROI analysis using coordinates extracted from classic ToM task studies yielded significant effects (i) in the left TPJ for 3pp > 1pp; (ii) in the TPJ bilaterally, the PCC and the mPFC for 3pp+ > 3pp; and (iii) in the mPFC and PCC for actor switch > actor repetition. Together, findings indicate that parts of the ToM network are differentially enhanced

during observation of normal everyday actions (a) when we see the action from the third person perspective, (b) when we can see the actor's face, and (c) when we see a new actor. These findings uncover the ToM network as being intimately involved in the perceptual analysis of ordinary action.

Observing Actions From the Allocentric Perspective (3pp)

Actions of others are typically perceived from an allocentric but rarely from an egocentric perspective, which is usually associated with own actions. Following our assumption that action observation uses not only the motor system but also draws on the ToM network, e.g., by considering the mental states of the actor, ToM regions were expected to be stronger activated for the allocentric perspective. Indeed, we found activations in regions associated with ToM for 3pp compared with 1pp, including the TPJ bilaterally, the mPFC, and the PCC. However,

these activations fell below a conservative statistical threshold, which may be indicative of spontaneously triggered mentalizing that is unconstrained and not explicitly required by the task. A ROI analysis using the mean of peak coordinates obtained by classic ToM tasks revealed a significant effect in the left TPJ (Fig. 3), thus we will limit our discussion of ToM effects to this region.

Among the ToM regions, TPJ is suggested to play a particular role in perspective taking [Ruby and Decety, 2001, 2004; Vogeley et al., 2004] and is considered to reflect that we mentally put ourselves in someone else's shoes [Abraham et al., 2008]. Recent findings suggest a role of the TPJ in control of shared motor representations to keep apart self- and other-caused actions [Brass et al., 2009]. TPJ was also reported in the context of visuospatial reorienting [Corbetta and Shulman, 2002]. The linkage between mentalizing and attentional reorienting implied by TPJ as the common node of the related networks is still puzzling, but both functions seem reconcilable [Corbetta et al., 2008]. Along these lines, TPJ activation in this study may reflect a visuospatial transformation during perspective taking to match the actor's spatial orientation.

Interestingly, activation of the TPJ was not found in other studies using 1pp and 3pp conditions in their experimental designs. Their foci of interest were the effect of perspective on static hands and body parts [Chan et al., 2004; Saxe et al., 2006a], and modulation of the premotor-parietal network during observation of placing [Hesse et al., 2009], grasping [Shmuelof and Zohary, 2008], and intransitive movements [Jackson et al., 2006]. All studies report a common overlap of activation in contralateral somatosensory and motor areas for 1pp, which is also consistent with our results. Regarding 3pp, increased activation was found in occipital regions [cuneus, lateral occipital [Chan et al., 2004], lingual gyrus [Hesse et al., 2009]] and/or in ipsilateral motor areas (right superior parietal lobe [Hesse et al., 2009; Shmuelof and Zohary, 2008], right precentral gyrus, and right EBA [Hesse et al., 2009]). Besides activation of the lingual gyrus, which is also activated in our study and might be explained by increased effort with regard to the analysis of the visual input, the differential activation of the motor system on the one hand, and the TPJ on the other, can be plausibly explained by two crucial aspects. First, grasping and placing can be performed with both hands equally well. Thus, corresponding motor representations are probably not lateralized to one hemisphere. Second, actions directed to a target object in space (often referred to as goal-directed actions [Bekkering et al., 2000]) are suggested to be remapped to the effector that would most efficiently replicate the action toward the relevant object so that a transformation into an observer-congruent reference frame is not required [cf. Shmuelof and Zohary, 2008]. In accordance, goal-directed actions observed from an allocentric perspective tend to be imitated in a mirrored fashion [Bekkering et al., 2000; Wohlschlagel et al., 2003], in line with the right-hemispheric activation found for grasping

and placing observation from 3pp. In contrast, bimanual object manipulation, which was used in this study, explicitly reveals the actor's handedness. Thus, actually taking the perspective of the actor may be a more suitable strategy for the analysis of bimanual object manipulation.

Observing and Recalling Actors

In a next step, we tested whether the opportunity to perceive not only the action itself but also the actor's face is a potential trigger for ToM activity as it provides person information. Contrasting 3pp+ with 3pp yielded extended activations in occipital and temporal areas as well as in the mPFC (BA 11) and in the right IFG (BA 45). The ROI analysis revealed significant enhancement in bilateral TPJ, mPFC (BA 10), and PCC (Fig. 3).

Other occipital and temporal activations were simply due to the sight of additional body parts (EBA, [Taylor et al., 2007]) and faces (pSTS [Allison et al., 2000; Puce and Perrett, 2003] and FFA [Kanwisher et al., 1997; Spiridon et al., 2006]). Coactivation of FFA with the amygdala and the PLV reflected components of the amygdalo-fusiform pathway [Smith et al., 2009] and suggests that our subjects dealt with the actors' faces, although this was not explicitly required by the task [Kouider et al., 2009; Pasley et al., 2004]. The right lateralized activation of the IFG, temporal pole, and aSTS supports this notion because these areas are suggested to belong to the so-called "extended" system for face perception [Barbeau et al., 2008; Ishai et al., 2005]. Moreover, temporal pole and aSTS were also found to play a role in ToM [Gallagher and Frith 2003], as well as mPFC and PCC. Interestingly, several studies on face recognition report activation of the precuneus, PCC, and mPFC when comparing recognition of familiar faces with recognition of unfamiliar faces [Gobbini et al., 2004; Leibenluft et al., 2004; Trinkler et al., 2009]. The precuneus and PCC are related to retrieval from long-term memory, whereas the mPFC was suggested for encoding information about personality traits of a familiar individual [Gobbini et al., 2004; Leibenluft et al., 2004]. It is possible that activation in the PCC was due to acquisition of visual familiarity over the course of the experiment (Gobbini and Haxby, 2006; Kosaka et al., 2003; Trinkler et al., 2009). Our post-session survey supports the assumption that the subjects acquired visual familiarity with the actors. Similarly, the mPFC may reflect the attempt of trait inference, as was reported for familiarity with faces [Gobbini et al., 2004; Trinkler et al., 2009]. The association of information about behaviour with faces was found to increase activity in the mPFC and pSTS [Todorov, 2007].

These findings and their interpretation were further substantiated by the analysis of trials with switched actors compared with trials with actors that were already presented in the preceding trial ($n-1$). Perceiving a new actor yielded enhanced activity in the mPFC, more precisely BA 10 and the ACC, as well as in the posterior precuneus.

Using ToM ROI coordinates of these regions yielded significant effects as well (Fig. 3).

Trinkler et al. [2009] reported activation of the retrosplenial cortex for the acquisition of visual familiarity and of the precuneus for the retrieval of personal knowledge. Thus, switching to a foreseen actor may trigger a recall of the visual image of the actor and episodic memory about the actions associated with this particular actor (e.g., this actor has squeezed an orange and sharpened a pencil before). Recalling characteristics of an actor could also include the retrieval of idiosyncrasies of movements or facial expressions, e.g., a person looks tired, ambivalent, or nervous, contributing to the formation of an actor-related knowledge, as reflected by activation of the mPFC. Moreover, the mPFC is often associated with the attempt to understand the reasons for a particular action [Van Overwalle, 2009]. In our study, actions were selected so that they did not imply some kind of long-term goal beyond or across several trials. However, coming up with a potential coherent global goal achieved by several single actions may be a spontaneous tendency during action observation. In contrast, the recollection of knowledge from remote trials showing the same actor is not required in actor repetitions and would be attenuated in comparison to actor switch trials.

CONCLUSIONS

Present findings suggest that perceiving actions from an allocentric perspective evokes mental inferences even when not required by the task. The particular selection of activated ToM areas reflected that the observers engaged in perspective taking as well as in the formation and retrieval of actor-related familiarity and knowledge. The latter aspect is especially elicited by both augmenting the “hands manipulating objects” scenario with the face of the actor, as well as by switching actors with regard to the preceding trial. Results indicate that the ToM network is intimately involved in the perception of ordinary actions.

REFERENCES

Abraham A, Werning M, Rakoczy H, von Cramon DY, Schubotz RI (2008): Minds, persons, and space: An fMRI investigation into the relational complexity of higher-order intentionality. *Conscious Cogn* 17:438–450.

Allison T, Puce A, McCarthy G (2000): Social perception from visual cues: Role of the STS region. *Trends Cogn Sci* 4:267–278.

Barbeau EJ, Taylor MJ, Regis J, Marquis P, Chauvel P, Liegeois-Chauvel C (2008): Spatiotemporal dynamics of face recognition. *Cereb Cortex* 18:997–1009.

Bekkering H, Wohlschlagel A, Gattis M (2000): Imitation of gestures in children is goal-directed. *Q J Exp Psychol A* 53:153–64.

Brass M, Ruby P, Spengler S (2009): Inhibition of imitative behaviour and social cognition. *Philos Trans R Soc Lond B Biol Sci* 364:2359–2367.

Brass M, Schmitt RM, Spengler S, Gergely G (2007): Investigating action understanding: Inferential processes versus action simulation. *Curr Biol* 17:2117–2121.

Chan AW, Peelen MV, Downing PE (2004): The effect of viewpoint on body representation in the extrastriate body area. *Neuroreport* 15:2407–2410.

Corbetta M, Patel G, Shulman GL (2008): The reorienting system of the human brain: From environment to theory of mind. *Neuron* 58:306–324.

Corbetta M, Shulman GL (2002): Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201–215.

Downing PE, Wiggett AJ, Peelen MV (2007): Functional magnetic resonance imaging investigation of overlapping lateral occipitotemporal activations using multi-voxel pattern analysis. *J Neurosci* 27:226–233.

Ferstl EC, von Cramon DY (2002): What does the frontomedian cortex contribute to language processing: Coherence or theory of mind? *Neuroimage* 17:1599–1612.

Gallagher HL, Frith CD (2003): Functional imaging of ‘theory of mind’. *Trends Cogn Sci* 7:77–83.

Gallagher HL, Happe F, Brunswick N, Fletcher PC, Frith U, Frith CD (2000): Reading the mind in cartoons and stories: An fMRI study of ‘theory of mind’ in verbal and nonverbal tasks. *Neuropsychologia* 38:11–21.

German TP, Niehaus JL, Roarty MP, Giesbrecht B, Miller MB (2004): Neural correlates of detecting pretense: Automatic engagement of the intentional stance under covert conditions. *J Cogn Neurosci* 16:1805–1817.

Gobbini MI, Haxby JV (2006): Neural response to the visual familiarity of faces. *Brain Res Bull* 71:76–82.

Gobbini MI, Koralek AC, Bryan RE, Montgomery KJ, Haxby JV (2007): Two takes on the social brain: A comparison of theory of mind tasks. *J Cogn Neurosci* 19:1803–1814.

Gobbini MI, Leibenluft E, Santiago N, Haxby JV (2004): Social and emotional attachment in the neural representation of faces. *Neuroimage* 22:1628–1635.

Friston KJ, Holmes AP, Worsley KJ, Poline JB, Frith CD, Frackowiak RS (1995): Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping* 2:189–210.

Hesse MD, Sparing R, Fink GR (2009): End or means—The “what” and “how” of observed intentional actions. *J Cogn Neurosci* 21:776–790.

Hynes CA, Baird AA, Grafton ST (2006): Differential role of the orbital frontal lobe in emotional versus cognitive perspective-taking. *Neuropsychologia* 44:374–383.

Ishai A, Schmidt CF, Boesiger P (2005): Face perception is mediated by a distributed cortical network. *Brain Res Bull* 67:87–93.

Jackson PL, Meltzoff AN, Decety J (2006): Neural circuits involved in imitation and perspective-taking. *Neuroimage* 31:429–439.

Kanwisher N, McDermott J, Chun MM (1997): The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311.

Kobayashi C, Glover GH, Temple E (2007): Children’s and adults’ neural bases of verbal and nonverbal ‘theory of mind’. *Neuropsychologia* 45:1522–1532.

Kosaka H, Omori M, Iidaka T, Murata T, Shimoyama T, Okada T, Sadato N, Yonekura Y, Wada Y (2003): Neural substrates participating in acquisition of facial familiarity: An fMRI study. *Neuroimage* 20:1734–1742.

- Kouider S, Eger E, Dolan R, Henson RN (2009): Activity in face-responsive brain regions is modulated by invisible, attended faces: Evidence from masked priming. *Cereb Cortex* 19:13–23.
- Leibenluft E, Gobbini MI, Harrison T, Haxby JV (2004): Mothers' neural activation in response to pictures of their children and other children. *Biol Psychiatry* 56:225–232.
- Lohmann G, Muller K, Bosch V, Mentzel H, Hessler S, Chen L, Zysset S, von Cramon DY (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.
- Mitchell JP (2008): Activity in right temporo-parietal junction is not selective for theory-of-mind. *Cereb Cortex* 18:262–271.
- Norris DG (2000): Reduced power multislice MDEFT imaging. *J Magn Reson Imaging* 11(4):445–451.
- Pasley BN, Mayes LC, Schultz RT (2004): Subcortical discrimination of unperceived objects during binocular rivalry. *Neuron* 42:163–172.
- Perner J, Aichhorn M, Kronbichler M, Staffen W, Ladurner G (2006): Thinking of mental and other representations: The roles of left and right temporo-parietal junction. *Soc Neurosci* 1:245–258.
- Puce A, Perrett D (2003): Electrophysiology and brain imaging of biological motion. *Philos Trans R Soc Lond B Biol Sci* 358:435–445.
- Rizzolatti G, Craighero L (2004): The mirror-neuron system. *Annu Rev Neurosci* 27:169–192.
- Ruby P, Decety J (2001): Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nat Neurosci* 4:546–550.
- Ruby P, Decety J (2004): How would you feel versus how do you think she would feel? A neuroimaging study of perspective-taking with social emotions. *J Cogn Neurosci* 16:988–999.
- Saxe R, Jamal N, Powell L (2006a): My body or yours? The effect of visual perspective on cortical body representations. *Cereb Cortex* 16:178–182.
- Saxe R, Kanwisher N (2003): People thinking about thinking people. The role of the temporo-parietal junction in “theory of mind”. *Neuroimage* 19:1835–1842.
- Saxe R, Powell LJ (2006): It's the thought that counts: Specific brain regions for one component of theory of mind. *Psychol Sci* 17:692–699.
- Saxe R, Schulz LE, Jiang YV (2006b): Reading minds versus following rules: Dissociating theory of mind and executive control in the brain. *Soc Neurosci* 1:284–298.
- Schubotz RI, von Cramon DY (2009): The case of pretense: Observing actions and inferring goals. *J Cogn Neurosci* 21: 642–653
- Shmuelof L, Zohary E (2008): Mirror-image representation of action in the anterior parietal cortex. *Nat Neurosci* 11:1267–1269.
- Smith CD, Lori NF, Akbudak E, Sorar E, Gultepe E, Shimony JS, McKinstry RC, Conturo TE (2009): MRI diffusion tensor tracking of a new amygdalo-fusiform and hippocampo-fusiform pathway system in humans. *J Magn Reson Imaging* 29: 1248–1261.
- Snodgrass JG, Corwin J (1988): Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *J Exp Psychol Gen* 117:34–50.
- Sommer M, Dohnel K, Sodian B, Meinhardt J, Thoermer C, Hajak G (2007): Neural correlates of true and false belief reasoning. *Neuroimage* 35:1378–1384.
- Spiridon M, Fischl B, Kanwisher N (2006): Location and spatial profile of category-specific regions in human extrastriate cortex. *Hum Brain Mapp* 27:77–89.
- Talairach J, Tournoux P (1988): Co-planar Stereotaxic Atlas of the Human Brain. New York: Thieme.
- Taylor JC, Wiggett AJ, Downing PE (2007): Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *J Neurophysiol* 98:1626–1633.
- Todorov A, Gobbini MI, Evans KK, Haxby JV (2007): Spontaneous retrieval of affective person knowledge in face perception. *Neuropsychologia* 45(1):163–173.
- Trinkler I, King JA, Doeller CF, Rugg MD, Burgess N (2009): Neural bases of autobiographical support for episodic recollection of faces. *Hippocampus* 19:718–730.
- Van Overwalle F (2009): Social cognition and the brain: A meta-analysis. *Hum Brain Mapp* 30:829–858.
- Van Overwalle F, Baetens K (2009): Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. *Neuroimage* 48:564–584.
- Vogeley K, Bussfeld P, Newen A, Herrmann S, Happe F, Falkai P, Maier W, Shah NJ, Fink GR, Zilles K (2001): Mind reading: Neural mechanisms of theory of mind and self-perspective. *Neuroimage* 14(1 Pt 1):170–181.
- Vogeley K, May M, Ritzl A, Falkai P, Zilles K, Fink GR (2004): Neural correlates of first-person perspective as one constituent of human self-consciousness. *J Cogn Neurosci* 16:817–827.
- Ugurbil K, Garwood M, Ellermann J, Hendrich K, Hinke R, Hu X, Kim SG, Menon R, Merkle H, Ogawa S and others (1993): Imaging at high magnetic fields: initial experiences at 4 T. *Magn Reson Q* 9(4):259–277.
- Wakusawa K, Sugiura M, Sassa Y, Jeong H, Horie K, Sato S, Yokoyama H, Tsuchiya S, Inuma K, Kawashima R (2007): Comprehension of implicit meanings in social situations involving irony: A functional MRI study. *Neuroimage* 37:1417–1426.
- Wohlschlager A, Gattis M, Bekkering H (2003): Action generation and action perception in imitation: An instance of the ideomotor principle. *Philos Trans R Soc Lond B Biol Sci* 358:501–515.
- Worsley KJ, Friston KJ (1995): Analysis of fMRI time-series revisited—again. *Neuroimage* 2(3):173–181.



Squeezing lemons in the bathroom: Contextual information modulates action recognition

Moritz F. Wurm^{*}, Ricarda I. Schubotz

Max Planck Institute for Neurological Research, Cologne, Germany
Psychological Institute II, Westfälische Wilhelms-Universität, Münster, Germany

ARTICLE INFO

Article history:
Received 15 July 2011
Revised 10 August 2011
Accepted 11 August 2011
Available online xxxx

Keywords:

Action perception
Action observation
Context
Ventrolateral prefrontal cortex
Inferior frontal cortex
Parahippocampal place area

ABSTRACT

Most every day actions take place in domestic rooms that are specific for certain classes of actions. Contextual information derived from domestic settings may therefore influence the efficiency of action recognition. The present studies investigated whether action recognition is modulated by compatibility of the context an action is embedded in. To this end, subjects watched video clips of actions performed in compatible, incompatible, and neutral contexts. Recognition was significantly slower when actions took place in an incompatible as compared to a compatible or a neutral context (Experiment 1). Functional MRI revealed increased activation for incompatible context in Brodmann Areas (BA) 44, 45, and 47 of the left ventrolateral prefrontal cortex (vlPFC; Experiment 2). Results suggest that contextual information - even when task-irrelevant - informs a high processing level of action analysis. In particular, the functional profiles assigned to these prefrontal regions suggest that contextual information activates associated action representations as a function of (in-) compatibility. Thus, incompatibility effects may reflect the attempt to resolve the conflict between action and context by embedding the presented action step into an overarching action that is again compatible with the provided context.

© 2011 Published by Elsevier Inc.

Introduction

Our daily activities take place in highly specialized places and rooms that are optimized for specific actions: In kitchens we prepare food, in bathrooms we engage in body care, but usually not vice versa. Since domestic settings are reminiscent of certain classes of actions, one may expect them to modulate the recognition of observed actions as well. Compatible contextual settings should constrain expectation in a facilitatory way, whereas incompatible ones should mislead expectation and thus cause interference. Its high relevance in real life notwithstanding, the influence of contextual information derived from domestic settings on action recognition has not been explored yet.

Action recognition has been defined as matching percepts of action onto corresponding action stored in semantic memory (Jeannerod, 2006). After Hamilton and Grafton (2007) actions can be described at three broad, hierarchically organized levels: the *muscle* level describes the pattern of muscle activity required for action execution, the *kinematic* level describes the characteristic movements of involved effectors in space and time, and the *goal* level describes the overarching desired outcome of the action. Contextual information may affect the perception-memory matching level of action recognition by activating

contextually related action memories and thereby modulate the matching process via spreading activation in the semantic network (Neely, 1977). In a similar vein, contextual information may affect higher levels of action analysis, e.g. inference of (long-term) goals and prediction of subsequent motor acts.

In the present experiments, both behavioral as well as neuroimaging methods were employed to investigate whether contextual information provided by domestic settings affects action recognition, and if so, at which level of action analysis.

In a behavioral study (Experiment 1), it was tested whether compatibility of context and action affects the speed of action recognition. We then used fMRI (Experiment 2) to identify brain regions that are modulated by contextual compatibility, and therewith to elucidate which level of action analysis is affected by contextual compatibility. We expected early steps of action analysis to be reflected in the occipitotemporal cortex (Beauchamp et al., 2002; Hamilton and Grafton, 2007; Jastorff and Orban, 2009; Jastorff et al., 2009; Kable and Charterjee, 2006) whereas inference of (long-term) goals and prediction of forthcoming motor acts to draw on prefrontal cortex as well as premotor-parietal loops (Iacoboni et al., 2005; Johnson-Frey et al., 2003; Kilner et al., 2007; Schubotz, 2007; Van Overwalle and Baetens, 2009; van Schie et al., 2006). Moreover, visual processing of rooms and places involves especially the parahippocampal place area (PPA; Epstein and Kanwisher, 1998). In monkeys, the parahippocampal cortex is connected to ventrolateral prefrontal, parietal and occipitotemporal

^{*} Corresponding author at: Max Planck Institute for Neurological Research, Motor Cognition Group, Gleueler Str. 50, 50931 Cologne, Germany. Fax: +49 221 4726 298. E-mail address: mwurm@nf.mpg.de (M.F. Wurm).

cortices (Schmahmann and Pandya, 2006; Schmahmann et al., 2007), making these regions likely candidates for the interaction of contextual and action information.

In both experiments, subjects watched video clips of context-specific everyday actions performed in domestic settings that were either compatible or incompatible with the action. In addition, we employed a neutral condition where actions were performed in settings without any interior. According to our approach, a video clip was contextually “neutral” when it did not provide a clear bias as to which room it pertains, and therefore could serve as a cue for contextual affiliation. Importantly, the neutral condition provided a baseline for evaluating facilitatory effects of the compatible and interference effects of the incompatible condition: Compatible but neither incompatible nor neutral contexts should bias action-relevant information and thereby facilitate action recognition. Similarly, incompatible but neither compatible nor neutral contexts should activate action-irrelevant information that interferes with the observed action.

In the behavioral study (Experiment 1), subjects were instructed to observe the action videos and to interrupt them by button press as soon as they recognized the presented action. We expected differences in the speed of recognition between the incompatible, the compatible, and the neutral condition: Facilitatory effects would manifest in faster responses to the compatible condition as compared to the incompatible and to the neutral condition because contextual associations of the domestic setting should *pre-activate the representation of the observed action* (Bar, 2007). In contrast, interference effects would manifest in delayed responses for the incompatible condition as compared to the compatible and to the neutral condition. In this case, *action recognition is hampered by pre-activation of incompatible action information that conflict with the observed action.*

In the fMRI Experiment, we tested whether, and if so, on which level, contextual information affects the neural signatures of action perception. As elaborated above, contextual modulation of the low-/perceptual level was expected to be signified by occipitotemporal activation. Modulation of the high/goal inference level was expected to be signified by prefrontal and/or premotor activation.

To this end, a different group of subjects observed the same set of actions that were presented in the behavioral study. To ensure that subjects paid attention to the videos, they occasionally had to confirm or reject an action description with respect to the preceding trial.

Influences of contextual compatibility on action observation were assessed using conjunction analyses: facilitatory effects would be signified by decreases of neural activity in the conjunction (compatible vs. neutral) \cap (compatible vs. incompatible), given that context-triggered pre-activation of the target action results in neural adaptation (Bedny et al., 2008; Gold et al., 2006; Henson, 2003). Similarly, interference effects would be signified by increases of neural activity in the conjunction (incompatible vs. neutral) \cap (incompatible vs. compatible) due to *activation of irrelevant, potentially conflicting information* (Bedny et al., 2008; Cardillo et al., 2004). In order to determine the functional specificity of contextual (in-) compatibility for action perception, we furthermore analyzed whether putative activation sensitive to contextual compatibility overlapped with regions sensitive to action recognition. Regions involved in action perception were identified by contrasting actions in the neutral context with a resting baseline.

Moreover, incompatibility between context and action was expected to *trigger a perceptual re-analysis of the domestic settings in a top-down manner after incompatibility of context and action has been detected.* In that case, the BOLD response should increase in scene processing regions, especially the PPA. The PPA as well as other context-specific regions (i.e. retrosplenial cortex (RSC) and transverse occipital sulcus (TOS)) was identified via a separate functional localizer for places and scenes (Epstein and Kanwisher, 1998). We also included a functional localizer for body parts to identify the extrastriate body area (Downing et al., 2001), and for objects to identify the lateral occipital complex (Grill-Spector et al., 2001). This was done because both functional regions are located in the

occipitotemporal cortex, which was part of our anatomical hypotheses (see *Methods* and Supplementary material for details).

All of the employed actions involved objects that were also compatible or incompatible to the context. To dissociate action-context compatibility from object-context compatibility effects, we additionally employed pantomime actions as a control. Pantomime actions used action-incompatible objects that were either context-compatible or not. Accordingly, object-context incompatible vs. object-context compatible pantomime should reveal effects of object incompatibility, if present, as well. If in contrast (in-) compatibility effects would be observed only for actions, not for pantomime, these effects could not be ascribed to object-context (in-) compatibility, but unambiguously to action-context (in-) compatibility.

Methods

Behavioral experiment

Subjects

Forty-nine healthy volunteers (19–28 years, mean 23.7 years, 31 females, 2 left-handed) participated in the study. Three subjects did not enter the analysis because of poor performance (error rates higher than 2 standard deviations). All participants had normal or corrected-to-normal vision. Participants gave written consent before the experiment. Data were handled anonymously.

Stimuli

Thirty context-specific actions (e.g. using a stapler, cracking an egg, sawing a plank) were filmed in compatible, incompatible, and neutral contexts. Actions were object manipulations specific for the contexts “office”, “kitchen” and “workshop” (see appendix). Each context was specified by the background, the working surface, and 3–5 context-specific stationary objects (e.g. computer screen, coffee machine, grinding machine). The neutral context was constructed by stretching a white sheet on a table forming a uniform surface without any corners. In each context, actions were filmed from an allocentric perspective (60° to the left of the actress) providing a convenient view on the object manipulation and the contextual setting (Fig. 1). Each video had a length of 3 s, a presentation rate of 25 frames per second and a display width and height of 720 × 576 pixels.

The actions started with grasping of the objects about 680 ms after video onset. Care was taken that the three action versions were performed in exactly the same manner for each of the three experimental conditions (see Supplementary material for examples in which the three action versions are displayed simultaneously).

To ensure context-specificity and context-incompatibility in the incompatible condition, a pilot study was performed where an independent group of subjects delivered a judgment on a Likert rating scale from 1 to 6, indicating how likely they find it to observe the presented action in the particular context. All actions scored significantly higher for compatible than for incompatible contexts. The same Likert rating was employed in a post-experiment survey following the fMRI session (see *Results*).

Design and procedure

Subjects were seated approximately 60 cm away from a computer screen and next to the experimenter. Trials started self-paced by pressing a target button with the right index finger, followed by a short fixation phase (500 ms blank screen, 500 ms fixation cross at the center of the screen). Videos appeared at the center of the screen (subtending approximately 13.6 × 10.5° of visual angle). Subjects were instructed to press the target button as soon as they recognized the presented action. After button press, the video was replaced by a question mark at the center of the screen, and subjects had to name the recognized action. The verbal response was recorded using a built-in program of the presentation software (Presentation 13.1, Neurobehavioral Systems). The

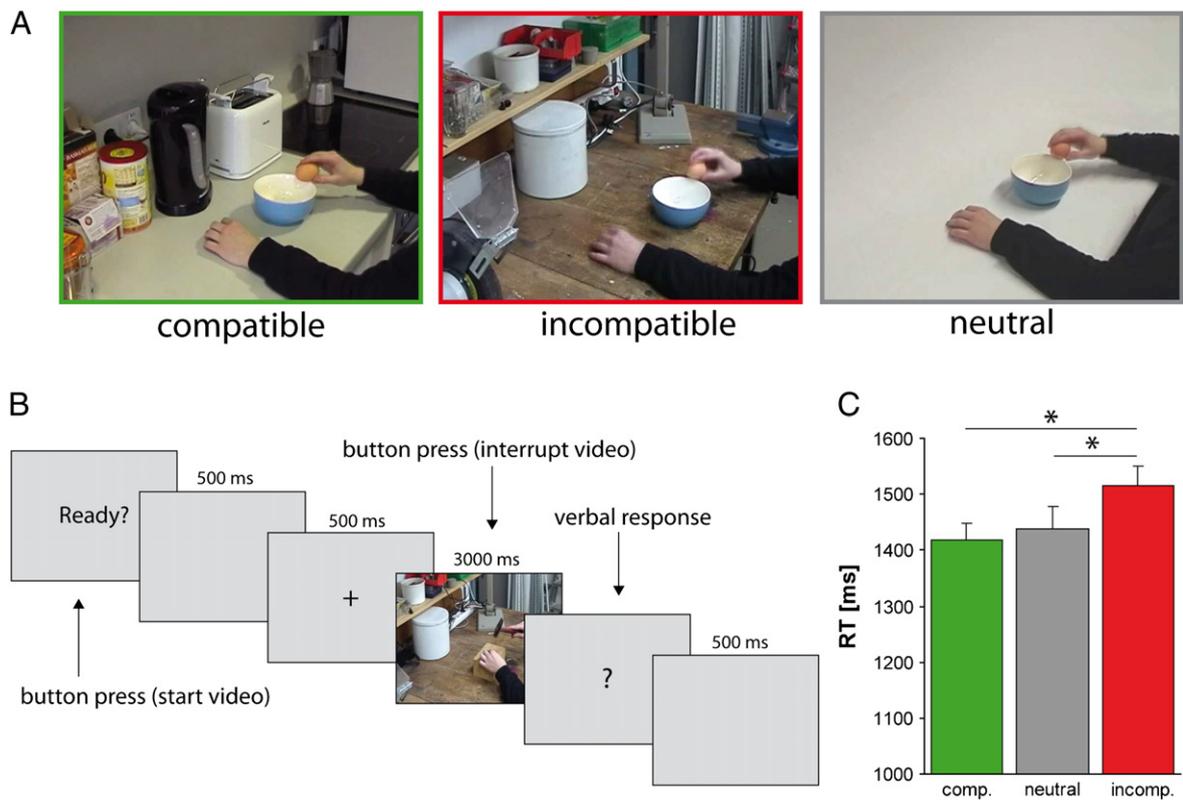


Fig. 1. Experimental conditions for both Exp. 1 and 2 (A). Trial design of the behavioral study (Exp. 1) (B). Results of Exp. 1; error bars indicate standard error of mean (C).

213 experimenter indicated by button press not visible to the subject
 214 whether the response was correct or not. Only correctly answered trials
 215 with RTs shorter than 3000 ms entered the statistical analysis. The trial
 216 ended with a blank screen for 500 ms.

217 Each subject watched each of the 30 actions only once during the ex-
 218 periment (10 trials per condition). Stimuli were hence balanced across
 219 three groups of subjects ($n = 15, 15, 16$). The occurrence of contexts
 220 was balanced within subjects so that each subject saw each of the
 221 three domestic settings 6–7 times and the neutral setting 10 times
 222 (Table 3). The trial order was pseudorandomized, so that transitions
 223 of contexts and transitions of conditions occurred equally often; i.e.,
 224 for each subject (= 30 trials) each of the 16 possible transitions be-
 225 tween the 4 contextual settings occurred 1–2 times, and each of the 9
 226 possible transitions between the 3 conditions occurred 3–4 times.

227 fMRI experiment

228 Subjects

229 Eighteen healthy volunteers (21–29 years, mean 24.9 years, 12 fe-
 230 males) were enrolled in the fMRI experiment. Subjects were right-
 231 handed according to the Edinburgh Inventory Manual Preference
 232 (Oldfield, 1971). All participants had normal or corrected-to-normal
 233 vision, and were native German speakers. No participant had a histo-
 234 ry of neurologic or psychiatric disorder. Participants gave written
 235 consent before fMRI measurement. The experimental standards
 236 were approved by the local ethics committee of the Medical Faculty,
 237 University of Cologne, Germany. Data were handled anonymously.

238 Stimuli and task

239 Stimuli of the behavioral experiment were used. Subjects were pre-
 240 sented with 30 trials per condition, ensuring that any within-subject
 241 differences between conditions could not be attributed to idiosyncra-
 242 sies of the actions. In addition to the conditions of the behavioral exper-
 243 iment we also employed additional conditions with pantomime actions

(i.e. actions performed with action-incompatible objects; Schubotz and
 244 von Cramon, 2009). Pantomime actions were either compatible or in-
 245 compatible to the context and involved objects that were either com-
 246 patible or incompatible to the context.

247
 248 In the present study, pantomime actions served to control for puta-
 249 tive object-context effects. We hence contrasted (1) object-context in-
 250 compatible with object-context compatible trials when actions were
 251 compatible to the context, and (2) object-context incompatible with
 252 object-context compatible trials when actions were compatible to the
 253 context. In a third contrast (3), we collapsed action-compatible and
 254 action-incompatible trials. The resulting two conditions were pantom-
 255 mime actions involving context-compatible objects and pantomime ac-
 256 tions involving context-incompatible objects, irrespective of action-
 257 context compatibility.

258 Subjects were instructed to observe the presented actions. They
 259 were informed that some of trials (20%) were followed by a verbal ac-
 260 tion description (question trial). Question trials were employed to en-
 261 sure that subjects paid attention to the actions. Because questions
 262 occurred well after presentation of the action clips, they were not
 263 expected to yield significant performance differences between condi-
 264 tions. Subjects had to indicate by button press whether the descrip-
 265 tion matched or did not match the action of the preceding trial ($n =$
 266 1). The description matched the action in 50% of the question trials.
 267 Every second question trial was followed by an empty trial in which
 268 subjects were instructed to fixate a small square in the center of the
 269 screen providing a resting baseline.

270 All trials had a length of 6 s, starting with a variable jitter (0, 500,
 271 1000, or 1500 ms) to enhance the temporal resolution of the BOLD re-
 272 sponse. Videos and verbal action descriptions were followed by a fix-
 273 ation cross until the next trial started.

274 The order of conditions per action was balanced across subjects to
 275 rule out habituation effects with respect to the actions. The trial order
 276 was pseudorandomized to equate transition probabilities between con-
 277 texts and transition probabilities between conditions. Furthermore, the

Table 1
Macroanatomical area, Brodmann area, Hemisphere (R, right; L, left), Talairach coordinates (x, y, z), and maximal Z scores of activation peaks in contrasts of experimental conditions (z-thresholded at 2.576, corrected cluster threshold $p < .05$).

Macroanatomical area	Brodmann area	Hemisphere	Talairach coordinates				mm ³
			x	y	z	Z	
<i>Incompatible ≥ compatible</i>							
Precentral gyrus (superior PMv)	6	L	-35	6	36	2.94	3078
IFG/IFS	44	L	-44	9	18	3.47	l.m.
IFG	44	L	-41	21	18	2.93	l.m.
IFG	45	L	-41	27	6	2.92	l.m.
IFG	47	L	-35	24	0	3.09	l.m.
<i>Incompatible ≥ neutral</i>							
IFG/IFS	44/45	L	-44	24	21	3.50	3105
IFG	47	L	-32	21	3	3.16	l.m.
PHC	35/36/37	L	-29	-51	0	4.42	14310
Cuneus/Fusiform gyrus	37	L	-23	-78	0	4.41	l.m.
PHC	35/36/37	R	22	-48	0	4.59	29646
Cuneus/Lingual gyrus	17/18	R	7	-84	0	6.69	l.m.
TOS	18/19	R	28	-78	27	3.18	l.m.
Thalamus, pulvinar		R	22	-27	0	3.21	l.m.
<i>Compatible ≥ neutral</i>							
PHC	35/36/37	L	-26	-51	0	4.85	2025
Cuneus/Lingual gyrus	17/18	L	-23	-78	0	4.40	11664
PHC	35/36/37	R	25	-45	0	4.68	32373
Cuneus/Lingual gyrus	17/18	R	10	-84	3	7.20	l.m.
Thalamus, Pulvinar		R	16	-28	0	3.40	l.m.

occurrence of question trials and baseline trials was balanced so that in each condition, there were the same numbers of trials preceded and followed by question and baseline trials.

Functional localizers

Following the main experiment, a functional localizer experiment adapted from Downing et al. (2006) was conducted: subjects viewed images from 4 categories in 16 blocks (4 blocks per category, randomized block order) with a length of 17 s, separated by a fixation screen for 8 s.

Categories were: places (indoor and outdoor), body parts (hands, arms, legs, face parts, torso), objects (tools and household articles), and scrambled objects (24×24 pixels). 40 full-color 720×576-pixel images per category were used; each image was presented twice in the experiment. Within each block, 20 images were presented for 300 ms; with an inter-stimulus-interval (ISI) of 650 ms. Subjects were instructed to indicate via button press with the right index finger when two identical images were presented in direct succession (one-back task, 1 or 2 repetitions per block).

Table 2
Macroanatomical area, Brodmann area, Hemisphere (R, right; L, left), Talairach coordinates (x, y, z), and maximal Z scores of activation peaks in conjunction analyses (* z-thresholded at 2.576, corrected cluster threshold $p < .05$).

Macroanatomical area	Brodmann area	Hemisphere	Talairach coordinates				mm ³
			x	y	z	Z	
<i>Conjunction (incompatible ≥ compatible) ∩ (incompatible ≥ neutral)*</i>							
IFG/IFS	44/45	L	-41	15	21	2.94	1296
IFG	47	L	-32	21	0	2.68	81
<i>Conjunction (compatible ≥ incompatible) ∩ (compatible ≥ neutral)</i>							
IFG	47	L	-41	30	3	-2.49	1107
Posterior MTG	37	L	-47	-57	6	-2.30	1026
	37	R	37	-51	-6	-2.59	2808

Table 3
Macroanatomical area, Brodmann area, Hemisphere (R, right; L, left), Talairach coordinates (x, y, z), and maximal Z scores of activation peaks in localizer contrasts (corrected cluster threshold $p < .05$).

Macroanatomical area	Brodmann area	Hemisphere	Talairach coordinates				mm ³
			x	y	z	Z	
<i>Actions in neutral context vs. fixation baseline</i>							
IFG/IFS	44/45	L	-38	33	21	4.29	104382
Precentral gyrus (superior PMv)	6	L	-50	0	36	5.76	l.m.
Precentral gyrus (PMd)	6	L	-20	0	51	5.95	l.m.
Anterior IPS	7	L	-53	-24	42	6.35	l.m.
Posterior IPS	7	L	-26	-54	54	6.17	l.m.
Occipito-temporal cortex	37/19	L	-40	-67	3	7.44	l.m.
IFG/IFS	44/45	R	40	30	15	4.29	86562
Precentral gyrus (superior PMv)	6	R	40	9	24	5.76	l.m.
Precentral gyrus (PMd)	6	R	31	-3	57	6.04	l.m.
Anterior IPS	7	R	28	-36	51	6.25	l.m.
Posterior IPS	7	R	19	-51	57	6.33	l.m.
Occipito-temporal cortex	37/19	R	37	-60	3	8.08	l.m.
SMA	6	L	1	-3	57	5.32	675
Thalamus (pulvinar)		L	-20	-30	6	5.33	1620
Cerebellum (VIIb/VIIIa)		L	16	-27	6	5.34	1944
		R	-29	-54	-42	6.40	6183
		R	16	-54	-39	6.55	5589
<i>Places ≥ objects, body parts</i>							
PHC (PPA)	35/36/37	L	-26	-45	-3	6.86	20412
RSC	26/29/30	L	-17	-57	21	5.86	l.m.
TOS	18/19	L	-23	-87	24	5.92	l.m.
PHC (PPA)	35/36/37	R	22	-42	-3	7.77	24543
RSC	26/29/30	R	10	-51	18	6.62	l.m.
TOS	18/19	R	10	-93	21	6.41	l.m.
<i>Body parts ≥ objects</i>							
Posterior MTG (EBA)	37/39/19	L	-50	-63	15	6.63	10206
		R	46	-60	6	6.59	15147
Fusiform gyrus	37	R	40	-48	-9	4.38	l.m.
Posterior IPS	7	L	-32	-42	54	5.49	19926
Anterior IPS	7	L	-47	-24	39	4.83	l.m.
Precentral gyrus (PMd)	6	L	-23	-3	48	4.60	l.m.
Precentral gyrus (superior PMv)	6	L	-50	12	30	4.62	l.m.
Posterior IPS	7	R	31	-30	45	5.67	15822
Anterior IPS	7	R	46	-24	42	4.54	l.m.
<i>Objects ≥ scrambled objects</i>							
Posterior MTG (LO)	37/19	L	-50	-69	12	6.68	32967
TOS	18/19	L	-41	-75	30	4.99	l.m.
Fusiform gyrus/PHC (pFs)	37	L	-32	-39	-9	6.37	l.m.
Posterior MTG (LO)	37/19	R	37	-72	9	6.73	32562
Fusiform gyrus/PHC (pFs)	36	R	25	-45	-3	5.94	l.m.

MRI data acquisition

After instruction and a short practice of the main task (not the localizers' task) outside the scanner, subjects were placed supine on the scanner bed with the right index or middle finger positioned on the response buttons. They were provided with earplugs and headphones to attenuate scanner noise. Form-fitting cushions were used to prevent arm and head motion.

The experiment was performed in a 3-T Siemens Trio system (Erlangen, Germany) using a standard birdcage head coil. 26 axial slices (192 mm field of view; 64×64 pixel matrix; 4 mm thickness; 1 mm

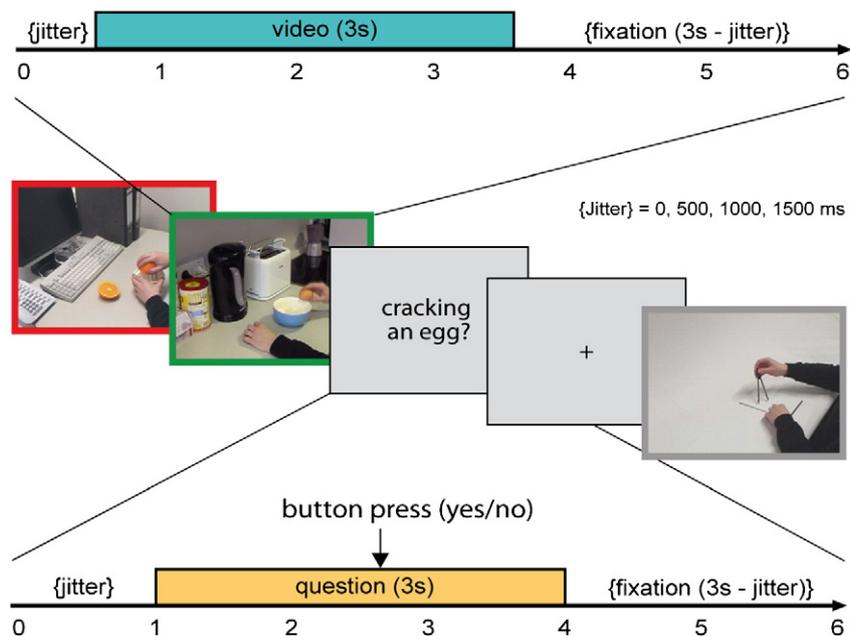


Fig. 2. Experimental design of fMRI experiment (Exp. 2).

spacing; in-plane resolution of 3×3 mm) covering the whole brain were acquired using a single-shot gradient EPI sequence (2000 ms repetition time; 30 ms echo time; 90° flip angle; 116 kHz acquisition bandwidth) sensitive to BOLD contrast. Functional imaging (90 images per condition of the main experiment, 32 images per category of the localizer experiment) was followed by acquisition of 26 anatomical T1-weighted MDEFT images (Norris, 2000). In a separate session, high resolution whole-brain images were acquired from each subject using a T1-weighted 3-D-segmented MDEFT sequence.

MRI data analysis

fMRI data were processed using the software package LIPSIA 1.5.0 (Lohmann et al., 2001). An offline motion-correction was performed with the Siemens motion-correction protocol PACE (Siemens, Erlangen, Germany). A cubic-spline interpolation was employed to correct for the temporal offset between slices acquired in one image. A temporal high-pass filter with a cut-off frequency of 1/80 Hz was used. Spatial smoothing was performed with a Gaussian filter of 5.65 mm FWHM. To align the functional data slices with a 3-D stereotactic coordinate reference system a rigid linear registration with six degrees of freedom (three rotational, three translational) was performed. An optimal match between the 3-D individual dataset and the EPI-T1 slices was achieved by using the rotational and translational parameters acquired on the basis of the MDEFT and the EPI-T1 slices. The MDEFT volume dataset with 160 slices and 1 mm slice thickness was standardized to the Talairach stereotactic space (Talairach and Tournoux, 1988). Rotational and translational parameters were subsequently transformed by linear scaling to a standard size. The resulting parameters were then used to transform the functional slices using trilinear interpolation, so that the resulting functional slices were aligned with the stereotactic coordinate system, thus generating output data with a spatial resolution of $3 \times 3 \times 3$ mm (27 mm^3). The statistical evaluation was based on a least-squares estimation using the general linear model for serially autocorrelated observations (Friston et al., 1995; Worsley and Friston, 1995). For the main experiment the design matrix was generated with a gamma function, convolved with the hemodynamic response function. Brain activations were analyzed time-locked to onset of the movies, and the analyzed epoch comprised of the full duration (3 s) of the presented movies, the duration of the null events (6 s) and the reaction time in question trials (max. 3 s), respectively. Analyzed epochs of the localizer

experiment comprised the duration of blocks (17 s) and were convolved with a Gaussian function. The model equation, including the observation data, the design matrix, and the error term, was convolved with a Gaussian kernel of dispersion of 4 s FWHM to account for the temporal autocorrelation (Worsley and Friston, 1995). In the following, contrast images, i.e. beta value estimates of the raw-score differences between specified conditions, were generated for each participant. As all individual functional datasets were aligned to the same stereotactic reference space, the single-subject contrast images were entered into a second-level random effects analysis for each of the contrasts.

One-sample *t* tests were employed for the group analyses across the contrast images of all subjects that indicated whether observed differences between conditions were significantly distinct from zero. The *t* values were subsequently transformed into *Z* scores.

To correct for false-positive results, in a first step, an initial voxelwise *z*-threshold was set to $z = 2.576$ ($p = .005$). In a second step, the results were corrected for multiple comparisons using cluster-size and cluster-value thresholds obtained by Monte Carlo simulations at a significance level $p = .05$, i.e., the reported activations are significantly activated at $p < .05$, corrected for multiple comparisons at the cluster level.

Conjunctions were calculated by extracting the minimum *z*-value of the two input contrasts for each voxel (Nichols et al., 2005).

Results

Behavioral experiment

Behavioral effects were considered significant at an α level of 0.05. Mean reaction times (RTs) were calculated for all conditions and each subject. Results are reported in Fig. 1. A repeated measures ANOVA indicated a significant effect of condition on RT ($F_{(2, 90)} = 4.39$, $p = 0.015$). RTs were significantly slower for the incompatible condition (mean \pm standard error of mean, 1515 ± 34 ms) compared to the compatible (1418 ± 30 ms; one-tailed paired *t* test, $t_{45} = 3.64$, $p < 0.001$) and neutral condition (1438 ± 41 ms; one-tailed paired *t* test, $t_{45} = 3.64$, $p = 0.021$). RTs between neutral and compatible condition did not differ significantly (one-tailed paired *t* test, $t_{45} = -0.5$, $p = 0.31$).

No significant effects were found for error rates (compatible: $6.0 \pm 1.1\%$, incompatible: $3.8 \pm 0.9\%$, neutral: $2.9 \pm 0.8\%$; repeated measures ANOVA: $F_{(2, 90)} = 2.30$, $p = 0.106$).

383 After the experiment, subjects were asked whether they noticed
 384 anything or anything odd in the videos. Only 26% (12 of 46 subjects)
 385 spontaneously recognized that some actions were performed in
 386 action-incompatible contexts.

387 *fMRI experiment*

388 *Behavioral performance*

389 Behavioral performance was assessed by analyzing responses to
 390 question trials. These referred to details of the object manipulation,
 391 not to contextual compatibility. Question trials were implemented
 392 to control for subjects' attention to the actions. Overall error rates
 393 were low (compatible: $3.3 \pm 1.8\%$, incompatible: $0.8 \pm 0.8\%$, neutral:
 394 $2.8 \pm 1.9\%$), ensuring that subjects paid attention to the presented

395 actions. Differences between conditions were not significant (repeat-
 396 ed measures ANOVA: $F_{(2, 34)} = 0.65, p = 0.528$).

397 Following the fMRI session, subjects evaluated the likelihood of
 398 each action to take place in each context. The Likert scale ranged
 399 from 1 (very unlikely) to 6 (very likely). Overall, context-compatible tri-
 400 als scored significantly higher (mean \pm standard deviation, 5.88 ± 0.25)
 401 than context-incompatible trials (2.19 ± 0.81 ; paired-samples t -test:
 402 $t_{17} = 34.17, p < 0.001$, one-tailed), corroborating our stimulus manipula-
 403 tion. Importantly, using the rating scores as regression coefficients in a
 404 parametric fMRI design yielded activation identical to the contrast in-
 405 compatible vs. compatible context (see section below).

406 *fMRI effects of incompatibility*

407 Interference effects were hypothesized to manifest as increases of
 408 BOLD response in the incompatible as compared to the compatible

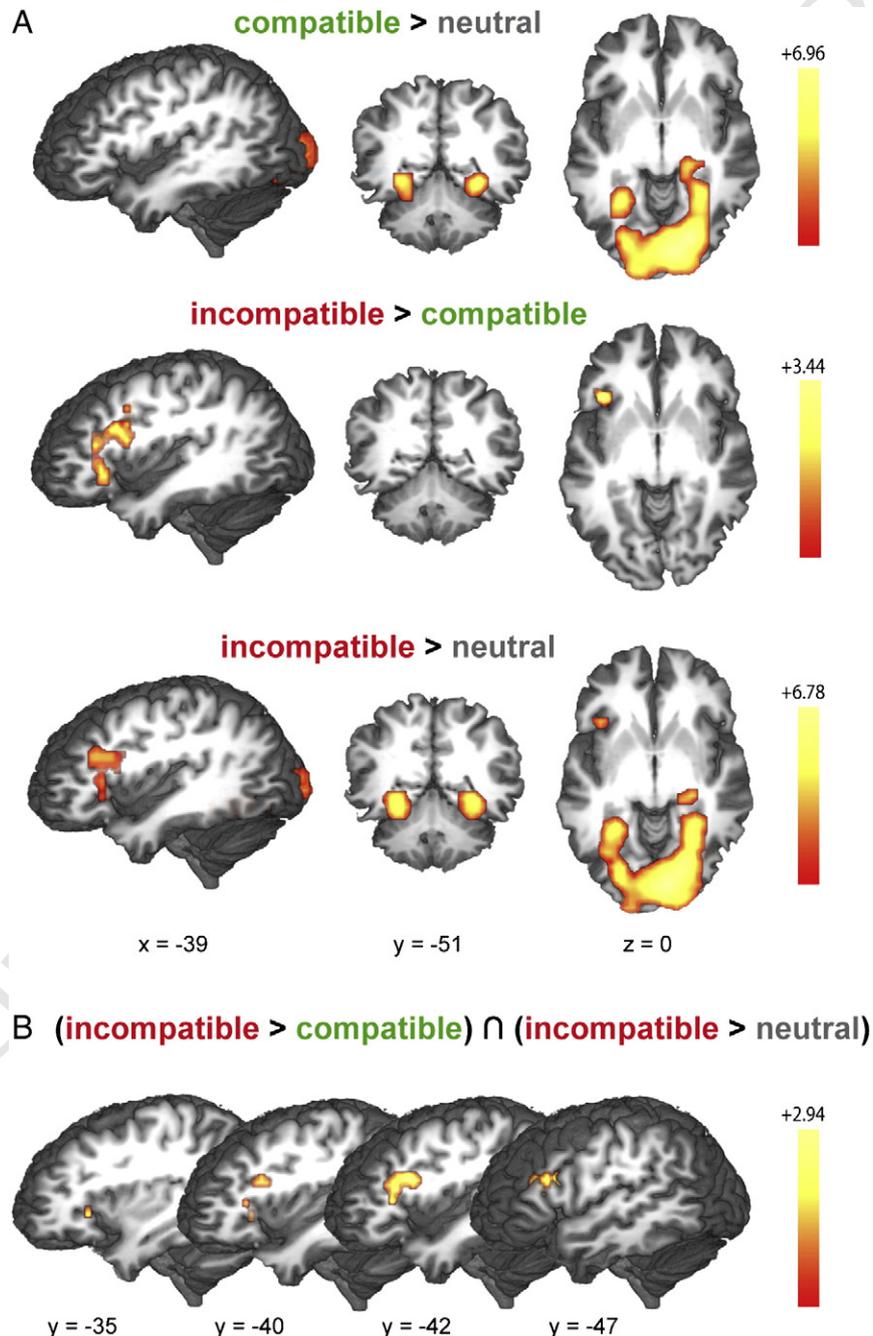


Fig. 3. Results of Experiment 2. (A) Direct contrasts between experimental conditions. (B) Conjunction of the contrasts (incompatible vs. compatible context) \cap (incompatible vs. neutral context), z-thresholded at 2.576, corrected cluster threshold $p = 0.05$.

and the neutral condition. Contrasting incompatible with compatible contexts revealed activation maxima in the left vIPFC (inferior frontal gyrus, IFG, and lower bank of inferior frontal sulcus, IFS) and the superior ventral premotor cortex (PMv, Fig. 2). A similar activation in the left vIPFC was also found in the contrast incompatible vs. neutral context (Fig. 2, Table 1). Further activation was found in inferior temporal cortex and parahippocampal gyri (PHG), and occipital regions including cuneus, lingual gyrus, and transverse occipital sulcus. These activations reflect different amounts of visual/contextual information in (in-) compatible vs. neutral contexts (cf. Supplementary material for a detailed discussion and ROI analysis).

No significant activation was found in the reverse contrast. The conjunction of both contrasts [(incompatible vs. compatible) \cap (incompatible vs. neutral)] confirmed the effect of incompatibility in the left vIPFC (Fig. 3).

fMRI effects of compatibility

Facilitatory effects were hypothesized to manifest as decreases of BOLD response in the compatible as compared to the incompatible and the neutral condition. The conjunction [(compatible vs. incompatible) \cap (compatible vs. neutral)] yielded a common overlap in the left IFG (BA 47) and in the occipitotemporal lobe bilaterally. The activations, however, were relatively weak (see Table 2). Even after lowering the significance level of the multiple comparisons corrections ($z = 1.96$) ($p = .025$), activations fell below the statistical threshold. Activations will hence not be discussed further.

Overlap of compatibility effects with regions sensitive to action recognition

To test whether regions sensitive to contextual compatibility of context and action overlap with regions sensitive to action recognition in general we computed the contrast actions in the neutral context vs. resting baseline. We then tested whether there is an overlap with the conjunction (incompatible vs. compatible) \cap (incompatible vs. neutral). Contrasting actions in the neutral context vs. rest yielded activation of dorsal and superior ventral premotor cortex, bilateral supplementary motor area, bilateral ventrolateral prefrontal cortex, left intraparietal sulcus, bilateral posterior occipitotemporal cortex, and right cerebellum. Activations overlapped with compatibility effects in left dorsal BA 44 (44d) and the posterior part of BA 45 (45p) but neither in the anterior part of BA 45 (45a; Amunts et al., 2010) nor in BA 47 (Fig. 4).

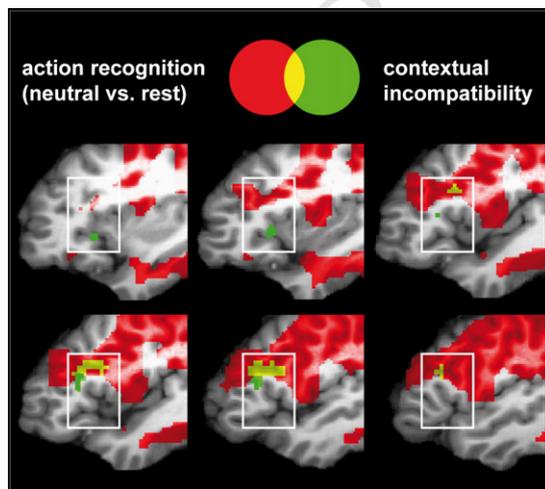


Fig. 4. Regions sensitive to contextual compatibility (green) overlap with regions sensitive to action recognition (red). Overlap of the conjunction (incompatible vs. neutral) \cap (compatible vs. compatible) with the contrast actions in neutral settings vs. rest; z -thresholded at 2.576, corrected cluster threshold $p = 0.05$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

fMRI effects object-context incompatibility

Since all actions involved objects, (in-) compatibility effects could in principle also arise before the action unfolds due to object-context (in-) compatibility. As a control, we contrasted object-context incompatible vs. object-context compatible trials in pantomime. This was done in three steps of analysis: (1) object-context incompatible vs. object-context compatible trials when actions were compatible to the context, (2) object-context incompatible vs. object-context compatible trials when actions were compatible to the context, and (3) object-context incompatible vs. object-context compatible trials irrespective of action-context (in-)compatibility (i.e. contrasts (1) and (2) were collapsed in a conjunction).

In neither case, there was increased activation for object-context incompatible trials.

Discussion

The present behavioral and fMRI experiments were carried out to investigate whether action recognition is modulated by contextual information derived from domestic settings. To this end, subjects watched context-specific everyday actions taking place in compatible, incompatible, and neutral contexts. The behavioral experiment revealed that subjects recognized actions significantly faster when they took place in a compatible than in an incompatible context. However, since recognition was also faster in neutral than in incompatible contexts, but not faster in compatible than in the neutral context, findings point to an interference effect of incompatible contexts, not to facilitatory effects of compatible ones. In accordance to this result, the fMRI experiment yielded evidence for interference effects. Increased activation was recorded for incompatibility in the left ventrolateral prefrontal cortex (vIPFC). Activation overlapped with regions sensitive to action perception in dorsal BA 44 and the posterior part of BA 45 but not in the anterior part of BA 45 and BA 47, pointing to a network extension in case of action-context conflict. There was no significant common decreasing activity in compatible as compared to either neutral or incompatible contexts, and hence no evidence in favor of facilitatory neural mechanisms.

Together, results suggest that incompatible settings exert interference on action recognition/analysis. In particular, activation of vIPFC in the fMRI experiment indicates that interference affected not the perceptual analysis of action properties but rather higher levels of action interpretation, i.e. the inference of hierarchically superordinate action structures, as will be discussed hereafter.

Slower responses to actions in incompatible contexts

In the behavioral experiment, reaction times reflected how long subjects needed to recognize the presented action. When actions were performed in an incompatible context, subjects were about 100 ms slower than for compatible contexts, and about 80 ms slower when compared to the neutral context. Interestingly, compatible contexts did not substantially facilitate action recognition, but incompatible ones interfered with it.

Knowledge of and thus habituation to a compatible context may have caused reduced usage of contextual information in compatible trials, resulting in reaction times comparable to the neutral trials. Alternatively, the absence of facilitatory effects may be due to behavioral ceiling effects. Thus, actions were always so easy to recognize that compatible settings could not further improve or accelerate their recognition. Another reason for the absence of facilitatory effects may be that background objects in the compatible and incompatible conditions might have distracted attention from the action. The lack of background in the neutral context might thus have positively affected the subjects' responses. Indeed, isolated objects are recognized more easily than objects embedded in a compatible context (Boyce et al., 1989; Davenport and Potter, 2004; Murphy and Wisniewski, 1989). We hence cannot rule out that putative facilitatory effects of compatible contexts were

masked/compensated by lower perceptual demands of neutral contexts. A possibility to overcome the problem of low-level perceptual differences between the neutral and the domestic settings could be to use a “scrambled” background in the neutral condition, providing the identical amount of visual information as the domestic settings. In addition, to account for mid-level perceptual differences (e.g. figure-ground segmentation, surface extraction, detection of contours and shapes, etc.) and recognition of background objects one would have to embed non-contextual (e.g. abstract) objects into the background. However, a background adjusted that way might be considered as unnatural and therefore evoke interference for that reason as well. Hence, interpretation and use of such a control condition would remain unclear. The choice of the appropriate background in a baseline condition is a generally unsolved problem that hampers a flawless differentiation between facilitation and interference in studies on visual scene effects (cf. Bar, 2004).

Neural effects of incompatibility

Observation of actions taking place in incompatible contexts as compared to either compatible or neutral ones increased activation in the left vlPFC, including BA 44/45 and BA 47. No significant effects were found in occipitotemporal regions, suggesting that recognition of perceptual properties of the action were not affected by contextual information. Likewise, contextual (in-) compatibility did not affect PPA activity, which was strongly activated by both action-compatible and action-incompatible contexts (cf. Supplementary material). In contrast, activation of the vlPFC rather suggests that contextual information influenced a high processing level during action analysis, for instance those relating to the inference of long-term goals. Dorsal BA 44 and posterior BA 45, but not BA 47 and anterior BA 45, overlapped with regions sensitive to action perception in general. This dissociation is in line with meta-analyses on action observation revealing involvement of BA 44 and BA 45 but not BA 47 (Caspers et al., 2010; Grezes and Decety, 2001; Van Overwalle and Baetens, 2009), and cytoarchitectonic studies that suggest a segregation of dorsal and ventral BA 44, and anterior and posterior BA 45 (Amunts et al., 2010).

Single-cell recordings in monkeys (Murata et al., 1997; Rizzolatti et al., 1988) as well as functional imaging (Rizzolatti and Craighero, 2004) and neuropsychological studies in humans (Buxbaum and Kalenine, 2010) suggest that left BA 44 of the vlPFC is involved in structuring sequences to realize particular outcomes (Fiebach and Schubotz, 2006; Grafman, 2002). During action perception, the same process may be exploited to generate top-down expectations about upcoming actions steps (Csibra, 2007; Fagg and Arbib, 1998; Friston et al., 2011; Schubotz and von Cramon, 2008). Here, contextual cues may modulate the concatenation of motor acts that are most likely to follow next. Indeed, observation of action seems to not only activate representations of the concurrent action but also of semantically related actions (Fogassi et al., 2005; Iacoboni et al., 2005). An imaging study where grasping objects in different situational contexts implied different actions to follow (grasping to drink and grasping to clean) engaged right BA 44 (Iacoboni et al., 2005). Related to this observation, we found that subjects, when asked to verbally indicate the actions in the behavioral experiment, they sometimes indicated not the actually presented action step, but rather a corresponding overarching action goal; i.e., they said “making pancake” instead of “cracking an egg”.

The activation of associated action representations may be a function of the probability that a given action usually occurs in a defined sequence. In terms of computations, domestic settings may shift the probability weights between semantically related action representations. For instance, observing someone squeezing a lemon triggers action memories that are likely to precede (halving the lemon) and to follow (pouring juice, preparing a drink, etc.). The likelihood of related actions to follow is equal in the compatible and in the neutral condition, because the latter does not negatively modulate the (high) probability weights between observed and the related, implied actions or action

steps. In contrast, in order to make sense of “squeezing a lemon in the bathroom”, alternative logically related actions compatible with both the action and the contextual setting are searched for, resulting in the inference of the underlying intention: Maybe in this case, the actor wants to decalcify the tap. This mechanism offers an explanation of how context-driven modulation of associative strengths between actions might contribute to long-term goal inference. Since prediction of subsequent action steps depends on knowledge about long-term goals, currently perceived action steps are probably embedded into the most likely action sequence. Activation of BA 44 and 45p overlapped with activation for actions in neutral settings vs. rest, in line with according literature (Caspers et al., 2010; Grezes and Decety, 2001; Van Overwalle and Baetens, 2009). This finding hence supports the assumptions that concatenation of events is involved in action perception in general (Caspers et al., 2010; Schubotz and von Cramon, 2004). In contrast, BA 47 and 45a may only be required when additional resources are needed. Thus, to resolve context-action incompatibility, alternative action goals need to be found or generated, charging retrieval and selection of (semantic or episodic) memory. BA 47 has been shown to play an important role in the controlled access to stored conceptual representations, whereas BA 45 is activated for the selection among competitively activated representations (Badre et al., 2005; Gold et al., 2006; Moss et al., 2005). In our study, BA 47 could be required for the controlled retrieval of semantic information that is necessary to come up with plausible solutions (“What can be done with lemon juice in the bathroom?”). Anterior BA 45 may operate after retrieval to select appropriate action representations that accounts best for both the observed action and action representations triggered by the incompatible setting. Interestingly, when subjects actively contemplated about the likelihood to find actions in incompatible contexts, ratings were relatively moderate, i.e. they rarely judged context-incompatible actions as “very unlikely” (see scores of the post-session rating in the Results section). This observation suggests that subjects had a bias toward – and were quite successful in – finding plausible alternative long-term goals.

In summary, enhancement of activation in BA 44, 45 and 47 by contextual incompatibility points to enhanced memory search and the attempt to overcome the discrepancy by a plausible overarching action goal. We therefore suggest that RT differences measured in our behavioral experiment most probably evolved from processing stages where the perceptual properties of an action were already recognized and the activated representation in semantic memory conflicts with context-activated action representations on the goal level. An electrophysiological study on the effects of settings on object recognition measured earliest context compatibility effects around 300 ms after stimulus onset and the morphology and time course of the event-related potential resembled the N400 congruity signal typically observed in semantic context paradigms (Ganis and Kutas, 2003). We hence suggest that visual context information does not measurably affect action recognition before respective semantic information is activated, at least when actions are very easy to recognize.

Limitations

The task we employed required subjects to focus their attention on the action. Still, it is principally possible that interference effects arose by object-setting incompatibility. In this case, effects of incompatibility should have been presented in the control pantomime actions as well. However, this was not observed. We hence can conclude that reported incompatibility effects were exclusively driven by action-setting incompatibility.

With respect to the behavioral experiment, we cannot exclude object-setting effects. Since actions unfold in time, however, one may argue that facilitated object recognition would not substantially interact with action recognition. This is because objects can be recognized within 150 ms (Thorpe et al., 1996) and responses to object

recognition typically occur around 600 ms after stimulus onset (Ganis and Kutas, 2003; Thorpe et al., 1996). In our experiment, actions began about 680 ms after video onset, i.e., well after objects were recognized, and responses occurred after about 1400 ms after movie onset. Thus, when our actions started to unfold, objects most likely were already recognized in both compatible and incompatible settings. Putative facilitated object recognition which may have occurred in our experiment is therefore unlikely to substantially contribute to the effects measured in the behavioral experiment.

Further experiments are needed to unambiguously test whether action-context incompatibility delays the inference of the action goal, or object-context incompatibility, or both.

Concluding remarks

To sum up, domestic settings – although actually ignorable in our experiments – affected action recognition. This was revealed by slower recognition (Experiment 1) and by increased neural activity in the left ventrolateral prefrontal cortex (Experiment 2) when actions took place in incompatible settings. Effects are suggested to reflect interference between the currently observed action and actions implied by the incompatible setting. Increased neural responses likely reflect the embedding of the observed action into a logically possible action sequence (BA 44); this requires the retrieval of overarching action goals that are compatible with both the observed action and the action-incompatible setting (BA 47), and the selection among these alternatives in order to solve the experimentally induced conflict in the best-possible way (BA 45).

Supplementary materials related to this article can be found online at [doi:10.1016/j.neuroimage.2011.08.038](https://doi.org/10.1016/j.neuroimage.2011.08.038).

Acknowledgments

We cordially thank Christiane Ahlheim, Anne Kühn, and Anne-Marie Schiffer for helpful comments on the manuscript, Anna Schneider for support in generating the stimulus material, and Christiane Ahlheim and Anika Sierk for experimental assistance.

Appendix

List of employed actions and means (with standard deviations) of rating scores obtained in a post-session survey of the fMRI experiment. Subjects were asked for the likelihood to observe actions in the compatible and the incompatible setting. Ratings were from 1 (very unlikely) to 6 (very likely).

Action	Compatible	Incompatible
Whisking	5.72 ± 1.18	1.28 ± 0.46
Buttering a bread	6.0 ± 0.0	2.0 ± 1.08
Cracking an egg	6.0 ± 0.0	1.94 ± 1.0
Peeling a potato	6.0 ± 0.0	1.0 ± 0.0
Slicing a carrot	6.0 ± 0.0	1.17 ± 0.38
Squeezing a lemon	5.94 ± 0.24	2.06 ± 1.35
Grating cheese	5.83 ± 0.38	1.28 ± 0.67
Opening a bottle	5.67 ± 1.19	4.0 ± 1.03
Drying dishes	6.0 ± 0.0	2.56 ± 1.29
Opening a can	6.0 ± 0.0	2.28 ± 1.18
Drawing a line with a ruler	5.44 ± 0.62	4.56 ± 1.34
Using a compass	5.0 ± 1.03	4.5 ± 0.86
Sharpening a pencil	5.61 ± 1.20	4.50 ± 1.29
Erasing	5.83 ± 0.38	4.33 ± 1.14
Sticking a stamp on a letter	6.0 ± 0.0	1.11 ± 0.32
Opening a letter with paper knife	5.94 ± 0.24	3.39 ± 1.82
Unlocking a cash box	4.28 ± 1.45	2.89 ± 1.57
Stamping	5.89 ± 0.32	2.0 ± 1.28
Hole-punching	6.0 ± 0.0	2.17 ± 1.38
Stapling sheets	5.72 ± 0.96	2.17 ± 1.54

Appendix (continued)

Action	Compatible	Incompatible
Sawing	6.0 ± 0.0	1.0 ± 0.0
Screwing with screwdriver	5.94 ± 0.24	2.0 ± 0.97
Hammering nails	5.94 ± 0.24	1.94 ± 1.06
Planing wood	5.89 ± 0.32	1.0 ± 0.0
Painting	5.44 ± 0.98	2.06 ± 1.06
Rasping wood	5.89 ± 0.32	1.17 ± 0.51
Pulling out nail	5.89 ± 0.47	1.72 ± 0.67
Removing isolation of wire	5.8 ± 0.38	2.06 ± 1.06
Screwing with wrench	5.89 ± 0.32	2.28 ± 0.96
Isolating wire	5.67 ± 0.59	1.78 ± 1.0

References

- Amunts, K., Lenzen, M., Friederici, A.D., Schleicher, A., Morosan, P., Palomero-Gallagher, N., Zilles, K., 2010. Broca's region: novel organizational principles and multiple receptor mapping. *PLoS Biol.* 8.
- Badre, D., Poldrack, R.A., Pare-Blagoev, E.J., Inslar, R.Z., Wagner, A.D., 2005. Dissociable controlled retrieval and generalized selection mechanisms in ventrolateral prefrontal cortex. *Neuron* 47, 907–918.
- Bar, M., 2004. Visual objects in context. *Nat. Rev. Neurosci.* 5, 617–629.
- Bar, M., 2007. The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci.* 11, 280–289.
- 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, 149–159.
- Bedny, M., McGill, M., Thompson-Schill, S.L., 2008. Semantic adaptation and competition during word comprehension. *Cereb. Cortex* 18, 2574–2585.
- Boyce, S.J., Pollatsek, A., Rayner, K., 1989. Effect of background information on object identification. *J. Exp. Psychol. Hum. Percept. Perform.* 15, 556–566.
- Buxbaum, L.J., Kalenine, S., 2010. Action knowledge, visuomotor activation, and embodiment in the two action systems. *Ann. N. Y. Acad. Sci.* 1191, 201–218.
- Cardillo, E.R., Aydelott, J., Matthews, P.M., Devlin, J.T., 2004. Left inferior prefrontal cortex activity reflects inhibitory rather than facilitatory priming. *J. Cogn. Neurosci.* 16, 1552–1561.
- 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, 1148–1167.
- Csibra, G., 2007. Action mirroring and action interpretation: an alternative account. In: Haggard, P., Rosetti, Y., Kawato, M. (Eds.), *Sensorimotor Foundations of Higher Cognition: Attention and Performance, XXII*. Oxford University Press, Oxford, pp. 435–459.
- Davenport, J.L., Potter, M.C., 2004. Scene consistency in object and background perception. *Psychol. Sci.* 15, 559–564.
- Downing, P.E., Jiang, Y., Shuman, M., Kanwisher, N., 2001. A cortical area selective for visual processing of the human body. *Science* 293, 2470–2473.
- Downing, P.E., Chan, A.W., Peelen, M.V., Dodds, C.M., Kanwisher, N., 2006. Domain specificity in visual cortex. *Cereb. Cortex* 16, 1453–1461.
- Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. *Nature* 392, 598–601.
- Fagg, A.H., Arbib, M.A., 1998. Modeling parietal-premotor interactions in primate control of grasping. *Neural Netw.* 11, 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, 499–502.
- Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., Rizzolatti, G., 2005. Parietal lobe: from action organization to intention understanding. *Science* 308, 662–667.
- Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D., Frackowiak, R.S., 1995. Statistical parametric maps in functional imaging: a general linear approach. *Hum. Brain Mapp.* 2, 189–210.
- Friston, K., Mattout, J., Kilner, J., 2011. Action understanding and active inference. *Biol. Cybern.* 104, 137–160.
- Ganis, G., Kutas, M., 2003. An electrophysiological study of scene effects on object identification. *Brain Res. Cogn. Brain Res.* 16, 123–144.
- 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. *J. Neurosci.* 26, 6523–6532.
- Grafman, J., 2002. The structured event complex and the human prefrontal cortex. In: Stuss, D.T., Knight, R.T. (Eds.), *Principles of Frontal Lobe Function*. Oxford University Press, New York.
- Grezes, J., Decety, J., 2001. Functional anatomy of execution, mental simulation, observation, and verb generation of actions: a meta-analysis. *Hum. Brain Mapp.* 12, 1–19.
- Grill-Spector, K., Kourtzi, Z., Kanwisher, N., 2001. The lateral occipital complex and its role in object recognition. *Vision Res.* 41, 1409–1422.
- Hamilton, A.F., Grafton, S.T., 2007. The motor hierarchy: from kinematics to goals and intentions. In: Haggard, P., Rossetti, Y., Kawato, M. (Eds.), *Sensorimotor Foundations of Higher Cognition*. Oxford University Press, Oxford, pp. 381–408.
- Henson, R.N., 2003. Neuroimaging studies of priming. *Prog. Neurobiol.* 70, 53–81.
- 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, e79.

- 745 Jastorff, J., Orban, G.A., 2009. Human functional magnetic resonance imaging reveals
746 separation and integration of shape and motion cues in biological motion proces-
747 sing. *J. Neurosci.* 29, 7315–7329.
- 748 Jastorff, J., Kourtzi, Z., Giese, M.A., 2009. Visual learning shapes the processing of com-
749 plex movement stimuli in the human brain. *J. Neurosci.* 29, 14026–14038.
- 750 Jeannerod, M., 2006. *Motor Cognition: What Action Tells the Self.* Oxford University
751 Press, New York.
- 752 Johnson-Frey, S.H., Maloof, F.R., Newman-Norlund, R., Farrer, C., Inati, S., Grafton, S.T.,
753 2003. Actions or hand–object interactions? Human inferior frontal cortex and ac-
754 tion observation. *Neuron* 39, 1053–1058.
- 755 Kable, J.W., Chatterjee, A., 2006. Specificity of action representations in the lateral occi-
756 pitotemporal cortex. *J. Cogn. Neurosci.* 18, 1498–1517.
- 757 Kilner, J.M., Friston, K.J., Frith, C.D., 2007. Predictive coding: an account of the mirror
758 neuron system. *Cogn. Process.* 8, 159–166.
- 759 Lohmann, G., Muller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., Zysset, S., von Cramon,
760 D.Y., 2001. LIPSI—A new software system for the evaluation of functional magnetic
761 resonance images of the human brain. *Comput. Med. Imaging Graph.* 25, 449–457.
- 762 Moss, H.E., Abdallah, S., Fletcher, P., Bright, P., Pilgrim, L., Acres, K., Tyler, L.K., 2005.
763 Selecting among competing alternatives: selection and retrieval in the left inferior
764 frontal gyrus. *Cereb. Cortex* 15, 1723–1735.
- 765 Murata, A., Fadiga, L., Fogassi, L., Gallese, V., Raos, V., Rizzolatti, G., 1997. Object representation
766 in the ventral premotor cortex (area F5) of the monkey. *J. Neurophysiol.* 78, 2226–2230.
- 767 Murphy, G.L., Wisniewski, E.J., 1989. Categorizing objects in isolation and in scenes:
768 what a superordinate is good for. *J. Exp. Psychol. Learn. Mem. Cogn.* 15, 572–586.
- 769 Neely, J.H., 1977. Semantic priming and retrieval from lexical memory: evidence for fa-
770 cilitatory and inhibitory processes. *Mem. Cognit.* 4, 648–654.
- 771 Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.B., 2005. Valid conjunction infer-
772 ence with the minimum statistic. *Neuroimage* 25, 653–660.
- 773 Norris, D.G., 2000. Reduced power multislice MDEFT imaging. *J. Magn. Reson. Imaging*
774 11, 445–451.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inven- 775
776 tory. *Neuropsychologia* 9, 97–113.
- Rizzolatti, G., Craighero, L., 2004. The mirror-neuron system. *Annu. Rev. Neurosci.* 27, 777
778 169–192.
- Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., Matelli, M., 1988. 779
780 Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and
781 the control of distal movements. *Exp. Brain Res.* 71, 491–507.
- Schmahmann, J.D., Pandya, D.N., 2006. *Fiber Pathways of the Brain.* Oxford University 782
783 Press, New York.
- Schmahmann, J.D., Pandya, D.N., Wang, R., Dai, G., D'Arceuil, H.E., de Crespigny, A.J., Wed- 784
785 een, V.J., 2007. Association fibre pathways of the brain: parallel observations from dif-
786 fusion spectrum imaging and autoradiography. *Brain* 130, 630–653.
- Schubotz, R.I., 2007. Prediction of external events with our motor system: towards a 787
788 new framework. *Trends Cogn. Sci.* 11, 211–218.
- Schubotz, R.I., von Cramon, D.Y., 2004. Sequences of abstract nonbiological stimuli 789
790 share ventral premotor cortex with action observation and imagery. *J. Neurosci.*
791 24, 5467–5474.
- Schubotz, R.I., von Cramon, D.Y., 2009. The case of pretense: observing actions and in- 792
793 ferring goals. *J. Cogn. Neurosci.* 21, 642–653.
- Talairach, J., Tournoux, P., 1988. *Co-planar Stereotaxic Atlas of the Human Brain.* 794
795 Thieme, New York.
- Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. 796
797 *Nature* 381, 520–522.
- Van Overwalle, F., Baetens, K., 2009. Understanding others' actions and goals by mirror 798
799 and mentalizing systems: a meta-analysis. *Neuroimage.*
- van Schie, H.T., Toni, I., Bekkering, H., 2006. Comparable mechanisms for action and 800
801 language: neural systems behind intentions, goals, and means. *Cortex* 42,
802 495–498.
- Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited—again. *Neuro-* 803
804 *image* 2, 173–181.

TITLE: The context-object-manipulation (COM) triad: cross talk during action perception revealed by fMRI

ABSTRACT

To recognize an action, an observer exploits information about the applied manipulation, the involved objects and the context where the action occurs. Context, object, and manipulation information are hence expected to be tightly coupled in a triadic relationship (the COM triad hereafter).

The current fMRI study aimed to investigate the hemodynamic signatures of reciprocal modulation in the COM triad. Subjects watched short video clips of pantomime actions, that is actions performed with inappropriate objects, taking place at compatible or incompatible contexts. The usage of pantomime actions enabled the disentanglement of the neural substrates of context-manipulation (CM) and context-object (CO) associations. There were trials in which (1) both actions and objects, (2) only actions, (3) only objects, or (4) neither actions nor objects were compatible with the context.

Increased hemodynamic responses were only observed when both manipulations and objects were compatible with the context, compared to either manipulation-compatible or object-compatible trials. CM compatibility effects were found in an action-related network comprising bilateral ventral premotor cortex (PMv), supplementary motor area (SMA), left anterior intraparietal sulcus (aIPS), and bilateral occipito-temporal cortex (OTC). Conversely, CO compatibility effects were found in bilateral lateral occipital complex (LOC). An overlap of CM and CO effects was

observed in the dorsal attention network, i.e. superior frontal sulcus/dorsal premotor cortex (PMd/SFS) and superior parietal lobe (SPL).

Results indicate that contextual information is integrated into the analysis of actions. Manipulation and object information is linked by contextual associations as a function of co-occurrence in specific contexts. Activation of either CM or CO associations shifts attention to either action- or object-related relevant information.

1. INTRODUCTION:

Object manipulations normally take place in particular contextual places. There is hence a tight, triadic coupling between context (C), objects (O), and manipulations (M; the COM triad hereafter). From a neural perspective, and as a result of frequency-based learning, the COM triad is likely to be reflected by associative strengths between the three kinds of information (Turk-Browne et al., 2010). Indeed, contextual information has been demonstrated to affect recognition of both objects (Bar, 2004) and manipulations (Wurm and Schubotz, in press). However, it remains unclear how context-manipulation (CM) and context-object (CO) couplings interact during action recognition.

In the present fMRI study, we investigated the neural effects of manipulation-context and object-context associations in a unitary experimental design. We manipulated the compatibility (1) between context and manipulation information and (2) between context and object information independently. This was achieved by decoupling manipulation from their associated objects in the COM triad. To this end, subjects were presented videos showing pantomime actions, i.e. manipulation indicative of a particular action but actually performed with inappropriate objects (Schubotz and von Cramon, 2009). We thereby realized a 2x2 factorial design: First, manipulations were either compatible (comM) or incompatible (incomM) with the context. Second, pantomime actions involved objects that were either compatible (comO) or incompatible (incomO) with the context (see Figure 1). There were hence four conditions where (1) both manipulations and objects (comM/comO), (2) only manipulations (comM/incomO), (3) only objects (incomM/comO), or (4) neither manipulations nor objects (incomM/incomO) were compatible with the context.

A recent study on recognition of pantomime and real action found that the processing of manipulation information draws on the cortical motor system (left ventral premotor cortex (PMv), left anterior intraparietal sulcus (aIPS) and the occipito-temporal cortex (OTC), whereas the processing of object information involved the lateral occipital complex (LOC; Schubotz and von Cramon, 2009). We hypothesized that the same neural networks are differentially engaged when either manipulation or object information is compatible with the context or not. Thus, CM compatibility effects should manifest as BOLD response differences in the motor system. Likewise CO compatibility effects should manifest as BOLD response differences in the LOC.

Moreover, due to the triadic relationship between context, objects, and manipulation, we expected interactions between CM and CO compatibility effects. Of particular interest was the concurrence of both CM and CO compatibility (see upper left cell of Figure 1). Here, the manipulation (opening a tin) and the objects (whisk and bowl) are associated with the same context (kitchen). This constellation is potentially conflicting because both objects and manipulations are reinforced by the context, but objects as well as object-implied invalid manipulations have to be suppressed or ignored. Hence, we expected that the influence of context on object and manipulation processing becomes particularly evident in this high-conflict situation when manipulation and objects share the same contextual affiliation. Based on the same findings outlined above, we expected that in the contrast comM/comO vs. incomM/comO an increased BOLD response would be observed in the motor system due to conflicting activation/suppression of manipulation information. Likewise, in the contrast comM/comO vs. comM/incomO, an increased response was expected in the LOC due to conflicting activation/suppression of object information. Finally, both

contrasts were expected to reveal increased responses in regions that are generally involved in interference resolution, including the dorsolateral prefrontal cortex and the rostral cingulate zone (Nee et al., 2007).

2. METHODS

2.1 Subjects:

Eighteen healthy volunteers (21-29 years, mean 24.9 years, 12 females) participated in the fMRI experiment. All subjects were right-handed according to the Edinburgh Inventory Manual Preference (Oldfield, 1971), had normal or corrected-to-normal vision, and were native German speakers. No participant had a history of neurologic or psychiatric disorder. Participants gave written consent before fMRI measurement. The experimental standards were approved by the local ethics committee of the Medical Faculty, University of Cologne, Germany. Data were handled anonymously.

2.2 Stimuli and Task:

Stimuli consisted of thirty context-specific pantomime actions (e.g. pantomiming to open a tin with a rasp and a piece of wood instead of a tin and a tin opener) filmed in compatible and incompatible contexts. In addition, the same actions were also filmed in a neutral setting. Actions were specific for the contexts "office", "kitchen" and "workshop". Contexts were specified by the background, the working surface, and 3-5 context-specific stationary objects (e.g. computer screen, coffee machine, grinding machine) placed on the working surface. The neutral context

consisted of a white sheet stretched on a table forming a uniform surface without any corners. Actions were filmed from an allocentric perspective (60° to the left of the actress) ensuring a convenient view on the action and the background (Fig. 1).

The following action conditions were employed in the experiment: (1) context-compatible manipulations performed with context-compatible objects (comM/comO), (2) context-incompatible manipulations performed with context-compatible objects (incomM/comO), (3) context-compatible manipulations performed with context-incompatible objects (comM/incomO), (4) context-incompatible manipulations performed with context-incompatible objects (incomM/incomO), and (5) the same set of pantomime actions in neutral contexts. Each video had a length of 3 s, a presentation rate of 25 frames per second and a display width and height of 720 x 576 pixels. Care was taken that movement trajectories of each pantomime action were performed in exactly the same manner for each of the five conditions. 30 trials per condition were presented in a event-related design, intermixed with three conditions employing normal actions (Wurm and Schubotz, in press).

Subjects were instructed to infer the presented actions to ensure that subjects were able to recognize and paid attention to the actions. They were informed that some of trials (20%) were followed by verbal action descriptions (question trials). Subjects had to indicate by button press whether a description matched or did not match the action of the preceding trial (n-1). The descriptions matched the action in 50% of the question trials. Every second question trial was followed by an empty trial in which subjects were instructed to fixate a small dot in the center of the screen, providing a resting baseline.

All trials had a length of 6 s, starting with a variable jitter (0, 500, 1000, or 1500 ms) to enhance the temporal resolution of the BOLD response. Videos and verbal action descriptions were followed by a fixation dot until the next trial started.

The trial order was pseudorandomized to equate transition probabilities between settings and transition probabilities between conditions. Furthermore the occurrence of question trials and baseline trials were balanced so that in each condition there were the same number of trials preceded and followed by question and baseline trials. Finally, to rule out habituation effects with respect to the actions, the succession of conditions was balanced across subjects for each action.

2.4 MRI Data Acquisition:

After instruction and a short practice of the experiment outside the scanner, subjects were placed supine on the scanner bed, with the right index or middle finger positioned over the two buttons of a response box. To attenuate scanner noise subjects were provided with earplugs and headphones. Form-fitting cushions were used to prevent arm and head motion.

Imaging was performed in a 3-T Siemens Trio system (Erlangen, Germany) using a standard birdcage head coil. 26 axial slices (192 mm field of view; 64 x 64 pixel matrix; 4 mm thickness; 1 mm spacing; in-plane resolution of 3 x 3 mm) covering the whole brain were acquired using a single-shot gradient EPI sequence (2000 ms repetition time; 30 ms echo time; 90° flip angle; 116 kHz acquisition bandwidth) sensitive to BOLD contrast. Functional imaging (90 images per condition) was followed by acquisition of 26 anatomical T1-weighted MDEFT images (Norris,

2000). In a separate session, high resolution whole-brain images were acquired from each subject using a T1-weighted 3-D-segmented MDEFT sequence.

2.5 MRI Data Analysis

fMRI data were processed using the software package LIPSIA 1.5.0 (Lohmann et al., 2001). An offline motion-correction was performed with the Siemens motion-correction protocol PACE (Siemens, Erlangen, Germany). A cubic-spline interpolation was employed to correct for the temporal offset between slices acquired in one image. A temporal high-pass filter with a cut-off frequency of 1/80 Hz was used. Spatial smoothing was performed with a Gaussian filter of 8 mm FWHM. To align the functional data slices with a 3-D stereotactic coordinate reference system a rigid linear registration with six degrees of freedom (three rotational, three translational) was performed. An optimal match between the 3-D individual dataset and the EPI-T1 slices was achieved by using the rotational and translational parameters acquired on the basis of the MDEFT and the EPI-T1 slices. The MDEFT volume dataset with 160 slices and 1 mm slice thickness was standardized to the Talairach stereotactic space (Talairach and Tournoux, 1988). Rotational and translational parameters were subsequently transformed by linear scaling to a standard size. The resulting parameters were then used to transform the functional slices using trilinear interpolation, so that the resulting functional slices were aligned with the stereotactic coordinate system, thus generating output data with a spatial resolution of 3 x 3 x 3 mm (27 mm³). The statistical evaluation was based on a least-squares estimation using the general linear model for serially autocorrelated observations (Friston et al., 1995; Worsley and Friston, 1995). For the main experiment the design matrix was generated with a gamma function, convolved with the hemodynamic response function (HRF). Brain

activations were analyzed time-locked to onset of the movies, and the analyzed epoch comprised of the full duration (3 s) of the presented movies, the duration of the null events (6 s) and the reaction time in question trials (max. 3 s), respectively. Analysed epochs of the localizer experiment comprised the duration of blocks (17 s) and were convolved with a Gaussian function. The model equation, including the observation data, the design matrix, and the error term, was convolved with a Gaussian kernel of dispersion of 4 s FWHM to account for the temporal autocorrelation (Worsley & Friston, 1995). In the following, contrast images, i.e. beta value estimates of the raw-score differences between specified conditions, were generated for each participant. As all individual functional datasets were aligned to the same stereotactic reference space, the single-subject contrast images were entered into a second-level random effects analysis for each of the contrasts.

One-sample t tests were employed for the group analyses across the contrast images of all subjects that indicated whether observed differences between conditions were significantly distinct from zero. The t values were subsequently transformed into Z scores.

To correct for false-positive results, in a first step, an initial voxelwise z -threshold was set to $z = 2.576$ ($p = .005$). In a second step, the results were corrected for multiple comparisons using cluster-size and cluster-value thresholds obtained by Monte Carlo simulations at a significance level $p = .05$, i.e., the reported activations are significantly activated at $p < .05$, corrected for multiple comparisons at the cluster level.

Conjunctions were calculated by outputting the minimum z -value of the two input contrasts for each voxel (Nichols et al., 2005).

3. RESULTS

3.1. Behavioral results:

Behavioral performance was monitored during the experiment to ensure that subjects were paying attention to the presented actions. Performance was assessed by analysing the error rate to question trials. 88.5% (\pm 1.1%, standard error of mean) of question trials were answered correctly, indicating adequate performance. Differences in performance between conditions was not analysed because of low statistical power (20% of trials, i.e. 6 of 30 trials per condition, were followed by question trials).

3.2. fMRI results:

3.2.1. context-manipulation (CM) compatibility

Effects of CM compatibility were expected to manifest in (1) the contrast incompatible comM/incomO vs. incomM/incomO), and (2) comM/comO vs. incomM/comO).

With respect to (1), no significant increases or decreases in BOLD response were found. With respect to (2), increases in BOLD response were found in the motor system, i.e. bilateral superior PMv extending into inferior frontal gyrus in the left hemisphere, bilateral supplementary motor area (SMA), left supramarginal gyrus (SMG), left aIPS, and bilateral OTC. Furthermore, activation was found in bilateral

dorsal premotor cortex/superior frontal sulcus (PMd/SFS) and left superior parietal lobe (SPL; Fig. 2A).

3.2.2. context-object (CO) compatibility

Effects of CO compatibility were expected to manifest in the contrasts (3) *incomM/comO* vs. *incomM/incomO*), and (4) *comM/comO* vs. *comM/incomO*).

With respect to (3), no significant increased or decreased activation was found. With respect to (4), increased activation was found in bilateral lateral occipital cortex (LO) and bilateral posterior parahippocampal and fusiform gyrus (pFs). Activation (left LO: -38, -75, 3; right LO: 19, -87, 6; left pFs: -29, -48, -3; right pFs: 28, -37, -9) overlapped with activation obtained in a localizer scan for objects vs. scrambled objects (left LO: -50, -69, 12; right LO: 37, -72, 9; left pFs: -32, -39, -9; right pFs: 25, -45, -3; Wurm and Schubotz, in press). Additional activation was found in bilateral PMd/SFS and bilateral SPL, right superior temporal gyrus/SMG and right superior PMv (Fig. 2B).

3.2.3. Dissociation of CM and CO compatibility effects

CM- and CO-specific and common activations were identified by computing an overlap of the corrected contrasts *comM/comO* vs. *incomM/comO* (CM compatibility) and *comM/comO* vs. *comM/incomO* (CO compatibility).

Activation specific for the contrast *comM/comO* vs. *incomM/comO* (CM compatibility) was found in the left superior PMv, left SMG, and left IPS. Activation specific for the contrast *comM/comO* vs. *comM/incomO* (CO compatibility) was found in bilateral parahippocampal and fusiform gyrus and right SMG. An overlap

was established in bilateral OTC, bilateral PMd/SFS, and left SPL (Fig. 3A). Activation of PMd/SFS and SPL resembled activation found in imaging studies on selective attention (Cristescu and Nobre, 2008; Fink et al., 1997; Nobre, 2001; Wager et al., 2004; Yantis and Serences, 2003).

3.2.5. Actions in neutral contexts

We speculated that common activation of PMd/SFS, SPL and OTC was due contextual relatedness of to-be-suppressed object/manipulation information and to-be-inferred object/manipulation information (Fig. 3B and C): Inferring actions from pantomime requires (1) suppression of invalid objects and their associated manipulations, and (2) activation of manipulation-congruent actions and associated objects. In the condition comM/comO both manipulation and objects are context-compatible, i.e. they belong to the same contextual category. In the conditions comM/incomO and incomM/comO, manipulations and objects belong to different contextual categories. Therefore, in the former condition suppression and activation within the same contextual category may have caused stronger conflict than in the latter two contrasts where contextual associations between relevant and irrelevant manipulation and object information were weaker (potentially explaining common activation in the OTC). Manipulation-manipulation (M-M) and object-object (O-O) conflict might result in increased demands in focusing attention on relevant manipulation and object information while suppressing irrelevant manipulation and object information (potentially explaining common activation of the dorsal attention network).

If this explanation is true a similar activation pattern should emerge if visual context information is absent, but the contextual *relationships* between objects and manipulations remain the same (Fig. 4A). We tested this hypothesis by analyzing pantomime actions in neutral context. In this condition, pantomime actions involved objects that were either from the same or a different contextual category. We therefore divided the neutral condition into two subsets: Pantomime actions of the first subset involved objects from the same contextual category as the manipulation (sameCC hereafter), pantomime actions of the other subset involved objects from a different contextual category as the manipulation (diffCC hereafter).

The according contrast sameCC vs. diffCC revealed activation in the bilateral dorsal attention network (bilateral PMd/SFS and bilateral SPL) and bilateral OTC. Further activation was found in the right SMA, right paracentral gyrus, bilateral IPS, and right anterior MTG (Fig 4B).

3.2.5. Overlap of activation due to contextual category effects

To test whether the activation pattern of the contrast contrast sameCC vs. diffCC overlapped with the activation pattern of the contrasts comM/comO vs. incomM/comO, comM/comO vs. comM/incomO we computed a conjunction of the three contrasts. The conjunction revealed common activation in left SPL, left PMd/SFS, and bilateral OTC (Fig. 4C).

4. DISCUSSION

The present experiment investigated the relationship of context, object, and manipulation information during action recognition. We employed pantomime action to decouple actions from their associated objects. Thereby, it was possible to disentangle effects of CM and CO compatibility.

No main effects with respect to CM and CO compatibility were found. Instead, differential activation was found in (1) left PMv, aIPS and SMA for CM compatibility when objects were also compatible with the context and (2) the fusiform component of the LOC for CO compatibility when actions were also compatible with the context. Common activation for both contrasts was found in bilateral OTC and bilateral PMd/SFS-SPL loop. Crucially, this common network was also found when analyzing pantomime actions in neutral settings if objects from either the same or a different contextual category as the manipulation were used. The common activation pattern might reflect increased demands to simultaneously (1) suppress perceived invalid objects and associated manipulation information and (2) activate objects implied by the manipulation, if manipulation and objects belong to the same contextual category. Together, the findings indicate that contextual information affects object and manipulation information in an interactive way: Activation is strongest in either manipulation-associated or object-associated brain regions dependent on reinforced CM and CO associations in the COM triad. Differential activation for CM and CO compatibility clearly demonstrates that contextual information, although task-irrelevant, is integrated into the analysis of actions. A similar activation in the absence of visual context suggests strengthened associations between contextually related vs. unrelated manipulation and object information. Finally, when conflicting information is contextually related to relevant information, orienting attention is particularly demanding, signified by increased activation of the dorsal attention network.

4.1. Differential activation of the motor system and the LOC

CM compatibility effects were present only when objects were also context-compatible in both contrast conditions. Activation was found in bilateral superior PMv, left SMG/aIPS, SMA, and bilateral OTC. These regions form a network that is typically active during the perception of action (Caspers et al., 2010).

The extrastriate OTC comprises several functional areas, e.g. the EBA, hMT, and pSTS, i.e. regions involved in the perceptual analysis of manipulation information (Beauchamp et al., 2002; Jastorff and Orban, 2009; Kable and Chatterjee, 2006; Thompson and Parasuraman, 2011). According to Giese & Poggio (2003), information about form, e.g. postures of body parts and shape of objects, is processed by the form pathway in V1, V4, EBA, and finally pSTS where so-called snapshot patterns are combined with optic flow patterns detected by the motion pathway, i.e. V1, V2, hMT, and MST.

Predictive accounts interpret activation of the premotor-parietal network during action perception as reflecting the generation of internal models to anticipate forthcoming events (Csibra, 2007; Jacob, 2009; Schubotz, 2007). The SMA is involved in temporally sequencing of action steps (Mita et al., 2009; Tanji, 1994). The SMA is connected to the PMv (Luppino et al., 1990), thereby providing signals to update the generation of the internal action models (Fagg and Arbib, 1998; Schubotz, 2007). PMv and aIPS are reciprocally connected (Geyer et al., 2000). PMv thus receives perceptual updates of the ongoing action and sending predictions of visual consequences of the internally generated action model (Keysers and Perrett, 2004;

Kilner et al., 2007). Activation of these regions thus likely reflects increased demands on internal modeling of actions.

To understand why the neural basis of action perception shows an enhanced response particularly when both manipulation and objects are compatible with the context, we need to reconsider the COM triad: In upper left cells of Figure 1A and B both the presented manipulation (opening a tin) as well as the used objects (whisk and bowl) are associated with the visible context (kitchen). Thus, while the manipulation and the object are incompatible, they belong to the same contextual category. In the upper right cells of Figure 1A and B only the objects (pliers and cable) are associated with the context (workshop), whereas the manipulation (opening a tin) is not. Here, manipulation and object belong to a different contextual category. In both cases the observed manipulation interferes with the object-implied manipulation. However, in the former constellation, interfering representations are contextually related, resulting in a stronger conflict and hence an increased load on the motor system. This hemodynamic effect shows that both context and objects function as primers of manipulation classes.

Parallel to CM effects reported above, CO compatibility effects showed only when manipulations were also context-compatible in both contrast conditions, Activation was found in the lateral occipital complex (LOC), comprised of a lateral occipito-temporal component (LO) and a component in posterior and mid fusiform gyrus extending into the occipito-temporal sulcus (pFs; Grill-Spector et al., 2001). The LOC plays a central role in object recognition (Grill-Spector et al., 2001; Kourtzi and Kanwisher, 2001). Both of its subdivisions are sensitive to object shape, irrespective of low-level visual features such as color or texture (Grill-Spector et al., 1999). The LOC, particularly LO, is also involved in representation of semantic object knowledge

(Binder et al., 2009). With respect to manipulable objects, LO overlaps with the middle temporal gyrus tool area (MTG-TA), which is sensitive to visual perception of tools (Chao et al., 1999; Downing et al., 2006) and to rigid tool motion (Beauchamp et al., 2003).

During identification of the pantomimed action, invalid object information is suppressed. Simultaneously, object information matching hand posture (i.e. object shape) and hand movements (i.e. tool motion) is activated. Attention is therefore selectively focused on object information implied by hand posture and movements. In a similar manner as described above, there is hence a conflict of perceptually activated and manipulation-implied object information. When both manipulation and objects are compatible with the context both perceptually activated and manipulation-implied object information belong to the same contextual category. In conclusion, increased activity in LOC likely reflects stronger interference of concurring object representations when objects and manipulations belong to the same vs. a different contextual category.

4.2. Contextual categories and the dorsal attention network

For both CM and CO compatibility, neural activity was increased in bilateral PMd/SFS and SPL. These regions are reciprocally connected with each other (via SLF I; Geyer et al., 2000) and form the so-called dorsal attention network (Corbetta and Shulman, 2002, 2011). Notably, the SPL is also connected to the parahippocampal place area (Schmahmann and Pandya, 2006), a region involved in the recognition of places (Epstein and Kanwisher, 1998) and the encoding of contextual associations (Bar and Aminoff, 2003; but see Epstein and Ward, 2010). The dorsal attention

network is involved in directing attention to a target and in selecting appropriate responses (e.g. limb or eye movements; Corbetta and Shulman, 2002). According to the biased competition model, a top-down signal, determined by task demands, biases among mutually inhibitory sensory representations, so that an attended cue 'wins' the competition by producing more robust cortical activity than unattended cues (Desimone and Duncan, 1995). Parietal cortices are thought to act as priority maps (Bisley and Goldberg, 2010) that integrate bottom-up visual stimuli and top-down influences provided by corresponding frontal regions. Locations or objects with high priority are represented with greater neural activity, thereby guiding attention and appropriate behavioral responses, e.g. eye or limb movements (Bisley and Goldberg, 2010).

During identification of the pantomimed action, attention was directed to hand postures and movements and associated object information in semantic memory. This demands for object- or feature-based (Liu et al., 2003; Yantis and Serences, 2003) and semantic-based attention (Cristescu et al., 2006). We propose that orienting attention towards hand posture and movements and associated object information in semantic memory is more demanding when manipulation and object information share a common contextual category. This is the case when both object and manipulation are compatible with the context. On the other hand, when (i) only objects, but not the manipulation, are compatible with the context, or (ii) only the manipulation, but not objects, is compatible with the context, orienting attention is less demanding. Note that in both (i) and (ii), same vs. different contextual categories are contrasted (see Fig. 3B and C). However, in (i) there is no manipulation bias by context information (hence increasing inference demands on manipulation within the COM triad), and in (ii) the

objects were unbiased by context information (hence increasing inference demands on objects).

4.3. Contextual category effects in absence of context information

The proposed mechanism should yield similar neural responses when no bias by visual context information is introduced. We therefore analyzed pantomime actions in neutral conditions. Half of the actions involved objects from the same contextual category as the manipulation, the other half involved objects from a different contextual category. Contrasting these kinds of actions revealed activation in bilateral OTC, PMd/SFS, and SPL/IPS. The activation pattern closely overlapped with the dorsal attention network and OTC activation found for CM and CO compatibility. This result supports the assumption that the contextual compatibility effects reported above reflect a suppression/activation conflict, which is stronger when to-be-suppressed and to-be-activated information belongs to the same vs. different contextual categories.

The observed effects for same vs. different contextual categories resemble to some degree semantic distance effects during picture-word interference, i.e. subjects are slower in naming pictures of objects when distracter words with related meanings are presented than if distracter words have unrelated meanings (Glaser and Dungalhoff, 1984). Similar findings were obtained when categorically related vs. unrelated body, hand, and mouth actions were to be named (Vigliocco et al., 2002). These and other studies provide support for the idea that objects and actions are represented in categorical neighborhoods in semantic memory and share a greater number of perceptual or conceptual features if they belong to the same contextual

category than if they belong to a different contextual category (Caramazza and Shelton, 1998; Harris and Little, 2010; Vigliocco et al., 2004). To our knowledge this is the first study providing evidence for a fine-grained dissociability of category-specific items within the domains hand actions and tools, i.e. kitchen, workshop, and office actions and tools. The present findings hence corroborate and extend previous findings on category-specificity in semantic memory.

REFERENCES

- Bar, M., 2004. Visual objects in context. *Nat Rev Neurosci* 5, 617-629.
- Bar, M., Aminoff, E., 2003. Cortical analysis of visual context. *Neuron* 38, 347-358.
- 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, 149-159.
- Beauchamp, M.S., Lee, K.E., Haxby, J.V., Martin, A., 2003. FMRI responses to video and point-light displays of moving humans and manipulable objects. *J Cogn Neurosci* 15, 991-1001.
- 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. *Cereb Cortex* 19, 2767-2796.
- Bisley, J.W., Goldberg, M.E., 2010. Attention, intention, and priority in the parietal lobe. *Annu Rev Neurosci* 33, 1-21.
- Caramazza, A., Shelton, J.R., 1998. Domain-specific knowledge systems in the brain the animate-inanimate distinction. *J Cogn Neurosci* 10, 1-34.
- 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, 1148-1167.
- Chao, L.L., Haxby, J.V., Martin, A., 1999. Attribute-based neural substrates in temporal cortex for perceiving and knowing about objects. *Nat Neurosci* 2, 913-919.
- Corbetta, M., Shulman, G.L., 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3, 201-215.
- Corbetta, M., Shulman, G.L., 2011. Spatial neglect and attention networks. *Annu Rev Neurosci* 34, 569-599.
- Cristescu, T.C., Devlin, J.T., Nobre, A.C., 2006. Orienting attention to semantic categories. *Neuroimage* 33, 1178-1187.
- Cristescu, T.C., Nobre, A.C., 2008. Differential modulation of word recognition by semantic and spatial orienting of attention. *J Cogn Neurosci* 20, 787-801.
- 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.
- Desimone, R., Duncan, J., 1995. Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18, 193-222.
- Downing, P.E., Chan, A.W., Peelen, M.V., Dodds, C.M., Kanwisher, N., 2006. Domain specificity in visual cortex. *Cereb Cortex* 16, 1453-1461.
- Epstein, R., Kanwisher, N., 1998. A cortical representation of the local visual environment. *Nature* 392, 598-601.

- Epstein, R.A., Ward, E.J., 2010. How reliable are visual context effects in the parahippocampal place area? *Cereb Cortex* 20, 294-303.
- Fagg, A.H., Arbib, M.A., 1998. Modeling parietal-premotor interactions in primate control of grasping. *Neural Netw* 11, 1277-1303.
- Fink, G.R., Dolan, R.J., Halligan, P.W., Marshall, J.C., Frith, C.D., 1997. Space-based and object-based visual attention: shared and specific neural domains. *Brain* 120 (Pt 11), 2013-2028.
- Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D., Frackowiak, R.S., 1995. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping* 2, 189-210.
- Geyer, S., Matelli, M., Luppino, G., Zilles, K., 2000. Functional neuroanatomy of the primate isocortical motor system. *Anat Embryol (Berl)* 202, 443-474.
- Giese, M.A., Poggio, T., 2003. Neural mechanisms for the recognition of biological movements. *Nat Rev Neurosci* 4, 179-192.
- Glaser, W.R., Dungelhoff, F.J., 1984. The time course of picture-word interference. *J Exp Psychol Hum Percept Perform* 10, 640-654.
- Grill-Spector, K., Kourtzi, Z., Kanwisher, N., 2001. The lateral occipital complex and its role in object recognition. *Vision Res* 41, 1409-1422.
- Grill-Spector, K., Kushnir, T., Edelman, S., Avidan, G., Itzhak, Y., Malach, R., 1999. Differential processing of objects under various viewing conditions in the human lateral occipital complex. *Neuron* 24, 187-203.
- Harris, I.M., Little, M.J., 2010. Priming the semantic neighbourhood during the attentional blink. *PLoS One* 5, e12645.
- Jacob, P., 2009. The tuning-fork model of human social cognition: a critique. *Conscious Cogn* 18, 229-243.
- Jastorff, J., Orban, G.A., 2009. Human functional magnetic resonance imaging reveals separation and integration of shape and motion cues in biological motion processing. *J Neurosci* 29, 7315-7329.
- Kable, J.W., Chatterjee, A., 2006. Specificity of action representations in the lateral occipitotemporal cortex. *J Cogn Neurosci* 18, 1498-1517.
- Keysers, C., Perrett, D.I., 2004. Demystifying social cognition: a Hebbian perspective. *Trends Cogn Sci* 8, 501-507.
- Kilner, J.M., Friston, K.J., Frith, C.D., 2007. Predictive coding: an account of the mirror neuron system. *Cogn Process* 8, 159-166.
- Kourtzi, Z., Kanwisher, N., 2001. Representation of perceived object shape by the human lateral occipital complex. *Science* 293, 1506-1509.
- Liu, T., Slotnick, S.D., Serences, J.T., Yantis, S., 2003. Cortical mechanisms of feature-based attentional control. *Cereb Cortex* 13, 1334-1343.
- Lohmann, G., Muller, K., Bosch, V., Mentzel, H., Hessler, S., Chen, L., Zysset, S., von Cramon, D.Y., 2001. LIPSIA--a new software system for the evaluation of functional magnetic resonance images of the human brain. *Comput Med Imaging Graph* 25, 449-457.

- Luppino, G., Matelli, M., Rizzolatti, G., 1990. Cortico-cortical connections of two electrophysiologically identified arm representations in the mesial agranular frontal cortex. *Exp Brain Res* 82, 214-218.
- Mita, A., Mushiake, H., Shima, K., Matsuzaka, Y., Tanji, J., 2009. Interval time coding by neurons in the presupplementary and supplementary motor areas. *Nat Neurosci* 12, 502-507.
- Nee, D.E., Wager, T.D., Jonides, J., 2007. Interference resolution: insights from a meta-analysis of neuroimaging tasks. *Cogn Affect Behav Neurosci* 7, 1-17.
- Nichols, T., Brett, M., Andersson, J., Wager, T., Poline, J.B., 2005. Valid conjunction inference with the minimum statistic. *Neuroimage* 25, 653-660.
- Nobre, A.C., 2001. The attentive homunculus: now you see it, now you don't. *Neurosci Biobehav Rev* 25, 477-496.
- Norris, D.G., 2000. Reduced power multislice MDEFT imaging. *J Magn Reson Imaging* 11, 445-451.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97-113.
- 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 Cogn Sci* 11, 211-218.
- Schubotz, R.I., von Cramon, D.Y., 2009. The case of pretense: observing actions and inferring goals. *J Cogn Neurosci* 21, 642-653.
- Talairach, J., Tournoux, P., 1988. *Co-planar Stereotaxic Atlas of the Human Brain*. New York: Thieme.
- Tanji, J., 1994. The supplementary motor area in the cerebral cortex. *Neurosci Res* 19, 251-268.
- Thompson, J., Parasuraman, R., 2011. Attention, biological motion, and action recognition. *Neuroimage*.
- Turk-Browne, N.B., Scholl, B.J., Johnson, M.K., Chun, M.M., 2010. Implicit perceptual anticipation triggered by statistical learning. *J Neurosci* 30, 11177-11187.
- Vigliocco, G., Vinson, D.P., Damian, M.F., Levelt, W., 2002. Semantic distance effects on object and action naming. *Cognition* 85, B61-69.
- Vigliocco, G., Vinson, D.P., Lewis, W., Garrett, M.F., 2004. Representing the meanings of object and action words: the featural and unitary semantic space hypothesis. *Cogn Psychol* 48, 422-488.
- Wager, T.D., Jonides, J., Reading, S., 2004. Neuroimaging studies of shifting attention: a meta-analysis. *Neuroimage* 22, 1679-1693.
- Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited--again. *Neuroimage* 2, 173-181.
- Wurm, M.F., Schubotz, R.I., in press. Squeezing lemons in the bathroom: contextual modulation of action recognition. *Neuroimage*.

Yantis, S., Serences, J.T., 2003. Cortical mechanisms of space-based and object-based attentional control. *Curr Opin Neurobiol* 13, 187-193.

FIGURES

Figure 1: Experimental conditions. (A) Example frames for the employed pantomime action movies. Conditions constitute a 2x2 factorial design with the factors CM and object-scene relation. (B) Schema of the triadic relationship of action (A), object (O), and scene (S) information in the according conditions. Solid lines represent strong contextual associations, dotted lines represent weak contextual associations.

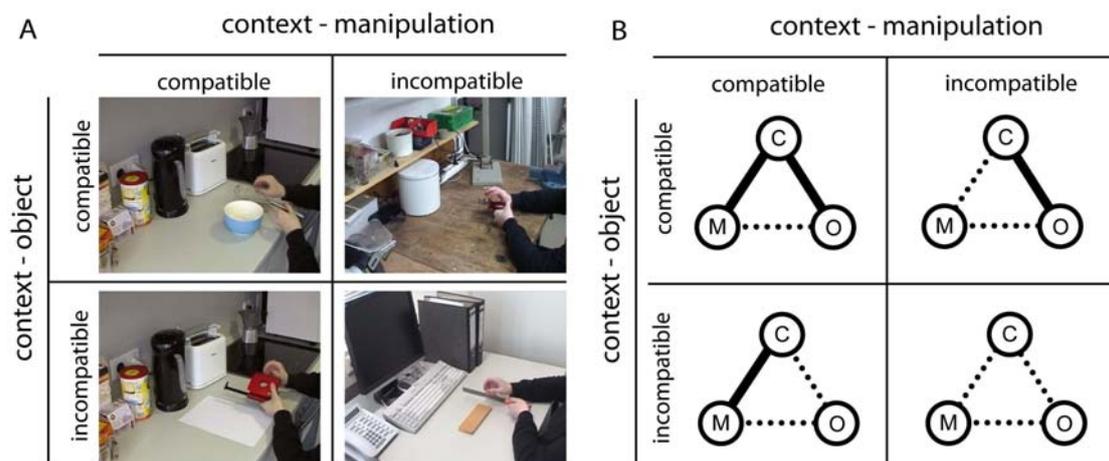


Figure 2: Context-manipulation (CM) and context-object (CO) compatibility effects. (A) CM compatibility (comM/comO vs. incomM/comO) enhanced neural responses particularly in the motor system (bilateral superior ventral premotor cortex (PMv), left supramarginal gyrus (SMG) extending into the intraparietal sulcus (IPS), supplementary motor area (SMA), and bilateral occipitotemporal cortex (OTC)). (B) CO compatibility (comM/comO vs. comM/incomO) enhanced activity particularly in the lateral occipital complex LOC (bilateral lateral occipital cortex (LO) and posterior fusiform and parahippocampal gyrus (pFs)). Contrasts are z -thresholded at 2.576; corrected cluster threshold $p < .05$.

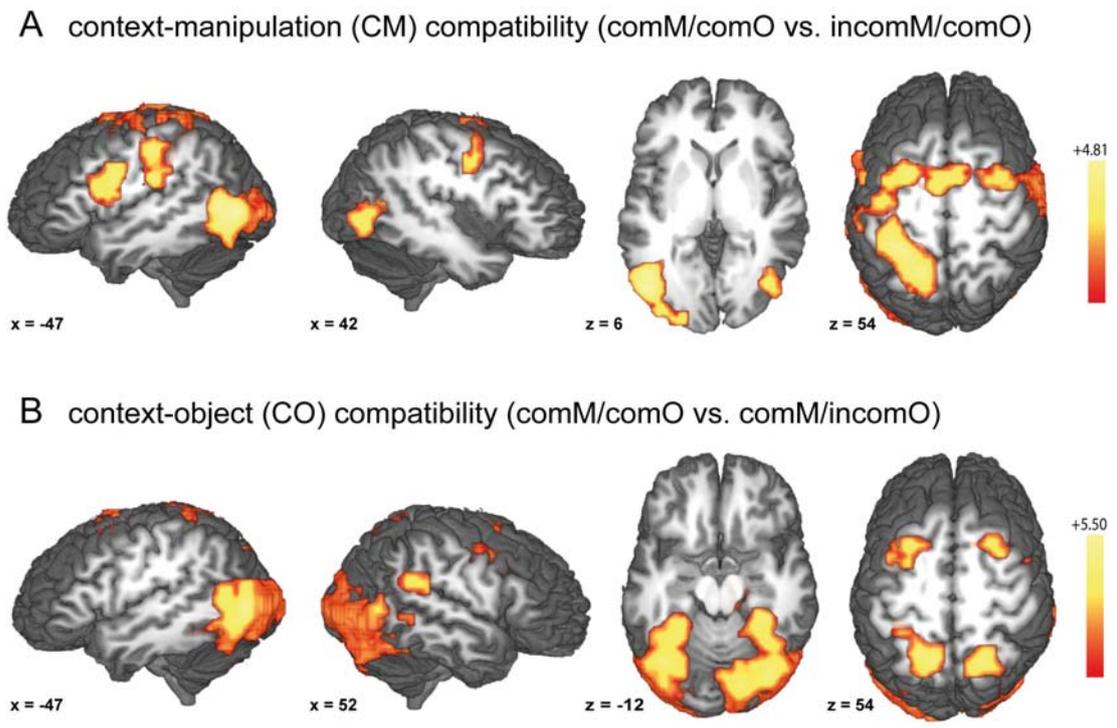


Figure 3: Influence of context on attentional focus in the COM triad. (A) Overlap of CM and CO compatibility effects (comM/comO vs. comM/incomO and comM/comO vs. incomM/comO). (B) Manipulation-manipulation M-M conflict arises when object-IMPLIED, invalid manipulation information is contextually related to the pantomimed action. (C) Object-object O-O conflict arises when perceived, invalid objects are contextually related to objects that are implied by the manipulation. Context information compatible with either manipulation or objects fixes contextual association on either the manipulation- or object-side of the triad. (Solid lines represent strong contextual associations, dotted lines represent weak contextual associations. Black circles represent present/cued context, object, or manipulation representations, grey circles represent object- or manipulation-IMPLIED (grey arrows) manipulation or object representations.)

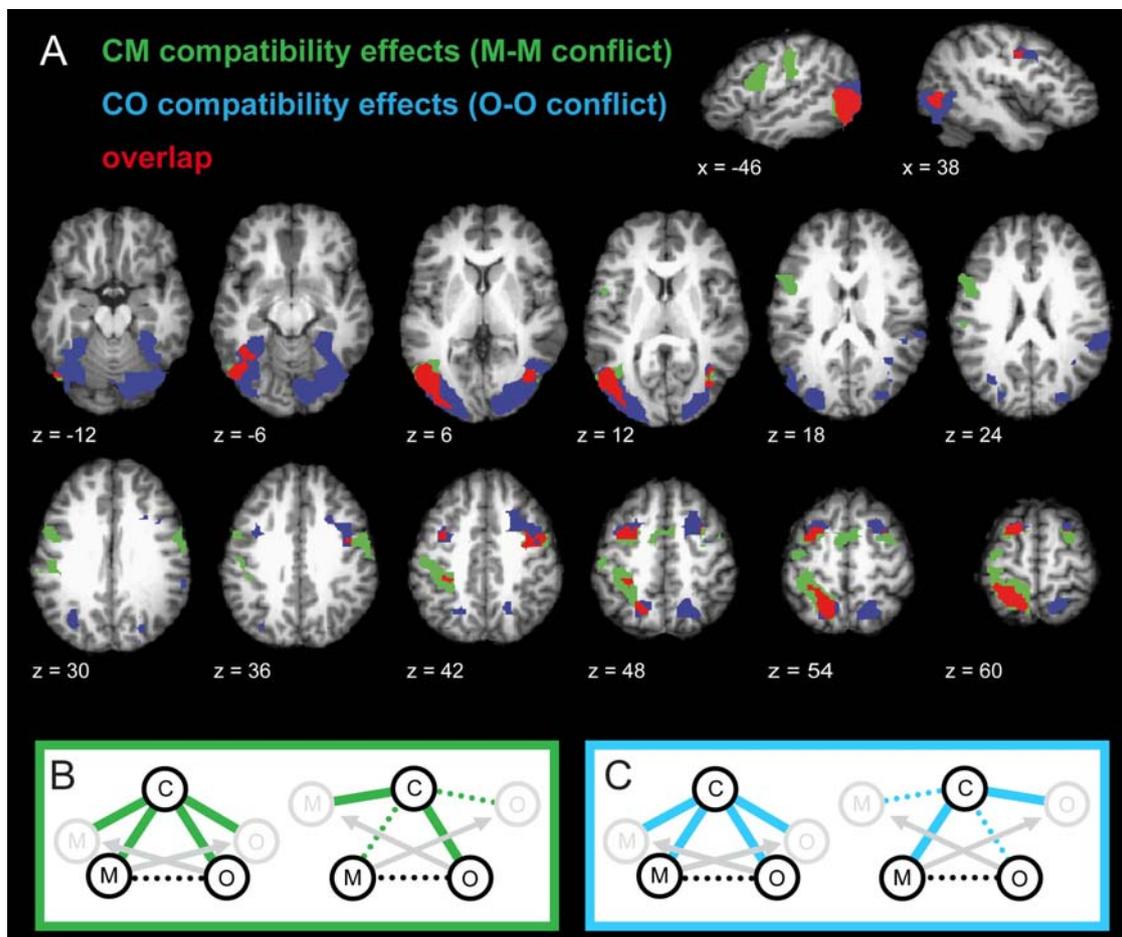
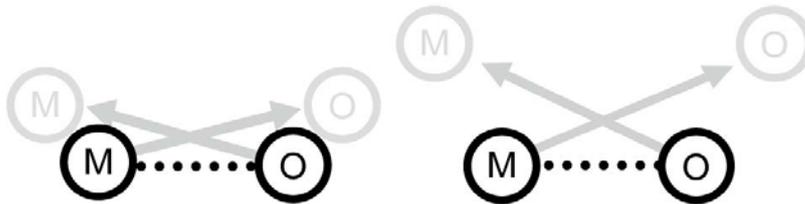
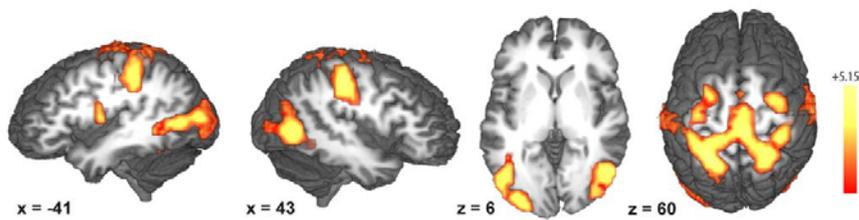


Figure 4: Same vs. different contextual category. (A) Schematic of actions in neutral settings employing objects of same vs. different contextual category (sameCC vs. diffCC). (C) conjunction of contrasts employing objects of same vs. different contextual category ((comM/comO vs. comM/incomO) \cap (comM/comO vs. incomM/comO) \cap (sameCC vs. diffCC)). Contrasts are z -thresholded at 2.576; corrected cluster threshold $p < .05$.

A same vs. different contextual category (no visual context)



B same vs. different contextual category (neutral condition)



C conjunction same vs. different contextual category

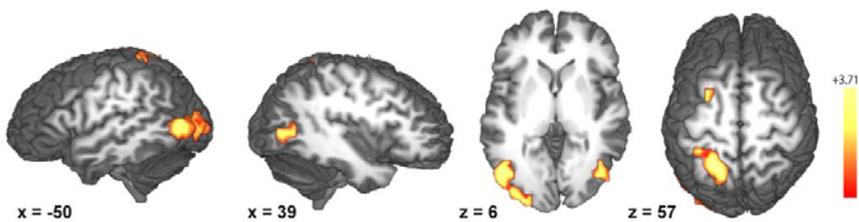


Table 1: Activations for CM and CO compatibility and same vs. different contextual categories

Area	BA	<i>x</i>	<i>y</i>	<i>z</i>	<i>Z</i>	mm ³
<i>CM compatibility (comM/comO vs. incomM/comO)</i>						
L pMTG/ITG (OTC)	18/19/37	-44	-72	6	4.79	13203
L		34	-69	6	3.46	2025
R superior PMv	44/6	-47	0	24	3.75	4266
R aIPS/SMG	40	-29	-42	48	4.95	19791
L SPL	7	-20	-54	63	4.59	l.m.
R PMd/SFS	6/8	-20	3	63	4.01	7074
L PMv		52	-6	35	4.06	5805
R		25	-3	60	3.81	l.m.
R SMA	6	-2	0	54	3.34	810
<i>CO compatibility (comM/comO vs. comM/incomO)</i>						
L pMTG/ITG (OTC)	18/19/37	-38	-75	3	5.63	32130
L fusiform gyrus	37	-29	-48	-3	4.11	l.m.
R pMTG/ITG	18/19/37	19	-87	6	5.67	35208
R fusiform gyrus	37	28	-36	-9	4.12	l.m.
L PMd/SFS	6/8	-32	0	36	4.21	5805
R		22	15	45	4.30	8100
L SPL	7	-17	-54	51	4.47	7155
R		16	-60	48	4.25	4995
R SMG	40	58	-36	21	4.25	2025
<i>actions in neutral context with objects from same vs. different contextual category (sameCC vs. diffCC)</i>						
L pMTG	18/19	-44	-81	12	3.61	5913
R		37	-69	22	3.66	4266
L SPL	7	-23	-51	60	4.64	21330
L IPS	40	-56	-24	39	3.91	l.m.
L PMd	6	-23	-3	51	3.54	l.m.
R		34	-24	42	5.14	26190
R SPL	7	10	-41	69	3.84	l.m.
paracentral gyrus	4	4	-33	63	4.21	l.m.
R		28	-9	51	3.58	l.m.
SMA	6	4	-3	51	3.45	l.m.

Hemisphere (L, R), Macroanatomical specification, Brodmann area (BA), Talairach coordinates (*x*, *y*, *z*), maximal *Z* scores (*Z*), and cluster volumes (mm³); *z*-thresholded at 2.576, corrected cluster threshold $p < .05$. Abbreviations: (*a*)*IPS* (anterior) intraparietal sulcus, *ITG* inferior temporal gyrus, *OTC* occipitotemporal cortex, *PMd* dorsal premotor

cortex, *pMTG* posterior middle temporal gyrus, *PMv* ventral premotor cortex, *SMA* supplementary motor area, *SMG* supramarginal gyrus, *SFS* superior frontal sulcus, *SPL* superior parietal lobe.

Curriculum Vitae

Name: Moritz Franz Nikolaus Wurm

Date of birth: 11/11/1981

Place of birth: Cologne, Germany

- 2011 Research assistant, Westfälische Wilhelms-Universität Münster, Germany (supervisor: Prof. Ricarda I. Schubotz)
- 2008-2011 Ph.D. student, Max Planck Institute for neurological research, Cologne, Germany (Motor Cognition Group, supervisor: Prof. Ricarda I. Schubotz)
- 2007 Diploma in Biology, University of Cologne, Germany (grade 1.0, supervisor: Prof. Diethard Tautz)
- 2007 Diploma student, Dept. of Evolutionary Developmental Biology, University of Cologne, Germany
- 2006 – 2008 Student assistant, Dept. of Developmental Genetics, University of Cologne (Advisor: Prof. Maria Leptin)
- 2004 Teaching assistant, Dept. of Botany, University of Cologne, Germany
- 2004 Teaching assistant, Dept. of Zoology, University of Cologne, Germany
- 2003 Student assistant, Dept. of Botany, University of Cologne, Germany (Advisor: Prof. Michael Melkonian)
- 2001 – 2006 Studies in Biology, University of Cologne, Germany
- 2001 Final secondary school examinations (Abitur), Gymnasium Rodenkirchen, Cologne, Germany

Eidesstattliche Versicherungen

Hiermit versichere ich, Moritz Wurm, 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, da ich hauptverantwortlich zur Entwicklung der Fragestellung und des experimentellen Designs, zur Datenerhebung und Datenauswertung, sowie zur Interpretation und Verfassung der hier genannten Manuskripte beteiligt war:
 - Wurm, M.F., von Cramon, D.Y., Schubotz, R.I. (2011). Do we mind other minds when we mind other minds' actions? an fMRI study. Human Brain Mapping, doi: 10.1002/hbm.21176. [Epub ahead of print].
 - Wurm, M.F., Schubotz, R.I. (2011). Squeezing lemons in the bathroom: contextual information modulates action recognition. Neuroimage, doi: 10.1016/j.neuroimage.-2011.08.038 [Epub ahead of print].
 - Wurm, M.F., von Cramon, D.Y., Schubotz, R.I. (in preparation). The context-object-manipulation (COM) triad: cross talk during action perception revealed by fMRI.

Köln, den _____

Unterschrift