In the Beginning was the Deed: From Sensorimotor Interactions to Integrative Spatial Encodings

Dissertation
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Dipl.-Psych.
Johannes Lohmann
aus Dresden

Tübingen
2017

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 14.07.2017

Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Martin V. Butz
2. Berichterstatter:	Prof. Dr. Bettina Rolke

Contents

Abstract

To change the environment in a goal-directed manner, it is necessary to generate reliable predictions regarding action-outcomes. These predictions are derived from models, which are formed, evaluated, and maintained during sensorimotor interactions with the environment. The *free energy principle* and the theories of *event segmentation* and *embodied cognition* allow to derive hypotheses regarding the identification and application of these predictive models. According to the free energy principle, cognitive systems constantly try to infer the causes of perceived sensations. This results in the formation of predictive models based on sensorimotor experience. While the free energy principle remains vague with respect to the underlying representational format, I will argue in favor of a integrative spatial code, relating different modalities in an abstract representation. The integration process is assumed to be biased towards behaviorally relevant modalities. Moreover, a striving for consistency is assumed to maintain unambiguous states. Besides the representational format, the prediction process itself is of central interest. According to the event segmentation theory, cognitive systems segment the stream of sensorimotor information along significant changes, so-called event boundaries. Hence, it seems likely that predictions are carried out in terms of a simulation of the next, desired event boundary within the proposed integrative spatial code. The spatial code might support mental simulation in general, providing sensorimotor grounding to higher cognitive functions – as proposed by theories of embodied cognition.

The presented experiments aimed at evaluating these assumptions. More precisely, the presented work covers the questions (i) if multisensory integration is biased towards the most behaviorally relevant modality, (ii) if and how representational consistency is preserved in the in face of conflicting sensory information, (iii) whether

anticipatory behavior control is realized in terms of an event-oriented prediction, and (iv) if mental rotation is realized by means of a spatial simulation. The obtained results show a goal-directed, biased integration of multisensory information, yielding a percept that is dominated by the most behaviorally relevant modality. Apart from this bias in favor of behaviorally relevant information, the results show how consistency in spatial representations is maintained by adapting multiple frames of reference. Furthermore, the results provided evidence for an event-like prediction, engaged in anticipatory behavior control. Finally, results regarding mental rotation yielded compatibility effects between mental rotation, concurrent rotational stimulation, and motor responses.

The obtained results confirm the assumptions regarding the proposed integrative spatial code. The combination of the free energy principle and the theory of event segmentation seems a viable approach to account for the emergence of an predictive, integrative spatial code from sensorimotor interactions. The results allow the derivation of design principles for an artificial spatial reasoning system and the developed experimental paradigms allow further investigations of the causal role of spatial models in higher cognitive functions.

Zusammenfassung

Eine Grundvoraussetzung für zielgerichtetes Verhalten ist die Fähigkeit, die sensorischen Konsequenzen von motorischen Aktionen vorherzusagen. Diese Vorhersagen beruhen vermutlich auf internen, prädiktiven Modellen die durch sensomotorische Interaktionen zwischen Organismus und Umwelt geformt und evaluiert werden. Vorhersagen bezüglich der Identifikation und der Arbeitsweise dieser Modelle lassen sich einerseits aus dem *free energy princinple*, andererseits aus der *event segmentation* Theorie und der Theorie der *embodied cognition* ableiten. Das free energy principle postuliert, dass jedes kognitive Systeme versucht die Ursachen der sensorischen Zustände die es erfährt, zu erschließen. Dieser Inferenzprozess führt zur Generierung von internen, prädiktiven Modellen, basierend auf den erfahrenen sensomotorischen Zusammenhängen. Bezüglich der Repräsentation, die diesen Modellen zugrunde liegt, trifft das free energy principle keine Aussage. Basierend auf verhaltensexperimentellen und neurophysiologischen Befunden erscheint mir eine räumliche Kodierung plausibel, die verschiedene modale Kodierungen in einem abstrakten, räumlichen Format vereint. Weiterhin nehme ich an, dass in diesem Integrationsprozess verhaltensrelevante Stimuli bevorzugt verarbeiten werden und ein globaler Konsistenzmechanismus für eindeutige Repräsentationen sorgt. Neben der Art der Repräsentation ist die Natur des Vorhersageprozesses selbst von Interesse. Laut der event segmentation Theorie unterteilen kognitive Systeme den kontinuierlichen sensomotorischen Datenstrom anhand von signifikanten Änderungen im Aktivitätsmuster, sogenannten event boundaries. Daher erscheint es mir wahrscheinlich, dass Vorhersagen von Verhaltenskonsequenzen ebenfalls bezüglich signifikanter sensorischer Änderungen getroffen werden. Falls die Annahme eines gemeinsamen räumlichen Formats für die internen Modelle zutrifft, sollten diese Vorhersagen

ebenfalls primär auf räumliche Informationen rekurrieren. Möglicherweise ist dieser Vorhersagemechanismus nicht auf die Verhaltenskontrolle beschränkt sondern bildet eine generelle Grundlage für höhere kognitive Funktionen, entsprechend der Theorie der embodied cognition.

Die in dieser Arbeit präsentierten Experimente zielten auf die Überprüfung dieser Annahmen ab. Im Detail wurde untersucht (i) ob sich bei multisensorischer Integration tatsächlich eine Dominanz von verhaltensrelevanten Reizen zeigt, (ii) ob sich Evidenz für den postulierten Konsistenzmechanismus im Fall von sensorischem Konflikt finden lässt, (iii) ob antizipative Verhaltenskontrolle auf der Vorhersage von event boundaries beruht und (iv) inwiefern ein räumlicher Vorhersagemechanismus in die Realisation einer höheren kognitiven Funktion, in diesem Fall mentaler Rotation, involviert ist. Die Annahmen wurden weitestgehend bestätigt. Die Ergebnisse implizieren eine Bevorzugung verhaltensrelevanter Stimuli bei multisensorischer Integration, ebenso wurden Hinweise für den angenommenen Konsistenzmechanismus gefunden. Die Untersuchungen zur antizipativen Verhaltenskontrolle bestätigten die Annahme eines Vorhersageprozesses, der auf die nächste event boundary ausgerichtet ist. In den Ergebnissen zur mentalen Rotation zeigten sich Kompatibilitätseffekte die sich im Sinne einer räumlichen Simulation erklären lassen.

Die Befunde sprechen allgemein für die angenommene, integrative räumliche Kodierung und ihre Verwendung in prädiktiven Modellen. Basierend auf dem free energy principle und der event segmentation Theorie lässt sich die Identifikation und Anwendung von prädiktiven räumlichen Kodierungen beschreiben. Weiterhin erlauben die Ergebnisse Vorschläge für die Implementierung eines artifiziellen Systems abzuleiten, welches in der Lage ist, rudimentäres räumliches Denken zu simulieren. Die entwickelte Methodik ermöglicht die weitere Untersuchung der kausalen Rolle, die prädiktive räumliche Kodierungen möglicherweise für die Realisierung von höheren kognitive Funktionen spielen.

1. Publications

1.1. Accepted papers

Schroeder, P. A., Lohmann, J., Butz, M. V., & Plewnia, C. (2015): **Behavioral bias for food reflected in hand movements: A preliminary study with healthy subjects**. *Cyberpsychology, Behavior, and Social Networking*. 19(2), 120-126.

Lohmann, J., Gütschow, J., & Butz, M. V. (2017): **Grasping Multisensory Integration: Proprioceptive Capture after Virtual Object Interactions**. Proceedings of the 39th Annual Meeting of the Cognitive Science Society. London, UK: Cognitive Science Society.

Lohmann, J., & Butz, M. V. (2017): **Lost in Space: Multisensory Conflict yields Adaptation in Spatial Representations across Frames of Reference**. *Cognitive Processing*, 1-18. doi: 10.1007/s10339-017-0798-5.

Lohmann, J., Rolke, B., & Butz, M. V. (2017): **In touch with mental rotation: Interactions between mental and tactile rotations and motor responses**. *Experimental Brain Research*, 235(4), 1063-1079. doi: 10.1007/s00221-016-4861-8.

1.2. Submitted manuscripts

Belardinelli, A., Lohmann, J., Farnè, A., & Butz, M. V.: **Mental Space Maps Into the Future**. *under review*.

1.3. Further Publications

Lohmann, J., & Butz, M. V. (2011). **Learning a Neural Multimodal Body Schema: Linking Vision with Proprioception**. In B. Hammer & T. Villmann (Eds.), Workshop New Challenges in Neural Computation 2011 (pp. 53-57). University of Bielefeld, Dept. of Technology CITEC.

Koryakin, D., Lohmann, J., & Butz, M. V. (2012). **Balanced Echo State Networks**. *Neural Networks*, 36, 35-45. doi:10.1016/j.neunet.2012.08.008

Lohmann, J., Herbort, O., & Butz, M. V. (2013). **Modeling the Temporal Dynamics of Visual Working Memory**. *Cognitive Systems Research*, 24, 80-86. doi:10.1016/j.cogsys.2012.12.009

Lohmann, J., & Butz, M. V. (2013). **Modeling Continuous Representations in Visual Working Memory**. Proceedings of the 35th Annual Meeting of the Cognitive Science Society (pp. 2926-2931). Berlin: Cognitive Science Society.

Lohmann, J., & Butz, M. V. (2014). **Memory disclosed by motion: predicting visual working memory performance from movement patterns**. Proceedings of the 12th Biannual Conference of the German Cognitive Science Society (KogWis 2014), 52-53.

Belardinelli, A., Lohmann, J., & Butz, M. V. (2016). **Human-object interaction understanding without objects**. Proceedings of the 38th Annual Meeting of the Cognitive Science Society (p. 2965). Austin, TX: Cognitive Science Society.

2. Introduction

> *It is not spoken language that is natural to man, but the faculty of constituting a language, that is, a system of distinct signs.*
>
> – Ferdinand de Saussure

According to Neisser's (2014) famous definition, *cognition* refers to all processes that are involved in the transformation of information. The research program defined by Neisser aimed at understanding the structured pattern of these transformations. The question *why* such a structured pattern should emerge was not pursued within the classic framework of cognitive psychology. As some researchers pointed out, however, cognitive systems did not emerge to understand the environment, but to change it in a goal-directed manner (e.g. Glenberg, 1997; Hoffmann, 1993; Prinz, 1990). Goal-directed behavior requires reliable predictions regarding action outcomes, which are inevitably based on structural, semantic knowledge.

As de Saussure pointed out, the human cognitive system seems to be equipped with an inference mechanism that fosters the generation of structure. I will argue that the modus operandi of this inference mechanism can be understood by asking the *why* question from an action-centered perspective. Behavior control unfolds within the continuous stream of sensorimotor information. In order to establish a predictive structure within this stream of information, it is necessary (i) to segment the continuous stream of information in time and space, (ii) to detect and abstract commonalities within the segments, and (iii) to assure consistency of the acquired structure. With replication and experience, consistent structures will condense into predictive models of sensorimotor contingencies. These models convey propositional information about states, like "grasped", "touched", or "behind each other", and the according transitions, like “how far”, “how big”, or “how fast”. Hence, these models provide a structured pattern of information transformation and I assume that they form the basis for de Saussure’s *system of distinct signs*.

The model presented in Fig. 2.1 combines these assumptions. According to this model, different modal codes are related through an *integrative spatial code* (ISC in the following). This code provides a common format that allows to detect and to preserve commonalities across modal codes in a more abstract representation. Driven by a striving for consistency, the ISC is assumed to maintain stable activation patterns across different modalities. Activation within the ISC is segmented along behaviorally relevant changes, like touching or grasping an object. Repeated experience of these changes allows to develop associations between these states and the experienced consequences. Inverse activation of these consequences, in terms of a desired goal state, is assumed to activate the ISC, yielding a prediction of the sensorimotor changes necessary to realize the goal state. Hence, the ISC can serve as a predictive code, translating a desired goal state into a sensorimotor prediction. The outlined structure is hierarchical, stretching from discrete goal states to continuous predictions of the stream of sensorimotor activation. Due to its assumed spatial format, the ISC naturally preserves spatial relations experienced through sensorimotor interactions.

The outlined approach can provide an explanation how propositional, essentially spatial knowledge arise from a general inference mechanism, focusing on sensorimotor interaction. The experiments presented in this work revolve around the investigation of some central assumptions regarding the ISC and its predictive properties. More precisely, the presented work covers the questions (i) how multisensory integration is biased towards the most behaviorally relevant modality, (ii) how consistency of spatial body representations is preserved in face of conflicting sensory information, (iii) how predictive models provide an anticipatory event-structure, which guides action, and (iv) how mental rotation is realized in terms of an abstract spatial transformation code. Implications for an artificial spatial problem solver are discussed, as well as the prospects of the developed experimental setups for the investigation of the relationship between the ISC and spatial reasoning. In the next sections, I will provide the theoretical background for the proposed predictive ISC.

2.1. Structure through Interaction

The recent version of an action-centered perspective on cognition is referred to as *embodied cognition* (Barsalou, 2008; Glenberg, 2010; Glenberg, Witt, & Metcalfe, 2013). According to this perspective, cognitive functions are deeply rooted in the sensorimotor system. Among other issues with the classic approach of cognitive psychology, the question how meaning is conveyed by symbols was one of the reasons for the development of the embodied perspective. According to the classic approach, meaning emerges from context, that is, symbols convey meaning because they are embedded in a semantic network. Especially Harnad (1990) showed that this approach is not sufficient and some kind of grounding is necessary. Lakoff and Johnson (1980a, 1980b) proposed that sensorimotor interaction can provide this grounding. For instance they related logical reasoning, like understanding the proposition “a or b”, to sensorimotor interactions with containers. According to Lakoff and Johnson the consistent way we interact with containers forms an abstract internal model, referred to as schema. The schema reflects the sensorimotor contingencies experienced during interactions, for instance, that an object is either inside a container or not. The extension of such basic schemata to abstract, but similarly structured concepts, such as the proposition “a or b”, allows to understand them in an embodied, metaphorical way. Accordingly, many studies have shown the involvement of the sensorimotor system in language understanding (Glenberg & Robertson, 2000; Glenberg & Kaschak, 2002; Kaschak & Glenberg, 2000). Especially Zwaan and Taylor (2006) could show the prominent role of motor activity in language understanding by providing evidence that interference between implied and overt movement is driven by verbs, which again highlights the role of sensorimotor knowledge in the understanding of meaning. In general, embodied cognition assumes that inference is accomplished in terms of sensorimotor simulations, which allow to generate grounded metaphors of abstract concepts. These simulations are driven by internal models based on action knowledge. This approach and the mentioned findings provide an elegant way to explain how symbols arise from

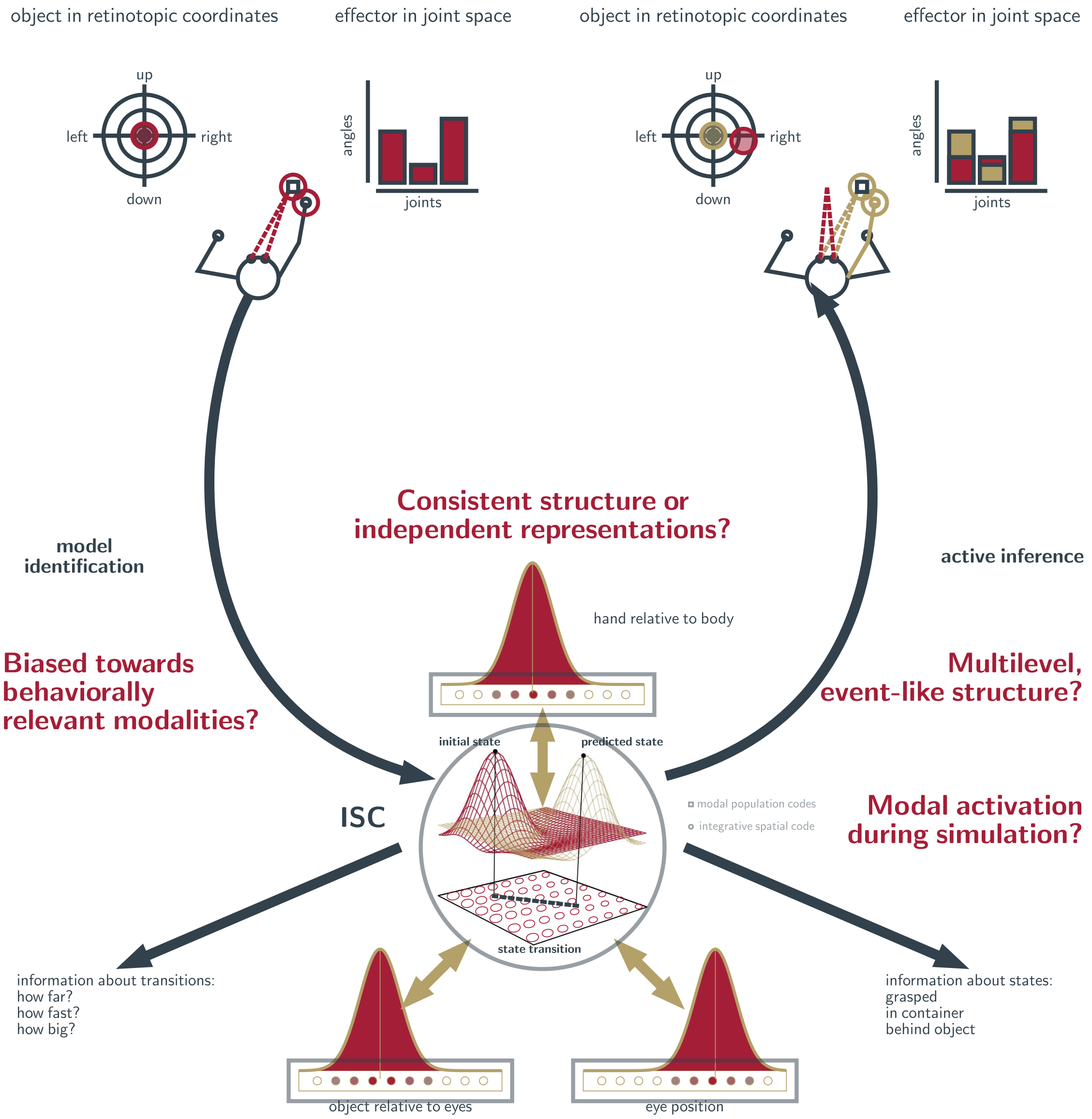


Figure 2.1.: The assumed predictive, integrative spatial code (ISC) and the main hypotheses pursued in the presented work. Sensorimotor interactions are assumed to produce a consistent event-like like structure in various modal codes and according frames of references (upper left panel). This structure is focused on the most behaviorally relevant modalities. Commonalities are preserved in an ISC which receives activation from modal codes (center). Consistency within the ISC is maintained by a striving for consistency across the involved frames of reference and modalities. Activation patterns that are associated with the realization of certain consequences, like the successful grasp of an object, are preserved in terms of predictive models. Inversely, after learning, the activation of a desired goal state is assumed to activate these models. This activation yields a trajectory through the ISC towards the desired state (center, upper right panel). Activation of the ISC is assumed to spread across frames of reference and the according modal codes, realizing a sensorimotor simulation. Hence, the model provides a hierarchical behavioral control structure, stretching from the desired goal state to the simulation of sensorimotor activation necessary to realize it.

sensorimotor experience. However, many aspects of the underlying process remain underspecified. For instance, it remains open which properties of the interactions are candidates for abstraction into models and which biases guide the formation of these models.

I argue that both questions can be answered from an action-centered perspective. If the models that are used in the sensorimotor simulations arise in the context of action control, they should be able to provide useful predictions of action outcomes. First, the predictions should be sparse and distinct, that is, they should only include relevant sensory changes at a certain point in time. Second, the predictions should be able to generalize over multiple, related situations. I assume that spatial changes associated with interactions are a suitable basis for general predictions and possible abstractions.

Cognitive systems interact in a complex and highly dynamic world, yet the perceptual system is able to extract rather stable representations and our actions can be directed towards distinct end states. According to Friston's (2010) *free energy principle*, to maintain a homeostatic equilibrium it is imperative to minimize the probability for an unexpected state transition to occur. The free energy principle provides a quantitative description of this minimization of uncertainty, which basically states that cognitive systems can suppress uncertainty by either changing sensory signals through interaction, or by adapting their internal predictive models. According to this approach, sensorimotor experience forms internal, predictive models, which are used to reduce uncertainty. Maintenance of these models requires constant evaluation through interaction. Counteracting uncertainty in predictions is essential for adaptive behavior, hence, structure formation on a sensorimotor level is a general bias in cognitive systems. Due to the two proposed mechanisms to reduce uncertainty - motor activation and model adaptation - the selection of relevant features integrated into models seems to emerge directly from task-relevance. Consequently, the sensory information that is relevant to achieve a certain goal state is preserved within the developing models. While the free energy principle allows to derive some predictions regarding the biases driving the formation of predictive models and the mechanisms of their maintenance, it remains vague with respect to the representational format of the models. Friston states that the internal models are generative, as they represent the causes of sensations. However, the nature of the involved inference process remains open. Seeing the vast amount of information conveyed by the stream of sensorimotor activations, some kind of segmentation seems necessary to draw reliable inferences. In this respect, significant changes in the stream of information seem to be most informative, where significance should be closely related to behavioral relevance. Hence, it seems intuitive that a causal inference model should focus on these changes and segment the stream of information along them. This is indeed one of the central claims of the *event segmentation theory* (Zacks & Tversky, 2001; Zacks, Speer, Swallow, Braver, & Reynolds, 2007). According to this theory, the perceptual system has evolved in a way that allows to detect *events*, which are considered segments of time at a certain location that are perceived to have a beginning and an end. The detection of events is assumed to be a basic mechanism that operates automatically. The theory is supported by various behavioral and imaging studies (e.g. Sridharan, Levitin, Chafe, Berger, & Menon, 2007; Zacks, Swallow, Vettel, & McAvoy, 2006; Zwaan, Langston, & Graesser, 1995). For instance, Speer, Zacks, and Reynolds (2007) let participants read stories describing everyday activities while recording their brain activity. After this, they let participants divide the stories into large and small events. The event structure provided by the participants correlated with changes in brain activation across different areas recorded in the free reading condition. As it was pointed out by Zacks and Tversky, the event detection mechanism produces a hierarchical coarse-to-fine structure, similar to the hierarchical structure of action planning. Seeing that events are conceptualized as spatiotemporal entities, the event segmentation theory allows to derive a hypothesis regarding the representational format of the predictive, generative models proposed by the free energy principle. If these models evolve to provide a hierarchical event-based structure, a spatial format seems plausible. According to this view, the models focus on the prediction of a change within the sensory input at a certain point in time and space.

2.2. Abstraction in Integrative Spatial Codes

There is indeed some evidence for an ISC that relates activation from different modal sources in the service of behavior control (see Cohen & Andersen, 2002, for an overview). The representation of the reachable space surrounding the body is characterized by highly interactive, adaptive spatial codes, which are used to control action within the *peripersonal space* (Holmes & Spence, 2004; Macaluso & Maravita, 2010). Peripersonal space encodings in premotor and parietal cortical areas respond to nearby stimuli relative to a particular body surface, anticipate approaching stimuli, and partially relocate their receptive fields during tool usage (Canzoneri et al., 2013; Fogassi et al., 1996; Iriki, Tanaka, & Iwamura, 1996). In their seminal study, Iriki et al. (1996) trained macaque monkeys to reach distant objects with a rake. Iriki et al. compared single cell recordings from neurons integration visual and proprioceptive information regarding the hands, before and after training. The visual receptive fields were extended to cover the space reachable by the rake. Apparently, the hand representation was adapted to the interaction possibilities offered by the tool. These results highlight the close relationship between the representation of bodily and external spaces. Such interactive spatial representations are fundamental for goal-directed behavior. For instance, to grasp an object, it is necessary to transform spatial information from different modalities and different frames of reference, that is, a visual location estimate, obtained in an eye-centered frame of reference, has to be transformed into a motor command, changing proprioceptive information in an effector- or body-centered frame of reference. The required computations are complex because the frames of reference are grounded in different sensory modalities, and differ with respect to the origin of the respective coordinate system. It seems that these transformations are at the core of the predictive ISC and that they are acquired through sensorimotor interaction. Hence, the predictive ISC provide an interactive action metric that operates on various modalities.

These spatial transformations are usually associated with activation of the posterior parietal cortex, and have been interpreted in terms of a common, eye-centered frame of reference, engaged in behavior control (Batista, Buneo, Snyder, & Andersen, 1999; Cohen & Andersen, 2000; Stricanne, Andersen, & Mazzoni, 1996). Single cell recordings in monkeys imply that this common frame of reference is encoded by neurons within the parietal reach region and the lateral intraparietal area. According to Cohen and Andersen (2002), this common frame of reference provides an abstract representation of the movement target relative to the eyes. Apparently, the representation is not motor based, since single-cell recordings have shown that the activation of the respective neurons stays the same, irrespective of the actual type of intended movement (e.g. eye or hand movement). Furthermore, the activation was not modulated by the sensory modality of the target (visual or auditory), implying that the representation is independent of the actual sensory modality. In contrast, eye position has been shown to affect activation in neurons encoding stimuli in other, for instance head-centered, frames of reference (e.g. Brotchie, Andersen, Snyder, & Goodman, 1995). This implies a central, integrative role for an eye-centered frame of reference, providing relative spatial mappings between effectors, eyes and objects. Thus, this common frame of reference seems a possible candidate for the neural realization of the assumed ISC.

The pivotal role of spatial transformations for higher cognitive functions was considered by Zacks and Michelon's (2005) in their *multiple system account*. According to Zacks and Michelon, there are three different kinds of reference frames that are relevant for interaction, namely object-based, egocentric, and environmental frames of reference. Object-based frames of reference are defined relative to external objects, egocentric frames of reference are defined relative to the self, while environmental frames of reference define a location relative to a fixed space. The brain seems to maintain multiple egocentric frames of reference. For interactions, those that represent objects relative to the observers perspective (eye-centered and head-centered frames of reference) and those that represent objects relative to the effectors, are most relevant. These frames of reference are continuously transformed, depending on changes in either one of them. For instance, head movements lead to an update of the head-centered frame of reference relative to object-based and environmental frames of reference. According to Zacks and Michelon, these transformation can be carried out or imagined, the latter case can be considered as a simulation within the ISC. Interactions yield changes in the spatial relations between egocentric and object-centered frames of reference, whereby the imagination of these changes is equivalent to the prediction of action outcomes. As it was pointed out by Zacks and Michelon, the ability to imagine or simulate transformations within and between frames of reference might be the basis for spatial manipulation skills, such as mental rotation. In the terminology of the multiple system account, mental rotation requires the transformation of an object-based frame of reference, relative to an egocentric and an environmental frame of reference. In mental rotation tasks, participants are usually requested to indicate the parity of a rotated object, or to perform a same / different classification of two objects, which are presented in different orientations. For both kinds of tasks, the response times increase linearly with the disparity, that is, the degree of rotation away from a canonical orientation (e.g. upright), or away from the object center. This implies that the required object-based transformation is realized by means of a continuous simulation, which follows the shortest path through object orientation space. When participants are required to mentally rotate displayed hands or feet, response times are elevated if the postures are difficult to achieve, with respect to biomechanical constraints (e.g. Ionta, Fourkas, Fiorio, & Aglioti, 2007). Hence, the mental rotation trajectory does not take the shortest path through Euclidean space, but considers factors like postural convenience. Furthermore, some neuroimaging studies have shown a contribution of premotor areas in mental rotation tasks (see Zacks, 2008, for an overview). Together, these results imply that the ability to simulate the spatial transformations involved in mental rotation is rooted within the sensorimotor system. If this is indeed the case, mental rotation would be one example for the realization of a higher cognitive function based on a simulation employing the proposed ISC.

Besides supporting spatial simulations, the proposed ISC might also provide the basis for relational knowledge. As it was pointed out by Walsh (2003), the ISC preserves sensorimotor relations. It does not simply provide information *where* an object is located relative to an egocentric frame of reference, but also relational information like *how far* it is away, or *how fast* it moves. This information is necessary for dynamic behavioral control and provides an analogue representation of quantity or magnitude. In his *theory of magnitude* (ATOM, Walsh, 2003), Walsh propose that this analogue metric for action can be used to represent arbitrary magnitudes and thus suggests a sensorimotor grounding for numerical cognition. Besides various developmental, neurophysiological, and behavioral studies supporting ATOM (see Bueti & Walsh, 2009, for an overview), the so-called *spatial numerical association of response codes* (SNARC, Dehaene, Bossini, & Giraux, 1993) effect is a vivid example for the overlap between spatial and numerical cognition. Judging a number to be comparatively small is faster in case of left response codes, while large number judgments are associated with right response codes. This effect has been observed in various environmental and effector-based frames of reference, highlighting again the flexibility of the ISC. ATOM is one example how understanding predictive sensorimotor encodings can yield insights regarding the grounding of higher cognitive functions. A further investigation of the properties of the involved ISC might reveal how spatial transformation models are abstracted from sensorimotor interactions and how they provide building blocks for symbolic representations.

2.3. Summary

The presented experiments and theories provide the theoretical background for the model outlined in Fig. 2.1. According to embodied cognition, cognitive functions are realized in terms of sensorimotor simulations. Based on the event segmentation theory, these sensorimotor simulations are assumed to focus on event boundaries, thereby providing a hierarchical, coarse-to-fine prediction of the next event boundary and the necessary sensorimotor activation to reach it. The different involved sensorimotor modalities are presumably related through an ISC. This notion is supported by research on the spatial codes involved in the representation of peripersonal space, a common, eye-centered frame of reference and the multiple systems account. It appears that this ISC provides a suitable format for behavior control since it allows abstraction from the actual sensory modalities to a certain degree, which in turn provides some means of generalization of spatiotemporal activation patterns. The formation of this code and its involvement in anticipatory behavior control can be described in terms of free energy minimization. Even if it is beyond the scope of the presented work, there is already evidence how such a predictive ISC can support higher cognitive functions like numerical cognition (as proposed by ATOM). Furthermore, since the event-predictive encodings, realized within the assumed ISC, convey abstract information about states and transitions between them, they might be considered precursors of symbolic representations. The main aim of the presented experiments was to investigate the properties of the proposed ISC. The pursued objectives are presented in more detail in the next section. After this, the different experiments are presented. A general discussion and an outlook regarding a possible implementation of the presented model conclude this thesis.

3. Objectives

According to the free energy principle (Friston, 2010), cognitive systems constantly try to infer the causes for their sensory input. This inference process yields the formation of predictive models. As it is shown in Fig. 2.1, I assume that the consistent changes in sensory input resulting from motor activity are processed through an ISC. This code is assumed to focus on the sensorimotor dynamics, which result in certain sensory states. An internal bias is assumed to highlight behaviorally relevant modalities, which provide the most informative signals regarding a certain state. These states are assumed to be characterized by discontinuities within the sensorimotor stream, hence they can be considered as event boundaries in terms of the event segmentation theory (Zacks & Tversky, 2001; Zacks & Swallow, 2007). Practice and experience allow to abstract from the current context, thus preserving the invariances within the dynamics that lead to a certain event boundary. According to the proposed model, an overall striving for consistency preserves the model structure. The acquired mappings relating sensorimotor dynamics and goal states allow goal-directed behavior control. Activation of a desired state is assumed to initiate the simulation of the sensorimotor dynamics necessary to achieve it. In order to do so, the ISC provides a multilevel prediction that involves the anticipation of the goal state, as well as the sensorimotor dynamics necessary to reach it. The ability to simulate action outcomes might be involved in higher cognitive functions, which are realized in terms of internal, spatial simulations, such as mental rotations.

The main aim of the presented work was to accumulate empirical evidence for these assumptions. More precisely, the presented experiments investigated (i) how the acquisition of structural knowledge is biased toward the task relevant modalities, (ii) how spatial models are kept consistent in case of conflicting sensory information,

(iii) how the ISC provides an anticipatory event-structure, which guides action, and (iv) whether the ISC is involved in higher cognitive functions (exemplified by mental rotation). The acquired data allow to derive conclusions how an abstract ISC can arise from sensorimotor experience, how it is maintained, and how it subserves behavioral control as well as higher cognitive functions.

All of the presented studies focused on the adaptation of the assumed internal models in case of sensory conflict. The introduction of localized multisensory conflict and the manipulation of experienced sensorimotor contingencies, for instance between vision and proprioception, required the implementation of novel, suitable experimental setups. Immersive virtual reality (VR) offers a reasonable compromise between experimental control and external validity. Two of the presented studies were carried out in an immersive VR, one study applied equipment and software designed for the VR setups in a classic display setup. In the next section, I give an overview of the developed VR paradigms and the according hardware, after this, I report the main studies.

4. Results

4.1. Natural Object Interactions in Virtual Reality

Algorithms and hardware described in this chapter have been applied in the following studies:

Schroeder, P. A., Lohmann, J., Butz, M. V., & Plewnia, C. (2015): **Behavioral bias for food reflected in hand movements: A preliminary study with healty subjects**. *Cyberpsychology, Behavior, and Social Networking*. 19(2), 120-126.

Lohmann, J., & Butz, M. V. (2017): **Lost in Space: Multisensory Conflict yields Adaptation in Spatial Representations across Frames of Reference**. *Cognitive Processing*, 1-18. doi: 10.1007/s10339-017-0798-5.

Lohmann, J., & Butz, M. V. (2016). **Multisensory Conflict yields Adaptation in Peripersonal and Extrapersonal Space**. Proceedings of the 13th Biannual Conference of the German Cognitive Science Society (KogWis 2016).

Lohmann, J., Gütschow, J., & Butz, M. V. (2017): **Grasping Multisensory Integration: Proprioceptive Capture after Virtual Object Interactions**. Proceedings of the 39th Annual Meeting of the Cognitive Science Society. London, UK: Cognitive Science Society.

Belardinelli, A., Lohmann, J., Farnè, A., & Butz, M. V.: **Mental Space Maps Into the Future**. *submitted*.

Development and evaluation of the algorithms and hardware was part of the following master and bachelor theses:

Gütschow, J. (2016): **Touching the Virtual: Dissociation of Vision and Touch in a VR Setup.** (Master's thesis, University of Tübingen, Tübingen, Germany).

Moghimi, S. (2016): **Interactive Visualization of Virtual Object Interactions in Unity®.** (Bachelor's thesis, University of Tübingen, Tübingen, Germany).

To investigate the role of sensorimotor contingencies in cognitive functions or the modus operandi of predictive models, it is necessary to manipulate sensorimotor mappings in a systematic way. This is difficult to achieve in real world setups, especially when manipulations of body representations are required. Classic paradigms, which are used to investigate the consequences of conflicting information regarding the own body representation, like the rubber hand illusion (RHI, Botvinick & Cohen, 1998), restrict the possibilities to study the effects of the conflict in an interactive setup. For instance, participants cannot perform an object interaction under visual control after the rubber hand illusion has been induced, since this would relieve the conflict. Other paradigms that were applied to investigate the adaptation of motor control in case of manipulated visual feedback - like experiments applying prism goggles - usually introduce a global sensory conflict in many different frames of reference at once. Hence, they do not allow to investigate how inconsistency in a single frame of reference, for instance hand space, is resolved. Virtual reality (VR) setups combined with real-time motion capture offer a solution to both issues, as they provide full control over the visual feedback received by participants as well as the visuomotor contingencies. Accordingly, VR setups have been successfully applied in the study of multisensory processing (Ernst & Banks, 2002), spatial representations (Gamberini, Seraglia, & Priftis, 2008; Linkenauger, Bülthoff, & Mohler, 2015) and bodily self-awareness (Ehrsson, 2007; Petkova & Ehrsson, 2008). The development of immersive, interactive VR setups that allow to study adaptation to multisensory conflict requires the integration of motion capture systems, the implementation of object interaction possibilities, and the integration of external stimulus devices to provide for instance vibrotactile feedback. In the presented experiments, I relied on the Unity® engine (Unity Technologies, San Francisco, California) which provides a powerful API equipped with a convenient C# interface. Unity® is distributed with the NVIDIA® PhysX® engine (Nvidia Corp, Santa Clara, California), which allows realistic physics simulations. In the next sections I give an overview how the mentioned requirements were realized within Unity®.

4.1.1. Motion Capture

To enable object interactions within VR, movements of the participants have to be recorded and displayed in real-time. For the reported experiments two different types of motion capture were used, namely optical and inertial tracking.

Optical motion capture relies on image processing of infrared data, it can be performed with infrared reflectors, so-called markers, or without. Systems relying on markers extract the position of the markers in space from the infrared image datastream. Such systems usually consist of multiple cameras which are configured to track a certain marker configuration. Systems without markers operate with less cameras and rely on online image recognition to identify the object of interest, which can either be a human body (in case of the Microsoft Kinect© sensor; Microsoft Inc., Redmond, Washington) or hands (in case of the Leap Motion© sensor; Leap Motion Inc, San Francisco, California). Markerless systems do not require calibration, however they are highly susceptible to occlusion. In contrast, systems applying markers can partly compensate for occlusion, since most of the time the relevant markers are visible for at least two cameras. Despite the problem of occlusion, markerless systems provide a fast and reliable way to map motor commands into VR. Plugins for Unity® allow a convenient integration of the Leap Motion Sensor© in VR scenarios.

Inertial tracking systems rely on IMUs (inertial measurement units) for motion capture. IMUs consist of accelerometers and gyroscopes, providing information with respect to their acceleration and orientation relative to gravity. For the motion capture approach used in this work, only orientation information was relevant. The system applied in the respective experiments consisted of a Synertial IGS-150 upper-body suit and IGS Gloves for the hands (Synertial UK Ltd., South Brighton, United Kingdom). Given that the sensor placement is known, the global orientation data provided by the IMUs can be used to calculate a kinematic chain. This kinematic chain can be used to animate an avatar in VR. This is a general purpose approach, the obtained orientation data can be applied to different kinematic chains. Inertial tracking systems do not suffer from occlusion and can provide detailed information for instance regarding the minute motor activity of the hands. Furthermore, tracking of the whole body kinematics allows to introduce multisensory conflict in different frames of reference. However, compared to markerless optical motion tracking, the inertial tracking system applied here required much more preparation time per participant. Even if occlusion is not an issue, inertial systems are susceptible to electromagnetic interference and movements can shift the sensors on the suit. Hence, the application of an inertial system is more complicated than a markerless optical system, but the data offers much more possibilities for experimental manipulations. No plugin was available for the applied inertial motion capture system, data collection and data broadcasting was realized through local network communication relying on TCP/IP.

Both approaches allow to visualize bodily movements in VR, but virtual object manipulations require an algorithmic interaction logic.

4.1.2. Object Interactions in VR

Both applied tracking methods allow to project a hand model into VR. To allow object interactions it is necessary to detect the on- and offset of grasps, furthermore, the logic of the bound movement of the grasped object has to be implemented. The VR setups presented in this work did not include force feedback, hence, all object interactions were performed without haptic feedback, grasping therefore purely relied on visuomotor control, sometimes augmented with visual or vibrotactile stimulation. Grasping without haptic feedback can be difficult, since the fingers are partially occluded by the object. Given this limitation, the aim was to derive a grasping algorithm that on the one hand feels rather natural during interactions, but on the other hand did not impose unbearable constraints with respect, for instance to finger placement on the target object. The algorithm presented here can be considered as a combination of pantomimic and natural grasping.

The algorithm constantly monitors the grip aperture of the relevant hand. Precision and power grasps are detected if the respective postural constraints are fulfilled. In case of precision grips, the distance between index finger and thumb has to be below a distance threshold. For the power grasp, the average distance between finger tips (except the thumb) and palm has to be minimal. If a grasp is detected and no object is grasped yet, the local space surrounding the hand is checked for graspable objects. If an object is detected, it is considered to be grasped and moves along with the grasp center, indicated by the center of thumb and index finger tip (precision grip), or the centroid of palm and finger tips (power grasp). Object movement is realized via simulated forces, which are applied to the object depending on the movement speed of the hand. The object is released when the hand is opened (power grasp), or when the grip aperture exceeds a certain threshold (precision grip). Upon releasing, the object maintains its current speed, hence the algorithm allows to simulate throwing as well as deliberate object placement. The algorithm can be adapted to consider certain, object-specific constraints with respect to grasping, like dual-hand grasping, or finger placement on the object.

Augmenting Interactions with Vibrotactile Feedback

As it was noted above, purely visual grasping control can be difficult, since the target object can partially occlude the fingers, preventing visual control. Hence, some kind of tactile feedback would be desirable to support grasping. Furthermore, the ability to provide tactile information regarding the extends of an object would allow to dissociate visual and tactile size perception, providing an additional experimental manipulation in the study of mulitsensory integration during object interactions. In order to so Jakob Gütschow and I assembled a vibrotactile stimulation device during his master thesis and implemented a software interface which allows to control it from external programs, like Unity®.

Basically, the device consists of a microcontroller which is used to control five vibration motors, which can for instance be placed under the fingertips of a participant. We decided to use an Arduino Uno microcontroller (Arduino S.R.L., Scarmagno, Italy), since it is affordable, provides a convenient programming interface, and comes with build-in, pulse-width modulation (PWM) connectors which can convert 8-bit [0 - 255]

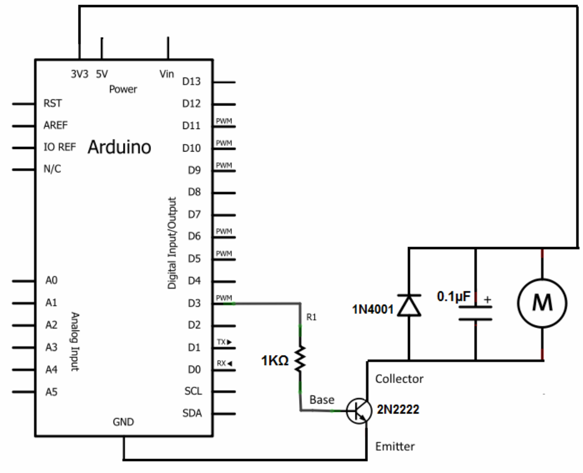


Figure 4.1.: The wiring diagram for the protective circuit. The design pattern is based on an example provided by http://learningaboutelectronics.com. This protective circuit has to be assembled for each motor connected to the microcontroller. The diode serves as surge protector, absorbing voltage spikes emitted by the motor. As an additional safety measure, the capacitor is intended to absorb any voltage spikes produced by the motor. The transistor is used to amplify the current output to reach a suitable voltage range. The resistor protects the motor from overvoltage.

inputs to simulated analog output for connected devices. Due to the PWM output, it is possible to control the vibration strength of the attached motors in a continuous fashion. The SDK of the controller allows programming in terms of simplified C++. Programs can be uploaded into the 32Kb flash memory via an USB-connection, this USB-connection can also serve as power supply. Programs stored in memory run continuously as long as the device is powered up.

Different types of shaftless vibration motors can be used. Depending on the applied hardware, it might be necessary to assemble a protective circuit. In a first version of the device we used LilyPad Vibe Boards (SparkFun Electronics, Niwot, Colorado) that combine a vibration motor along with a diode and a 33 Ω resistor on a circuit board. The diode provides surge protection against voltage spikes which might damage the microcontroller. The resistor limits the current flow through the motor. LilyPad Vibe Boards can be attached directly to the PWM outputs, but they are comparatively large (diameter of about 2 cm) and can interfere with motion tracking when attached to the finger tips. Hence, we used smaller vibration motors with a diameter of about 0.5 cm and a height of 0.3 cm in later versions of the device, however, this required the

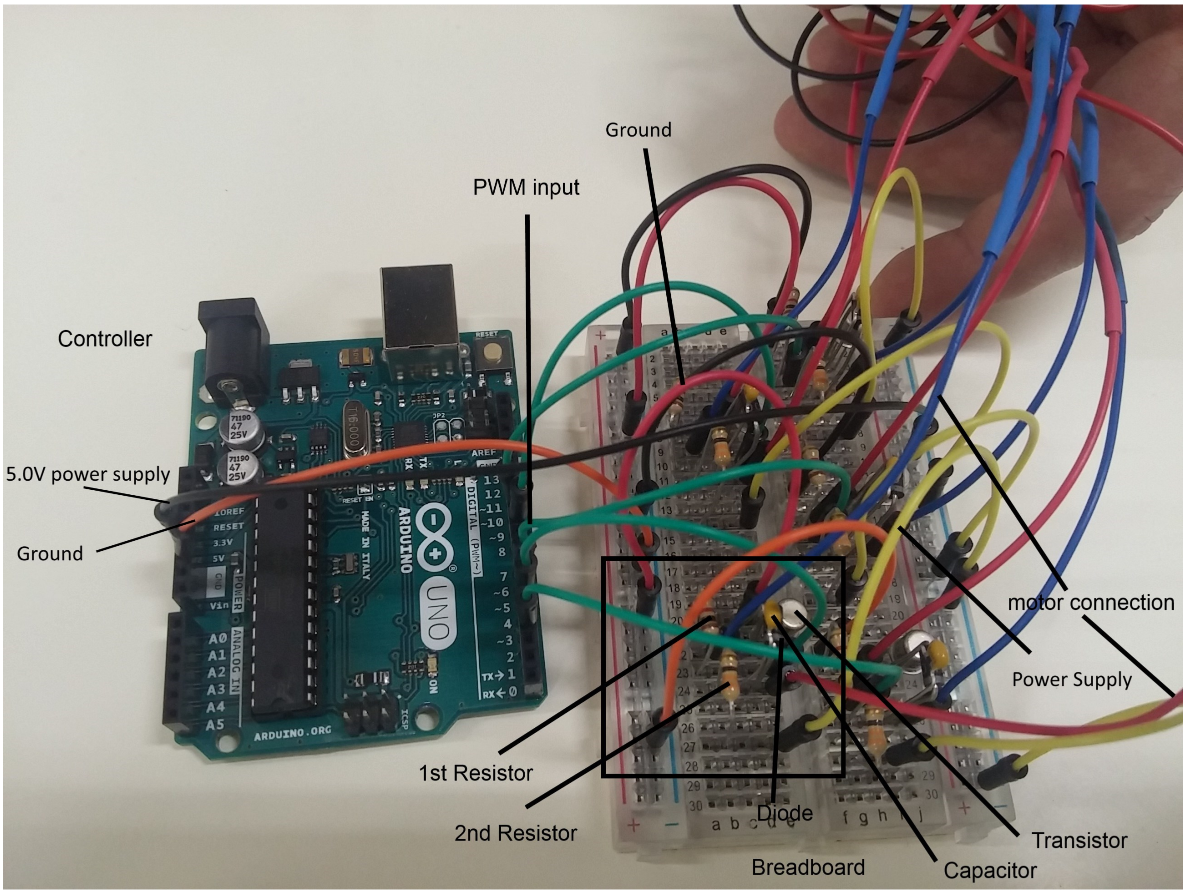


Figure 4.2.: The protective circuit for five motors arranged on a single breadboard. The microcontroller is shown on the left. Power supply is realized via the black cable, grounding is realized by the orange cable. Each of the motor circuits receives input from a PWM connector (green cables). One circuit is marked with a black square and the components are annotated. The long yellow cable provides the power supply, while the short red cable serves as grounding. The blue and the red cable are the connections to the motor. The diode is hard to see but it is directly behind the capacitor (cf. Fig. 4.1). Please note the additional second resistor which was included to further protect the motors when running the setup with the 5.0 V power output provided by the microcontroller. The shown assembly does not require soldering connections, the only mandatory soldering connections are the connections between motors and output cables.

separate assembly of the protective circuit. To realize a stable setup, we applied a more complex circuit design than the one realized on the LilyPad Vibe Boards. First, a diode has to be connected with the motor in parallel. The diode acts as a surge protector against voltage spikes which could damage the microcontroller. Second, a capacitor is connected in parallel to the motor, which absorbs voltage spikes produced by he motor. Voltage spikes are a common problem when running vibration motors. They usually emerge when the brush contact that provides current to the motor opens or closes. Third, a transistor is added to amplify the comparatively weak current output provided by the microcontroller. Fourth, a resistor is added to make sure that the applied current cannot damage the motor.[1] The motor and all the protective elements connected in parallel, have to be connected to the collector of the transistor. The base of the transistor receives the PWM input, while the emitter is connected to the grounding (see Fig. 4.1) The assembled circuit is shown in Fig. 4.2. At a current of 3.0 V, the motors produce a vibration with 200 rotations per second, the resulting vibration amplitude is about 0.75 g, which is well noticeable. Due to the 5.0 V power output provided by the microcontroller we added a second resistor to the protective circuit. As it was noted above, the vibration strength can be continuously scaled through the current applied to the PWM outputs.

The software running on the controller can be kept rather simplistic and consists of three parts: the definition of the pwm-to-motor mapping, a setup method defining the parameters of data transmission, and a continuously running loop. This controller program can be interfaced from external applications by directly writing data to the respective port. For instance, collision detection in Unity® can be used to trigger or stop vibration. Furthermore, due to the general purpose interface, the device can be used as stimulation device in experimental setups that require vibrotactile stimulation.

A detailed description of the hardware setup, the controller software and an example for the integration into Unity®, can be obtained from: http://www.wsi.uni-tuebingen.de/lehrstuehle/cognitive-modeling/staff/staff/johannes-lohmann.html

Validation

Usability and validity of the derived algorithm have been investigated in different experiments. Furthermore, effects of the vibrotactile stimulation have been evaluated.

---

[1] The transistor amplifies the base current about 100 times. Hence, the range of the suitable current becomes rather small. The resistor is necessary to keep the current arriving at the base within this range.

Perceived Naturalness of Object Interactions In her bachelor thesis, Sara Moghimi compared the basic grasping algorithm described above with a more sophisticated extension, considering actual finger placement on the target object as an additional constraint. In both cases, grasping was augmented with vibrotactile feedback. The extended version of the grasping algorithm required different grasps, depending on the shape of the object. She let participants perform a grasp and carry task, recording movement onset and interaction times, as well as success of the interaction. Furthermore, she recorded self-reports regarding usability, difficulty, and perceived naturalness of the object interaction on a Likert scale ranging from -3 to +3. Participants were instructed to provide their rating in comparison to actual natural grasping. An ANOVA with respect to the time measures and error rates revealed longer interaction times for the more sophisticated grasping algorithm. Error rates, comprising object destruction due to small apertures and involuntary releases during the carry phase, were low in general and did not differ with respect to the grasping algorithm. The self reports only yielded significant differences with respect to perceived naturalness. Participants judged the more sophisticated grasping algorithm to be more natural than the simpler one. Even if the results imply that including object-specific grasping constraints increases the perceived naturalness of the grasping, it remains open how participants would judge the interactions in VR in direct comparison to natural object interactions.

Effects of Vibrotactile Feedback To investigate whether vibrotactile feedback can indeed facilitate virtual object interactions, Jakob Gütschow compared the described grasping algorithm in a reaching and a grasp and carry task, either with vibrotactile feedback or without. An ANOVA analysis of the error rates revealed improved performance in case of vibrotactile feedback, especially in case of small objects, requiring precise control of the grip aperture (see Fig. 4.3, left panel). The experiment was carried out with an early version of the grasping algorithm, hence the overall error rate was high. Comparing the error rates in the two feedback conditions, it became clear that vibrotactile feedback strongly reduced the likelihood of involuntary object destruction (see Fig. 4.3, right panel). These results show that vibrotactile feedback can indeed support virtual object interactions, but again, it remains open how this performance relates to natural object interactions.

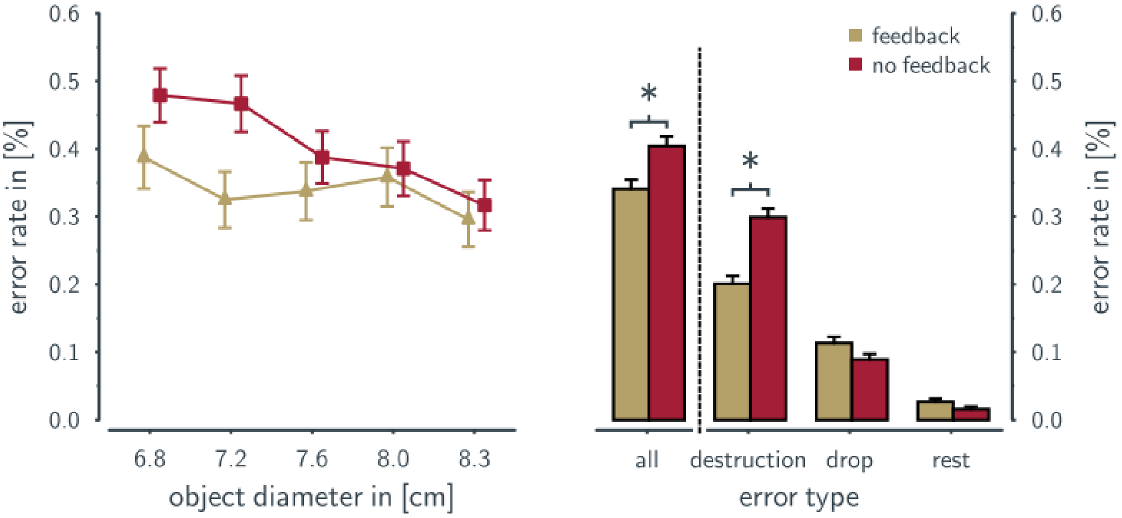


Figure 4.3.: Performance in a reaching and a grasp and carry task, carried out either with (yellow lines and bars), or without (red lines and bars) vibrotactile feedback. Especially for small object, vibrotactile feedback can improve performance (left panel). A separate analysis of the error types show that while vibrotactile feedback reduces the overall error rate, this improvement is due to less involuntary object destructions (right panel).

Kinematics in Virtual Object Interactions Since all VR studies involving object interaction required motion tracking, it was possible to collect kinematic data during the virtual object interactions. Still, a direct, statistical comparison with reference data (e.g. Jeannerod, 1986; Jeannerod, Arbib, Rizzolatti, & Sakata, 1995) is not possible, due to the differences in the setup. However, the obtained trajectories can be compared qualitatively with the reference data. Average grip aperture between thumb and index finger and wrist velocity profiles from 26 participants from a virtual object interaction task (the VR version of the study described in section 4.4; not published yet) are shown in Fig. 4.4. The profiles are normalized with respect to an arbitrary time scale ranging from 0 to 100. Velocity profiles were obtained via a five-point stencil applied to the positional trajectories of the right wrist. The light area around the curves indicates the standard error of the mean. Two typical kinematic features of grasping movements are preserved in the trajectories for the virtual object interaction. First, the velocity profile shows the typical inverted u-shape. Apparently, participants were able to use the visual depth information, provided by the stereoscopic presentation to plan and carry out the transportation component of the prehension movement in a rather natural way. Second, the grip aperture profiles show the typical overshoot in the maximum grip aperture just before reaching the object. This implies that also the manipulation component of the prehension movement was carried out in a way similar to natural object interactions. The interindividual variance with respect to velocity and aperture profiles is low, as it is indicated by the small standard errors. This again implies that the object interactions could be carried out in a straightforward, rather natural way and that the kinematics did not depend on individual differences in coping with the VR setup.

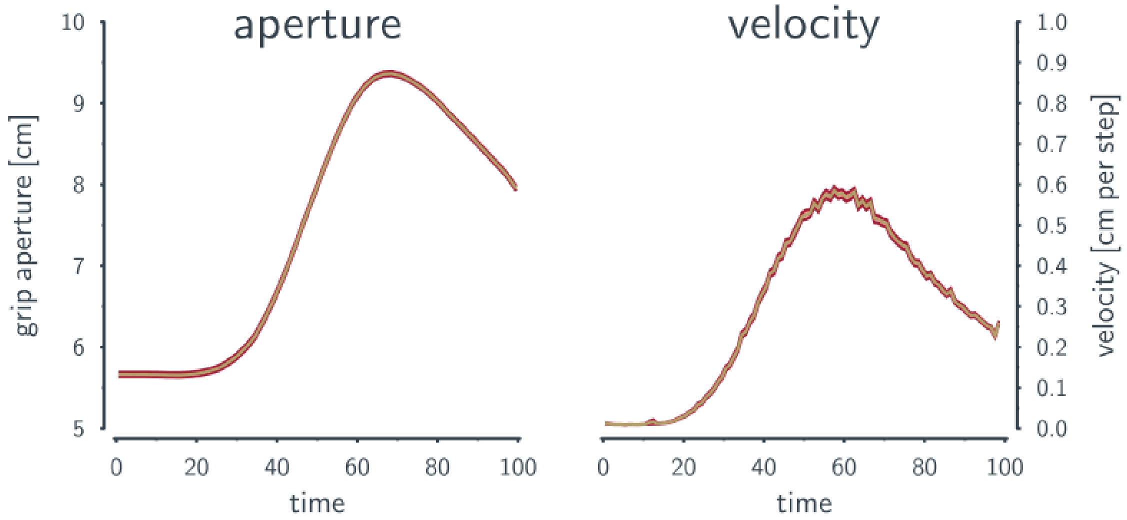


Figure 4.4.: Grip aperture and velocity profiles during virtual object interactions. Profiles were obtained by aggregating trajectories from 26 participants. Red lines indicate the standard error of the mean, yellow lines indicate the mean values. Trajectories were normalized on an arbitrary time axis with 100 steps before aggregation. The aperture data reproduces the typical overshoot before reaching the object. The velocity profile shows the typical, inverted u-shape. Both features are characteristic for natural object interaction (cf. Jeannerod, 1986).

4.1.3. Online Cognition Disclosed by Motion

Kinematic and dynamic parameters of movement execution are an extremely rich data source that entails information of both, planning as well as control processes. With respect to mouse tracking, the correlation between movement execution and cognitive control has already been shown (J. Freeman, Dale, & Farmer, 2011). Compared to mouse tracking, virtual object interactions allow a more natural setup, which for instance allows to vary movement outcomes and interaction goals in a more deliberate manner. Such setups allow to study goal-directed behavior on multiple levels, ranging from sensorimotor control process to top-down attentional priors. In the first published study applying the derived grasping algorithm, we investigated whether habitual behavioral bias affect movement trajectories in object interactions. In a cooperation with colleagues from the Neurophysiology & Interventional Neuropsychiatry group at the University Clinic of Psychiatry and Psychotherapy, we investigated the interplay between motor execution and cognitive control when interacting with either virtual food, or neutral stimuli. In this experiment, participants had to grasp or ward high-calorie food or neutral, but grasp-affording ball objects. In case of grasping, participants had to carry the objects into a container object, located near themselves. For both types of stimuli and interactions, we measured movement parameters like movement onset, object contact, and - in case of grasping movements - the collection time. Compared to ball objects, food items were collected significantly faster. Movement onsets for food stimuli were significantly faster than for ball objects. Furthermore, the differences in collection time correlated with BMI and eating related attitudes. These findings dovetail with previous results, showing an attention bias for food stimuli, compared to neutral stimuli (see, Loeber et al., 2012, for an overview). Usually, mainly visual tests, like dot-probe (Yokum, Ng, & Stice, 2011), or stroop tasks (Davidson & Wright, 2002), have been used to investigate attentional biases in the processing of food related stimuli. Our results show a stable behavioral bias with respect to more natural prehension movements.

4.1.4. Summary

The derived grasping algorithm - in combination with vibrotactile stimulation, or without - proved to be appropriate for the study of object interactions in VR. So far the evaluations imply that a rather natural grasping experience can be implemented in

VR setups. If object specific grasping is considered, the perceived naturalness of the interaction is significantly increased. Vibrotactile augmentation can further improve the grasping experience and performance. Obtained kinematic profiles resemble profiles obtained in actual object interactions. However, a systematic comparison of virtual and natural grasping is still pending. Such a study would require a setup that allows the execution of the same interactions in VR and the real world. Furthermore, tracking of the target object would be desirable to investigate whether participants interact with virtual objects in the same way as with real object. The Leap Motion sensor© cannot track external objects, furthermore, an object partially occluding the hand interferes with the tracking. The inertial tracking system allows precise motion tracking, irrespective of occlusion, however, external objects cannot be tracked by this system. Hence, a combination of the tracking suit and an external, marker based motion capture setup seems most suitable for a direct comparison of natural and virtual object interactions. Such a setup will allow to directly compare kinematic and dynamic properties of both types of interaction. The data might disclose quantitative differences between both kinds of interaction, like it has been shown with respect to natural and pantomimic interactions (Goodale, Jakobson, & Keillor, 1994).

In general, virtual object interactions seem to be a viable approach to investigate goal-directed behavior on multiple levels. Besides the continuous tracking data, VR setups allow controlled manipulations of the visuomotor mapping which are difficult to realize in real-world setups (see the studies described in section 4.2 and 4.3). Apart from its application in basic research, interactive VR setups seem a promising tool for psychotherapy and experimental clinical psychology (D. Freeman et al., 2017; Riva, 2005). Considering these prospects regarding the application of VR in psychology, neuroscience, psychiatry, and in possible interdisciplinary cooperations, Philipp Schröder, Prof. Butz, Prof. Plewnia and I organized the firtst **VECTOR** workshop on **V**irtual **E**nvironments: **C**urrent **TO**pics in psychological **R**esearch in Tübingen in 2016[2].

---

[2] Details can be found at the workshop website: http://vr-workshop-tuebingen.org/

4.1.5. Personal Contribution

I implemented the basic version and different extensions of the grasping algorithm. I also implemented the TCP/IP interface for the inertial tracking suit. Together with Jakob Gütschow and Fabian Schrodt, I assembled the first version of the vibrotactile stimulation device. I assisted Jakob Gütschow and Sara Moghimi during the implementation and the setup of their experiments. Together with Philipp Schröder, I developed the paradigm for the study on behavioral biases during interactions with food stimuli. I implemented the VR setup and assisted Philipp Schröder during the data analysis and the preparation of the manuscript. Together with Philipp Schröder, I wrote the proposal for the funding of the VECTOR workshop by the Future Concept of the University of Tübingen. The organization and realization of the workshop was mostly done by Philipp Schröder and me. Silke Bieck assisted us during the organization, Verena Heußer, Simone Kurek, Johannes Palagy, and Fedor Schlegel helped us on-site at the workshop.

4.2. Inference Through Interaction: Top-Down Biases in Multisensory Integration

Results presented in this section have been published in:

Lohmann, J., Gütschow, J., & Butz, M. V. (2017): **Grasping Multisensory Integration: Proprioceptive Capture after Virtual Object Interactions**. Proceedings of the 39th Annual Meeting of the Cognitive Science Society. London, UK: Cognitive Science Society.

According to the free energy principle, cognitive systems constantly try to minimize the probability for unexpected sensations. This can be realized by either adapting internal models, or by changing sensory signals through interaction. To enable efficient interactions, feature selection for internal models should be driven by task relevance. Applied to the perceptual domain, task-relevance should bias the estimation of object properties. To test this assumption, multisensory integration - referring to the process of creating a coherent percept from multiple sensory sources - seems a suitable test case. So far, most studies on multisensory integration focused on bottom-up influences, like the reliability of the combined information sources. In their seminal study Ernst and Banks (2002) dissociated vision and touch during a comparative size estimation. Participants explored a stimulus visually and by means of touch, afterwards a reference was presented and participants had to indicate if the initial stimulus was taller or smaller than the reference. In half of the trials, vision and haptics were dissociated during the intial exploration that is visual and haptic size of the stimulus did not match. The comparative judgments of the participants implied a dominance of visual information within in the combined percept. Besides the psychophysic assessment, Ernst and Banks could show that the observed multisensory integration occured in a statistically optimal fashion, considering the variances of the different modalities.

Based on these findings, multisensory integration is considered to be realized in terms of a maximum likelihood integrator that combines different sensory signals based on their individual reliability, providing a Bayesian estimate of the external world (Ernst & Bülthoff, 2004). Typically, the visual modality is the most reliable source of information and usually dominates in case of conflicting sensory information, a phenomenon referred to as visual capture. Accordingly, multisensory integration has been usually described as an automatic process. This interpretation has been questioned based on findings implying a modulation of multisensory integration by top-down attentional mechanisms (see Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010). In a study on attentional control over perceptually ambiguous visual stimuli, van Ee, van Boxtel, Parker, and Alais (2009) could show that the presence of a task relevant stimulus in another modality, which matched the temporal rate of either one of the visual stimuli, increased the ability of the participants to hold the respective visual stimulus dominant. This implies that focusing on a spatiotemporal stimulation pattern can serve as a top-down prior for multisensory integration. Further evidence for top-down modulations of multisensory integration by spatial attention comes from EEG studies (e.g. Talsma & Woldorff, 2005). According to Talsma et al. top-down attentional effects on multisensory integration can usually be observed in case of high competition between stimuli and in case of comparatively complex environments. Natural object interaction qualifies for both criteria, since it requires the consideration of different, competing modalities. Seeing that most studies on multisensory integration focused on perceptual tasks, I aimed at investigating the integration process in a more natural setup, where top-down priors can be related directly to task demands. In order to investigate whether task-relevance can bias the size estimation of an object in favor of task relevant, proprioceptive information, Jakob Gütschow and I combined an object interaction with a size estimation task. Participants performed a grasp and carry task in an immersive virtual reality. On a trialwise basis, seen and felt grip aperture were dissociated, inducing conflict between vision and proprioception. After the object interaction participants had to indicate the size of the object they interacted with either visually or based on the grip aperture. Due to task relevance, the noisy proprioceptive information should dominate the percept compared to the more reliable, but less relevant visual information.

Besides the investigation of possible top-down biases in multisensory integration, the proposed setup also allows to probe the storage format which is used in multisensory representations. Previous studies on the maintenance of multisensory information showed interactions between different modalities. For instance Morey and Cowan (2005) could show that verbal load in a visual working memory task interfered with the maintenance of the visual information. Morey and Cowan interpreted these results in terms of a general attentional resource which is involved in the maintenance of information. The nature of this general resource remains elusive, even if Morey and Cowan argued that visuo-spatial maintenance and verbal retrieval seemed to share resources. Besides this evidence for a - possible spatial - shared resource involved in working memory maintenance, a recent study showed that unimodal retrieval from multisensory representations is affected by previous modal encodings. Thelen, Talsma, and Murray (2015) investigated the performance in a continuous recognition task, where participants had to indicate whether a visual or auditory stimulus appeared for the first or the second time in a block of trials. Some of the objects were presented in a unimodal fashion (either visual or auditory) or in a multisensory fashion (visual and auditory). Auditory recognition was enhanced if initial multisensory presentations were congruent, that is, both the irrelevant visual and the relevant auditory stimuli were not encountered earlier. The results imply the involuntary generation of a multisensory representation which could interfere with auditory retrieval in case of incongruent stimulus pairings. These results imply a multisensory representation which fits well with the assumption of multisensory internal models proposed by the free energy principle. These internal models provide a Bayesian estimate for the environment and are continuously updated given the internal top-down priors and the bottom-up sensory evidence. The overall goal should be effective object interaction, hence, model updates should be biased towards task relevant information. Hence, the free energy principle predicts multisensory memory and provides an account how top-down and bottom-up processes contribute to an integrated representation (see also Quak, London, & Talsma, 2015, for similar reasoning).

The experimental setup proposed here allows to test the prediction of an integrated representation, biased towards task relevant modalities. Retrieval was always unimodal, that is, participants either had to give a visual or a proprioceptive estimate of the object size. If modal information is represented independently, the veridical visual information should be accessible without being affected by the manipulated proprioceptive information. If visual estimates are biased in a similar way as proprioceptive estimates, this would indicate a joint representation of visual and proprioceptive information.

4.2.1. Method

The aim of this study was to identify possible influences of task relevance on the retrieval of multisensory information. To allow for the relevant dissociation between felt and seen grip aperture, the experiment was realized in terms of a VR setup. Participants had to perform a grasp and carry task, the object interaction was augmented with vibrotactile feedback, which signaled when the relevant object was grasped. This allowed to dissociate the seen and felt grip aperture, inducing conflict between vision and proprioception. The target object was a virtual cube with an edge length of either 7.0 cm, 7.35 cm, 7.7 cm, 8.05 cm or 8.4 cm. Following the object interaction and after applying a visuo-tactile mask, participants had to judge the size of the object they interacted with either visually or based on the grip aperture. The schedule of a single trial is shown in Fig. 4.5. The multisensory conflict between seen and felt grip aperture was realized in terms of an angular offset at the visual root joints of the index finger and the thumb, increasing or shrinking the visual grip aperture accordingly. In one third of the trials, the fingers were rotated towards each other by 10°, in one third of the trials, the fingers were rotated away from each other, in the remaining trials, no multisensory conflict was induced. The different angles applied and the resulting grip apertures are shown in Fig. 4.5 (lower right). Due to the manipulation, participants had to compensate for the visual offset during the object interaction. To compensate a visual offset shrinking the grip aperture, the grip aperture had to be wider, while a visual offset extending the grip aperture required a closer grip aperture. Successful grasping always yielded the same visual effect, that is, the visual grip aperture always matched the actual extends of the object, while the felt (tactile and proprioceptive) aperture differed from this impression. After carrying the object into the target container, participants had to return into the initial position, after this a visual mask was applied along with a vibrotactile mask. Visual masking was realized in terms of randomly sized cubes spawning at the initial cube position, while tactile masking consisted of random activations of the vibration motors. Then, the screen went black and one of the two reproduction conditions was presented. In the visual reproduction condition, a slider appeared along with a reference cube. Participants were instructed to replicate the size of the cube they interacted with by moving the slider. In the proprioceptive reproduction condition, participants were requested to maintain the a grip aperture that fit the size of the cube they interacted with. Due to the multisensory conflict, the proprioceptive estimates should be biased according to the actual, felt grip aperture. Hence, visual manipulations shrinking the seen grip aperture, should yield overestimation, while manipulations extending the grip aperture should yield underestimation. If these effects also occur in visual reproduction, this would imply a multisensory representation and a bias towards the task relevant, proprioceptive modality. This effect should be visible in a two-way interaction between cube size and visual offset condition.

4.2.2. Results

Size estimates were analyzed with a repeated measure ANOVA based on the applied 5 (cube sizes) × 3 (visual offset condition) × 2 (reproduction condition) within-subject design. Both proprioceptive and visual estimates were biased towards the actual grip aperture, necessary to grasp and carry the target object. However, the variance within the visual estimates was much lower and the bias towards the actual grip aperture was smaller compared to the proprioceptive estimates, yielding a significant three-way interaction instead of the expected two-way interaction (see Fig. 4.6). To check whether visual offsets shrinking the seen grip aperture always led to overestimation and visual offsets increasing the seen grip aperture always led to underestimation, with respect to each other and the control condition, post-hoc t-tests were performed, separately for both reproduction conditions. The respective t-tests yielded significant differences for all comparisons in case of proprioceptive reproduction. Proprioceptive estimates were significantly larger in case of shrunk visual grip apertures, both

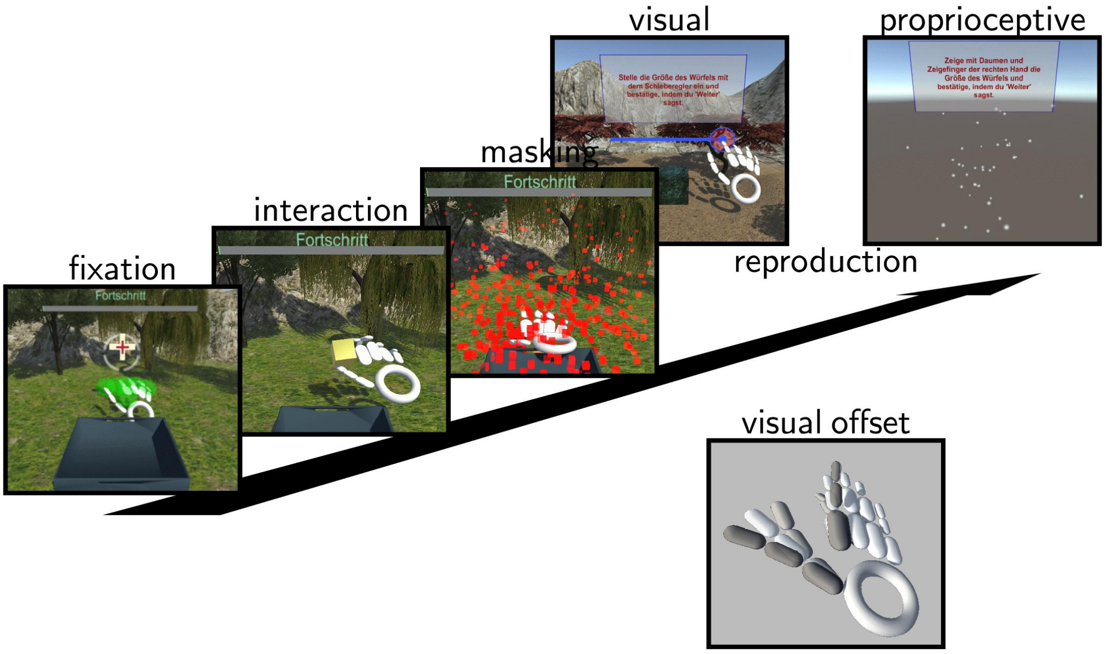


Figure 4.5.: Schedule of a single trial. Participants had to maintain a start position (indicated by the green spheres) and to center their field of view on a fixation cross. Then, a cube appeared. Participants were requested to grasp the cube and to carry it into the container in front of them. After placing the cube in the container, participants had to return to the initial position. After this, a visuo-tactile mask was applied and the scene faded out. Then, either a visual, or a proprioceptive reproduction scene faded in. In the visual reproduction scene, participants had to reproduce the size of the cube they interacted with using a slider. In the proprioceptive scene, participants were requested to maintain the a grip aperture that fit the size of the cube they interacted with. Participants confirmed their estimate by a verbal response. After this, the next trial started. Multisensory conflict between seen and felt grip aperture was realized in terms of an angular offset at the visual root joints of the index finger and the thumb. Examples for inward (light gray) and outward (dark gray) offsets, along with the condition without offset (white) are shown in the lower right corner.

compared to the control condition and trials with increased visual grip apertures. For increased visual grip apertures the opposite was true. In case of visual estimates, only the differences between the offset conditions were significant, both offset conditions did not differ significantly from the control condition. As noted above and shown in Fig. 4.6, this pattern of results was further modified by a significant three-way interaction. Together, the results show that visual estimates are a biased towards the

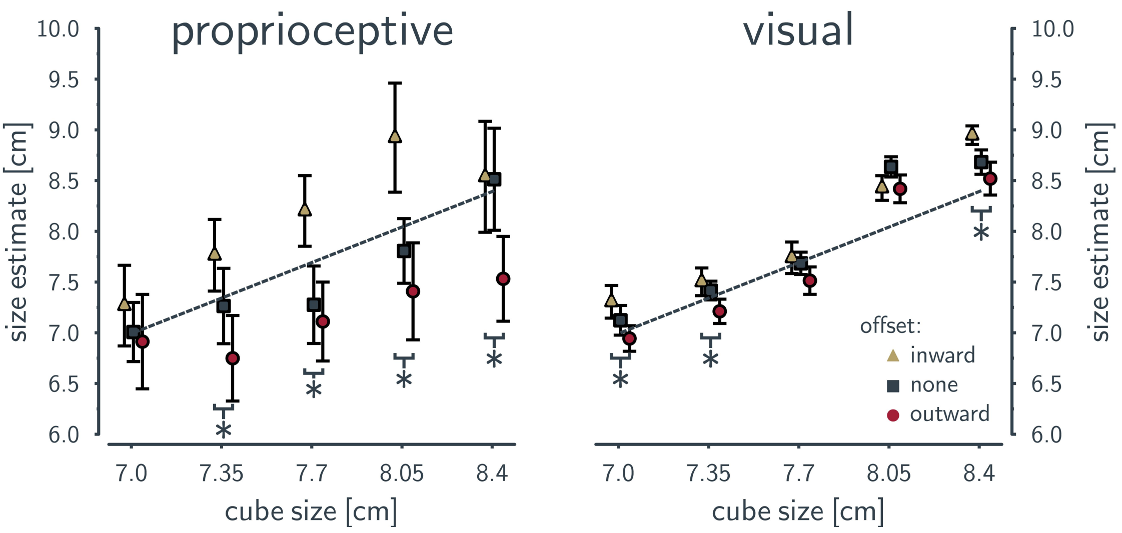


Figure 4.6.: Three-way interaction between reproduction condition, cube size and visual offset. Differences between the inward and outward offset visual conditions were analyzed with directed t-tests (inward > outward). Significant differences with p < .05 are indicated by an asterisk. Tests were adjusted for multiple comparisons. The actual cube size is indicated by the dashed line. Both visual and proprioceptive estimates are biased towards the actual grip size, implying a dominance of proprioceptive information in the estimate.

felt grip aperture. However, this effect was much more pronounced in proprioceptive estimates. Proprioceptive estimates were much noisier compared to visual estimates, implying that reliability of the modal information cannot account for the observed proprioceptive capture.

4.2.3. Summary

The aim of the presented experiment was to investigate whether multisensory integration can be biased by task relevance and whether this bias is visible in unimodal retrieval. The results show a bias towards task relevant proprioceptive information both in proprioceptive size estimates as well as in visual size estimates. The strength of the bias differed between both modalities, proprioceptive estimates were affected to a stronger degree than visual estimates. This implies that task relevance indeed modulated the formation of the multisensory memory trace. It is noteworthy that the pattern of results apparently cannot be accounted for by the reliability of the proprioceptive information. In general, proprioceptive estimates were much noisier than visual estimates. The results fit the free-energy perspective put forward in the outlined model (see Fig. 2.1) as well as the multisensory account for working memory, proposed by (Quak et al., 2015). According to the free energy principle, a Bayesian estimates of the environment is used to guide goal-directed interaction and to minimize uncertainty accordingly. These estimates are continuously updated, especially if model predictions conflict with actual sensory input. Since these models represent a multisensory environment, they are assumed to be multisensory as well. To keep these internal models consistent, conflicting information should be accounted for by adapting estimates in different modalities. This adaptation should not only depend on bottom-up sensory information and the reliability of the respective information, but also on the purpose of the model, that is, behavior control. Hence, this view can account for the observed bias towards task relevant proprioceptive information in the general estimate, but also for the finding that the proprioceptive information did not completely dominate the estimate.

Besides a bias towards task relevant modalities, the results also imply a maintenance mechanism that tries to keep the internal representations consistent. This striving for consistency is assumed to be a basic property of the ISC outlined in Fig. 2.1. According to the proposed model, this mechanism should not be restricted to a single frame of reference, for instance an object based one, like in this study. This striving for consistency was further investigated in the study presented in the next section.

4.2.4. Personal Contribution

The experiment was conducted by Jakob Gütschow as part of his master thesis. Prof. Butz and I derived the paradigm together with Jakob Gütschow. Jakob Gütschow, Fabian Schrodt and I assembled the vibrotactile stimulation device. I implemented the motion capture interface and the grasping algorithm applied in the experiment. I assisted Jakob Gütschow in the data analysis and prepared the manuscript.

4.3. Consistency in Spatial Body Representations Across Frames of Reference

Results presented in this section have been published in:

Lohmann, J., & Butz, M. V. (2017): **Lost in Space: Multisensory Conflict yields Adaptation in Spatial Representations across Frames of Reference**. *Cognitive Processing*, 1-18. doi: 10.1007/s10339-017-0798-5.

A preliminary version of this work was presented at KogWis 2016 (Brain Products best paper award):

Lohmann, J., & Butz, M. V. (2016). **Multisensory Conflict yields Adaptation in Peripersonal and Extrapersonal Space**. Proceedings of the 13th Biannual Conference of the German Cognitive Science Society (KogWis 2016).

The presented results regarding multisensory object representations did not only show an influence of task relevance on the weighting of the involved modalities, but also implied a general striving for consistency in the overall representation. This striving for consistency can account for the observed bias in the visual estimates, which resembled the bias observed in proprioceptive estimates. However, the study was limited to effects of multisensory conflict on a single, object-based estimate. As it is shown in Fig. 2.1, I argue that this striving for consistency is a general mechanism subserving the maintenance of the ISC. To test this hypothesis, I conducted an experiment to investigate whether compensation for multisensory conflict yields adaptation in different representations across frames of reference. From an action-centered perspective, a mechanism which assures consistency across different spatial representations seems highly desirable. Considering a simple grasping movement, many different spatial representations have to be considered. A visual estimate of the object position within an eye-centered frame of reference has to be translated into a motor command that aligns an object- and an effector-centered frame of reference (see McGuire & Sabes, 2009). As mentioned in the introduction, the involved transformations are complex, since the frames of reference not only differ with respect to their origin but also with respect to the weighting of different sensory modalities. For instance it has been shown that weighting of visual information increases with distance to the body, while proprioception dominates the representation of space close to the body (Longo & Lourenco, 2006). As it was argued above, the required transformations are acquired through sensorimotor interactions and can be used to guide action planning and execution (see Butz, Herbort, & Hoffmann, 2007, for a computational model for this approach). This requires consistency in the mappings between frames of reference. If, for instance, positional estimates in hand- and eye-space differ, eye-hand-coordination would be disrupted. Especially spatial body representations, covering the peripersonal space, have been shown to be highly adaptive and to integrate multisensory information (Holmes & Spence, 2004). Seeing the fast adaptation of peripersonal space, for instance in case of tool-use (e.g. Canzoneri et al., 2013), peripersonal space representations seem a viable test-bed to evaluate adaption processes induced by multisensory conflict. Previous studies applying multisensory conflict paradigms could show that manipulating the visual hand position yields adaptation in reaching movements. One suitable approach is the so-called rubber hand illusion (RHI, Botvinick & Cohen, 1998). In this setup, participants watch a rubber hand placed in close proximity to their own, hidden hand being stroked. At the same time, participants receive synchronous strokes on their own hand. Usually, this conflicting visual and tactile stimulation cause participants to estimate their own hand position to be closer to the rubber hand. Recent studies have shown that this adaptation is not restricted to hand space, but that adaptation can also be observed with respect to the elbow angle (Butz, Kutter, & Lorenz, 2014). Apparently, the whole representation of the kinematic chain is adapted to compensate for the conflict in hand space. It is noteworthy that this implies that the multisensory conflict between vision and touch yields adaptation in another, unaffected modality, namely proprioception. These findings are in line with the proposed ISC and the assumed striving for consistency.

The aim of the of the present study was to further investigate the adaptation of spatial representations across frames of reference in case of multisensory conflict. In the classic RHI setup, it is not possible to probe performance of the manipulated hand in an interactive task, since the illusion vanishes as soon as the real hand can be seen again. Furthermore, the setup only allows to investigate the adaptation of the proprioceptive hand estimate, while a possible effect of the visual estimate cannot be measured. To overcome this issues and to allow a measurement of the adaptation of both visual and proprioceptive estimates in case of multisensory conflict, I applied a VR setup, combining immersive VR with online motion capture (by means of the LeapMotion® sensor). Participants had to perform a complex, bimanual interaction task during which an inconsistency between seen and felt hand position was introduced on a trialswise basis. Before and after the object interaction, participants were requested to perform a self- and an external localization. Hence, compared to RHI setups, this setup allowed to study how participants use their manipulated effectors, for instance in a localization task. Furthermore, the setup allowed to investigate separately, to which degree proprioceptive and visual representations adapt. Half of the localization trials had to be performed with the hands invisible, probing adaptation in proprioceptive estimates alone. The proposed striving for consistency should yield adaptation of localization estimates for self- and external localization. Seeing the different weighting of sensory modalities in different representations, hand visibility was expected to affect localization estimates.

4.3.1. Method

The trial schedule is shown in Fig. 4.7. Each trial consisted of three stages. First, participants had to locate themselves and an external reference with a pointing gesture (pre-localization in the following). Second, participants performed a bimanual interaction task, in which they had to pick petals from a flower and to put them into a basket. During the object interaction, the visual hand representation in the VR was shifted. To compensate for this offset, participants had to adapt their actual hand position accordingly. Hence, this manipulation introduced a mismatch between vision and proprioception. Adaptation was mandatory since participants could not perform the object interaction otherwise. Third, the participants repeated the localization task,

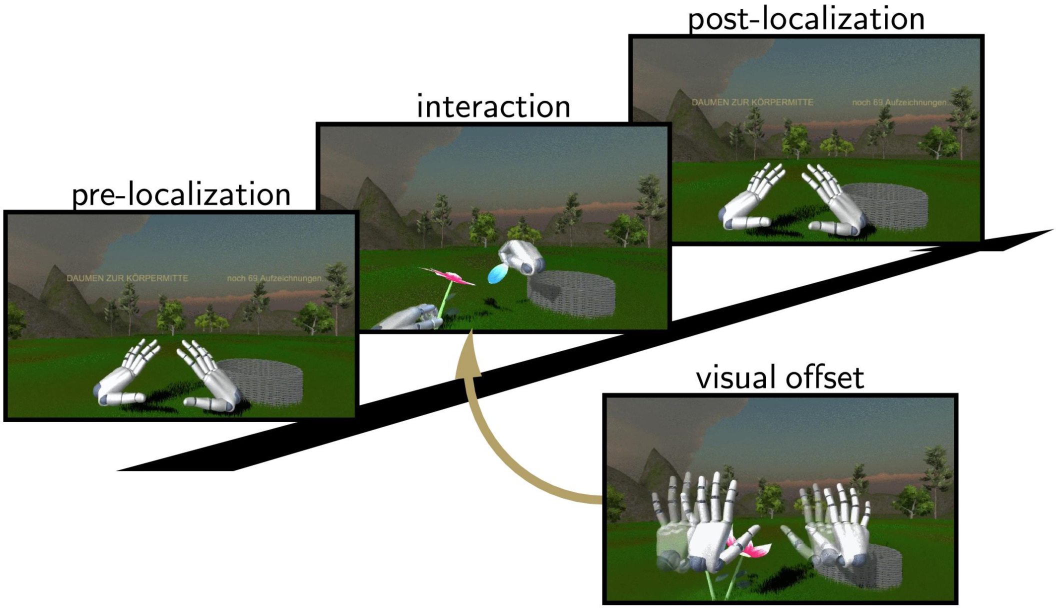


Figure 4.7.: Each trial consisted of three stages. First, participants had to locate themselves and an external reference (only the self-localization is shown in the figure). Second, participants had to perform a bimanual object interaction, during which the visual hand model was gradually shifted. The maximum amplitude of the shift on the horizontal axis is shown in the lower right part of the figure. The transparent hands indicate the actual position of the hands, the solid hands indicate the shifted position the participants saw. Please note that the transparent hands were included to show the effect of the offset, participants only saw the solid hands. Third, participants had to repeat the localization with the visual offset active. At the end of a trial, participants had to move their hands outside the sensor range and the offset was removed. The direction of the offset varied from trail to trial. In one block, the hands were invisible during the localization task, in the other block, hands were visible throughout the trial.

with the multisensory conflict active (post-localization in the following). At the end of the trial, participants were requested to move their hands out of the sensor range and the visual offset was removed accordingly. Hence, every trial started without multisensory conflict regarding the hand position. The experiment was divided into two blocks, in one block the hands remained visible during the localization task, in the other block, the hands were invisible during the localization.

To investigate possible adaptation of spatial representations due to the introduced multisensory conflict, the performance in the localization task was analyzed. The LeapMotion® sensor provides the actual joint positions of the hands, which were used to derive three dependent measures. First, the palm drift was calculated as the difference between the palm centroid in pre- and post-localization. Especially in case of localization with invisible hands, such a compensatory shift would imply an adaptation of the center of hand space. To compensate a possible drift, participants had to adapt their pointing direction, that is they had to rotate their hands differently in the pre- and post-localization. This angular disparity provides information about the compensation mechanism which applies when conflicting visual and proprioceptive information is integrated. If there is no compensation, this would imply that participants fully rely on their manipulated, visual hand representation and that the representation of the whole kinematic chain was rearranged to be in line with the manipulated information. Differences in the adaptation between conditions with visible and invisible hands would imply an adaptive weighting mechanism that considers the available information. To directly compare how much the estimated target position differed between pre- and post-localization, a positional discrepancy measure was derived. Together with the angular disparity measure, the positional discrepancy can reveal whether a possible compensation mechanism completely removes the biases introduced by the multisensory conflict. Significant results on both measures, would imply a partial compensation. If the weighting mechanism is as adaptive as assumed, the compensation should be complete in case of purely proprioceptive estimations that is self-localization with invisible hands.

4.3.2. Results

To investigate these hypotheses, the three dependent measures were analyzed with linear mixed effect models. The analysis focused on effects of hand visibility during the localization task (visible vs. invisible) with respect to the target (self or external). Aggregated data for the four combinations of visibility and localization target is shown in Fig. 4.8, the raw data is presented in Fig. 4.9. Analysis of the drift data yielded a main effect of visibility. The aftereffect in terms of an adaptation of the hand center

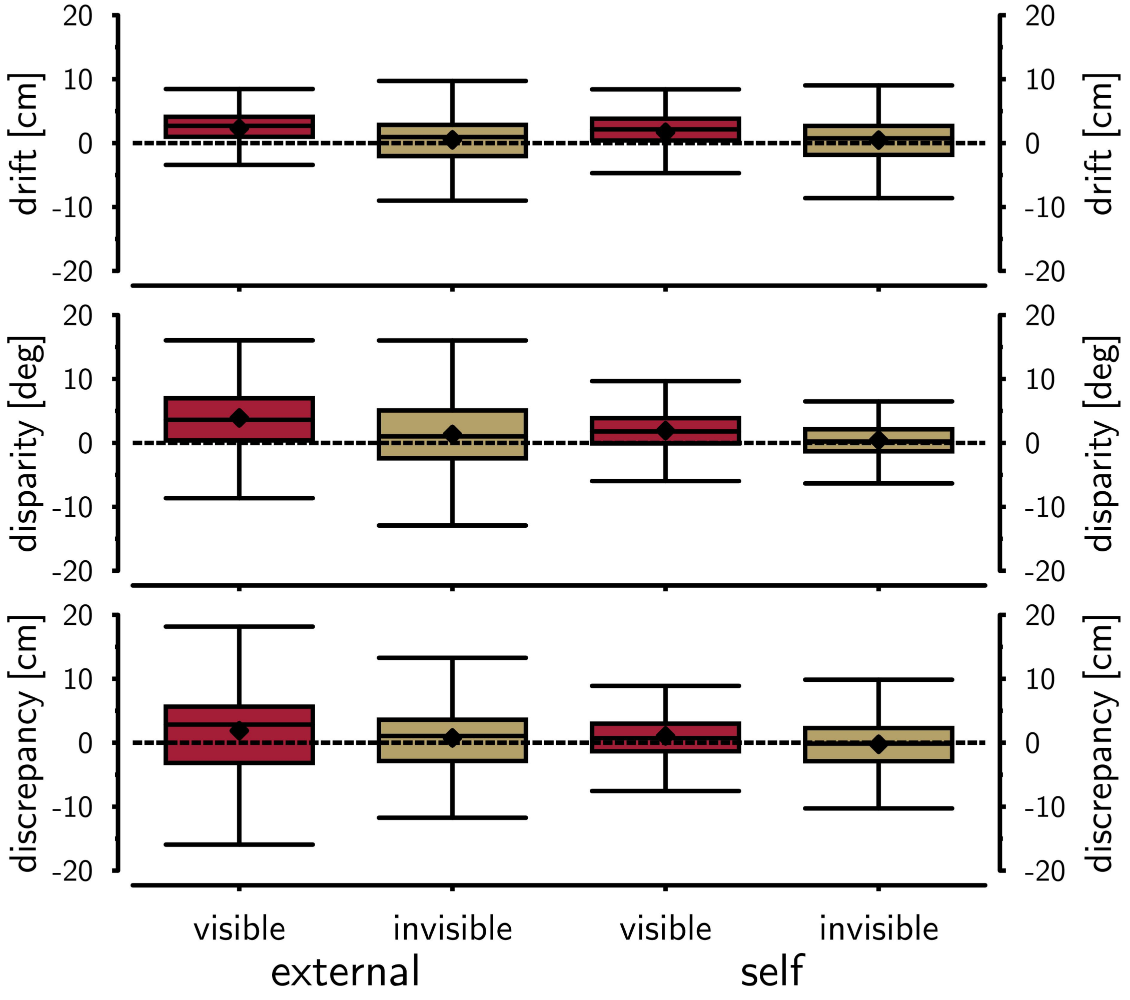


Figure 4.8.: Data aggregated with respect to the four combinations of hand visibility and localization reference. Horizontal lines within the boxes indicate the median, diamond markers refer to the mean. The lower hinge indicates the end of the first, the upper hinge the end of the third quartile. The notches cover ±1.5 of the interquartile range. All means differ from zero, with two exceptions. For self-localization with invisible hands the discrepancy score did not differ significantly from zero (p =.493), for external localizations the respective p-value only approached significance (p =.051).

was stronger in case of visible hands. Post-hoc t-tests showed that the drift differed from zero for all combinations of visibility and reference. Data regarding the angular disparity was best described by a model assuming main effects for visibility and reference. Disparity increased in case of visible hands and decreased when participants pointed towards themselves. Like the drift, the disparities differed significantly from

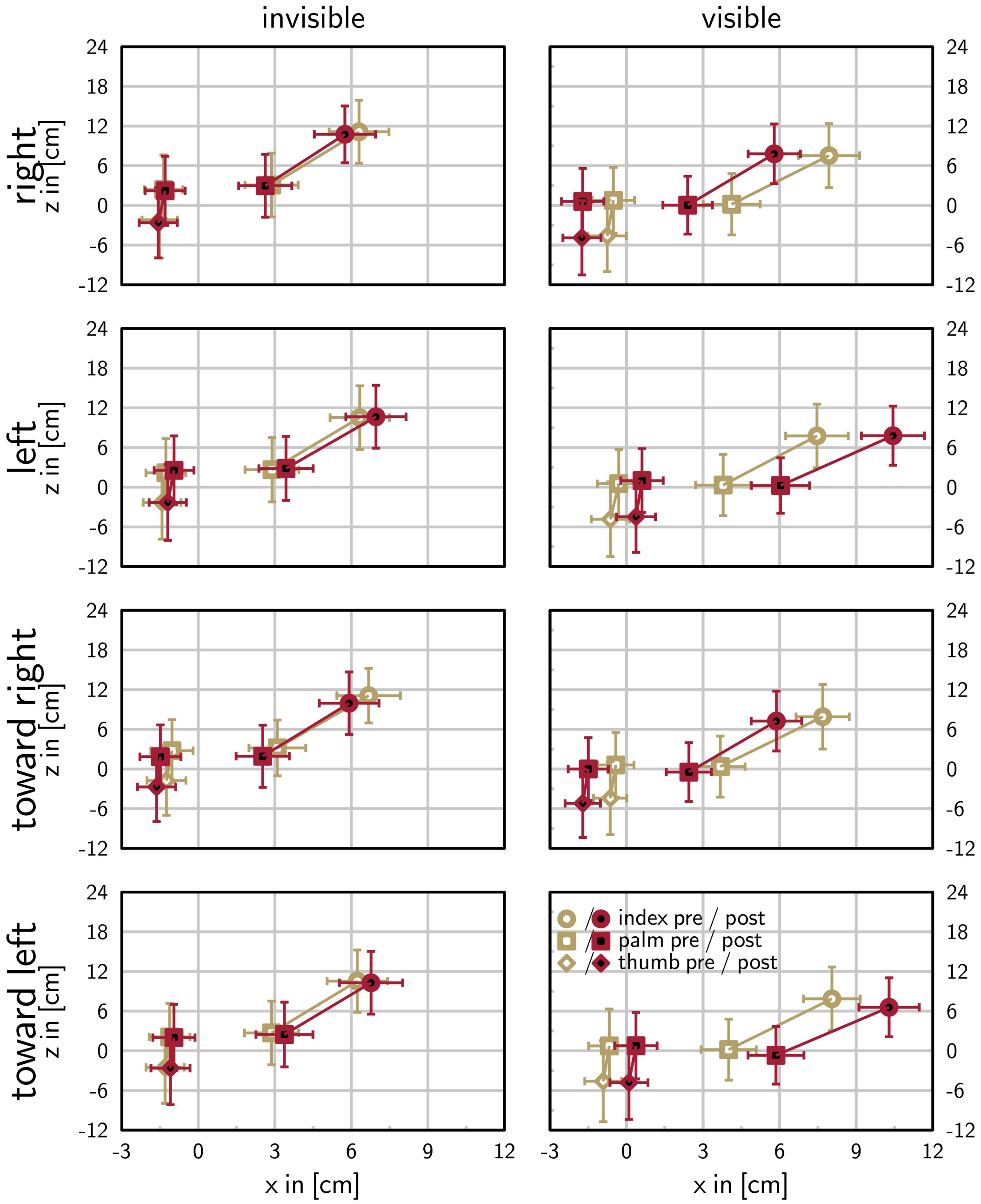


Figure 4.9.: Raw data with respect to the different offset conditions. Only the horizontal (x-axis) and sagittal (z-axis) are considered. Error bars indicate the standard error of the mean. Please note that the markers represent centroids, e.g. the yellow square indicating the palm position in the pre-loclaization was obtained as the centroid of both palms. Horizontal shifts between pre- and post-localization indicate the compensation of the visual offset. Compensation is opposed to the offset direction, hence, a visual offset to the right yields compensation to the left. The respective effects are more pronounced in case of visible hands.

zero for all combinations of visibility and reference. The pattern of results is more complex for the discrepancy, however, the best fitting model involved main effects for visibility and reference. Discrepancy was increased in case of visible hands and in case of external localization. Post-hoc t-tests showed that the discrepancy did not differ from zero in case of self-localization with invisible hands, implying a complete compensation in this condition.

For all considered measures a significant effect of hand visibility was observed. In case of visible hands, aftereffects were more pronounced, as well as the according compensation. Furthermore, visible hands yielded a stronger localization mismatch. The localization target only modulated compensation in terms of angular disparity and localization mismatch in terms of positional discrepancy. The effects of hand visibility are in line with the assumption of an adaptive weighting mechanism that tries to provide the most reliable spatial estimate given the available information. The changes in localization, induced by localized multisensory conflict regarding the hand position, imply an adaptation of spatial representations in different frames of reference, especially in those which rely on visual information. If saliency of proprioceptive information was increased - by hiding the hands and requesting a self-localization - the effects of the manipulated visual information vanished.

4.3.3. Summary

The results show how multisensory conflict in one frame of reference - in this case hand space - affects spatial representations in other, external, as well as egocentric frames of reference. Localization of the own body and an external reference changed in a way consistent with the adaptation in hand space necessary to compensate the conflict between vision and proprioception. Apparently, the different frames of reference are closely coupled, probably by means of a postural body schema (Holmes & Spence, 2004). The adaptation was modulated by the available information, stronger aftereffects and more pronounced changes in localization were observed if the manipulated visual hand models were visible. Furthermore, the effect of manipulated visual information was larger for external than for self-localization, dovetailing with previous results showing an increased weighting of visual information with increasing distance to the body (Longo & Lourenco, 2006, 2007). It seems that an adaptive weighting mechanism tries to provide the most consistent and reliable spatial estimate given the currently available data. With respect to goal directed behavior, such a mechanism seems extremely useful, since the transfer of information between different frames of reference is crucial for instance for visuo-motor control. However, the presented results only show how the mappings between frames of reference are kept consistent, not how these mappings are engaged in behavior control. This question was pursued in the study presented next.

4.3.4. Personal Contribution

Prof. Butz and I derived the paradigm. I carried out the implementation of the experiment as well as the data analysis. Data collection was done during a practical course in the winter semester 2015/2016. Prof. Butz helped me with the preparation of the manuscript and during the revision process.

4.4. Spatial Predictive Models in the Service of Action Control

Results presented in this section are submitted and are currently under review:

Belardinelli, A., Lohmann, J., Farnè, A., & Butz, M. V.: **Mental Space Maps Into the Future**. *under review*.

So far, the presented experiments have shown how task relevance biases the formation of internal models and how these models are kept consistent. As it is shown in Fig. 2.1 and as it was mentioned in the introduction, the proposed ISC is assumed to develop in the service of action control, providing internal models to drive active inference during interactions. The aim of the experiments presented here, was to investigate the predictive properties of the ISC in the service of anticipatory behavior control.

According to the model outlined in Fig. 2.1, the ISC combines sensory and motor activations in a spatial code. Such an integration seem very useful to predict the outcomes of motor actions. For instance, the activation of a certain goal state, like grasping an object should not only prepare the necessary motor commands to reach the object, but also the sensations which are likely to be experienced during the interaction. Hence, the ISC provides a common, event-oriented, and predictive format for sensorimotor activations. This conceptualization of a common, sensorimotor code is closely related to the assumptions of the *theory of event coding* (TEC, Hommel, Müsseler, Aschersleben, & Prinz, 2001a; Hommel, Müsseler, Aschersleben, & Prinz, 2001b). According to the outlined model, sensorimotor predictions are realized in terms of a continuous simulation. This simulation is expected to predict the state transition between event boundaries (see Fig. 2.1, center). This transition results in sensory changes in various frames of reference, for instance eye- and hand-centered ones. If these changes are indeed simulated during movement planning, this should result in a remapping of the involved frames of reference, that is, peripersonal space representations should be partially tuned to future sensory stimulations (see Butz, 2016). To investigate this anticipatory remapping, it is necessary to probe multisensory processing in the involved representations.

A classic method to investigate localization in peripersonal space is the *cross modal congruency* paradigm (Driver & Spence, 1998; Spence, Pavani, Maravita, & Holmes, 2004). In these setups, participants are requested to indicate the location of a sensory stimulation as fast as possible, thereby ignoring distractors in other modalities. Usually a combination of vibrotactile targets and visual distractors is used. Participants have to indicate whether their thumb or index finger received a vibrotactile stimulation which is accompanied with light flashes at either the thumb or the index finger. Response times are faster if the location of both stimuli match. Since the effect is modulated by the spatial distance between the stimuli, crossmodal congruency is usually interpreted as an effect of multisensory integration within the respective frame of reference. Hence, crossmodal congruency allows to investigate current spatial mappings between different sensory modalities. Accordingly, crossmodal congruency has been used to show the remapping of peripersonal space. For instance Holmes, Calvert, and Spence (2007) found crossmodal congruency effects in case of tool use. Here, visual stimuli at the tool tip elicited compatibility effects with vibrotactile stimulation at fingers at the same side of the tool. Besides results showing spatial remapping in case of tool use, previous findings have provided evidence for an anticipatory remapping in grasping movements (Brozzoli, Pavani, Urquizar, Cardinali, & Farnè, 2009; Brozzoli, Cardinali, Pavani, & Farnè, 2010). Brozzoli et al. combined a grasping task with a crossmodal congruency task to investigate the remapping of peripersonal space during movement control. Crossmodal congruency between the target position and the current hand position was observed already upon movement initiation. Apparently, peripersonal space was mapped into the future, yielding compatibility between visual stimulation at the target position and the future hand position. However, since the grasping task was blocked in this experiment, it is hard to judge whether this remapping reflects an online control mechanism, or a general preparation effect due to the predictability of the task. Irrespective of this issue, crossmodal congruency appears to be a suitable paradigm to investigate spatial remapping during natural object interactions.

Besides the anticipatory remapping of vision and touch, other reference frames should be involved in movement planning as well. The presented results regarding adaptation of spatial representations across frames of reference imply a close coupling of spatial representations throughout the kinematic chain. Previous eye-tracking results have shown that attention is directed towards the anticipated grasp location on an object, considering the subsequent manipulation, before the actual hand movement was initiated (Belardinelli, Stepper, & Butz, 2016). Considering the proposed model, this implies an eye-centered frames of reference is involved in the anticipatory remapping, along with peripersonal space encodings. If this is indeed the case, vibrotactile stimulation, which should result in bottom-up, attentional capture should yield a shift in gaze towards the future finger position.

To investigate these hypotheses, we combined a crossmodal congruency task with an object interaction. Besides response times for the crossmodal congruency task, we also collected gaze data.

4.4.1. Method

The applied dual task paradigm is shown in Fig. 4.10. Participants had to perform a pantomimic grasp and carry task with respect to a bottle presented on a touch screen. In order to do so, they had to touch the bottle, displace it and touch the screen again to put it down in an upright position. During the initial movement towards the bottle, participants received a vibrotactile stimulation on their index finger or thumb. Concurrently, a visual distractor was presented on the left (one third of the trials) or right (one third of the trials) side of the bottle. To investigate the effect of vibrotactile stimulation alone, in the remaining third of the trials, no visual distractor was presented. Participants were requested to verbally report the stimulated finger as fast and accurate as possible. Response times were recorded as the onset of the verbal response via speech recognition software.

Since we were interested in dynamic movement planning, we did not instruct a certain grasp type, but manipulated the orientation of the bottle, which appeared upright in half of the trials and upside down in the remaining trials. The orientation is known to

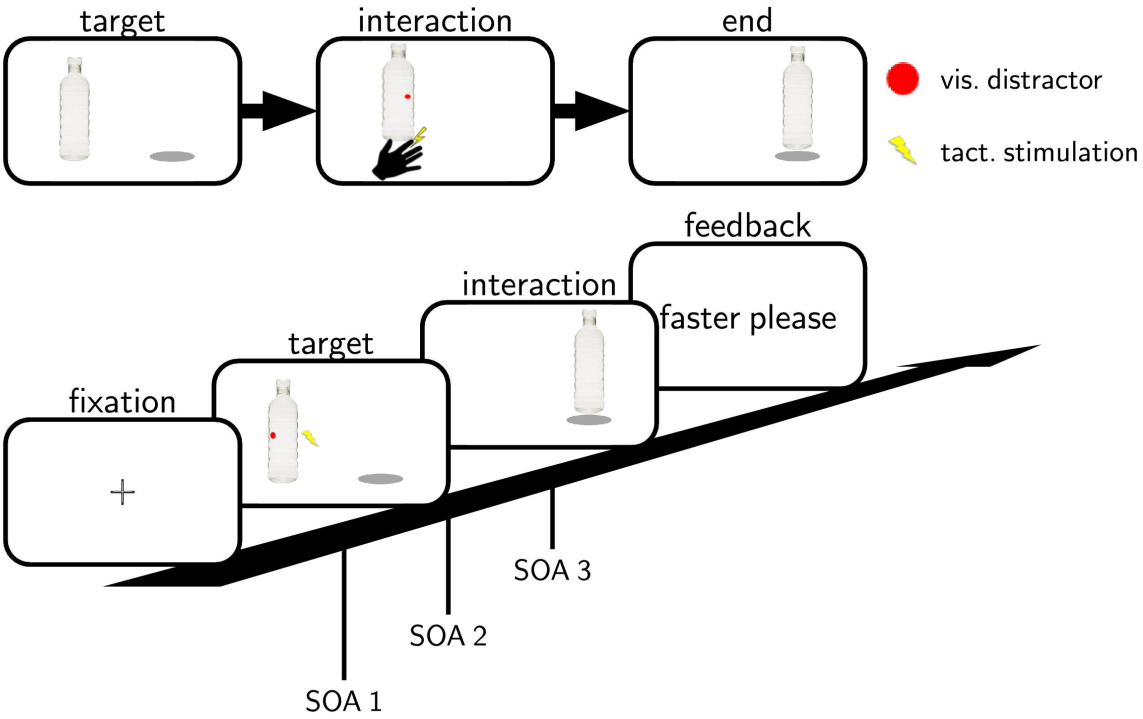


Figure 4.10.: Trial schedule in the dual task setup. The schedule for a single trial is shown in the bottom half. After the presentation of a fixation cross, the image of a plastic bottle appeared. Participants had to perform a pantomimic grasp and carry task, by touching the screen. The target position was indicated by a gray circle. During the interaction, tactile (indicated by the yellow flash) and visual stimulation (indicated by the red dot) were delivered concurrently either 200 ms after visual onset of the bottle (SOA1), at motion onset (SOA2), or 200 ms after motion onset (SOA3). Participants were requested to respond as fast as possible to the tactile stimulation by naming the stimulated finger. The upper half of the figure shows the interaction sequence in more detail. After visual onset of the bottle, which could be either oriented upright or upside down, participants performed a pantomimic grasp. After touching the screen, the bottle disappeared. It reappeared when participants touched the target location.

elicit either overhand or underhand grasps, to match a comfortable postural end state (Rosenbaum et al., 1990; Herbort & Butz, 2011). The anticipatory spatial remapping should match the end state, hence it should change with the bottle orientation. Accordingly, we expected a difference in crossmodal congruency depending on the bottle orientation. The different compatibility conditions are shown in Fig. 4.11. Furthermore, gaze data was expected to reveal a bias towards the future finger position upon vibrotactile stimulation.

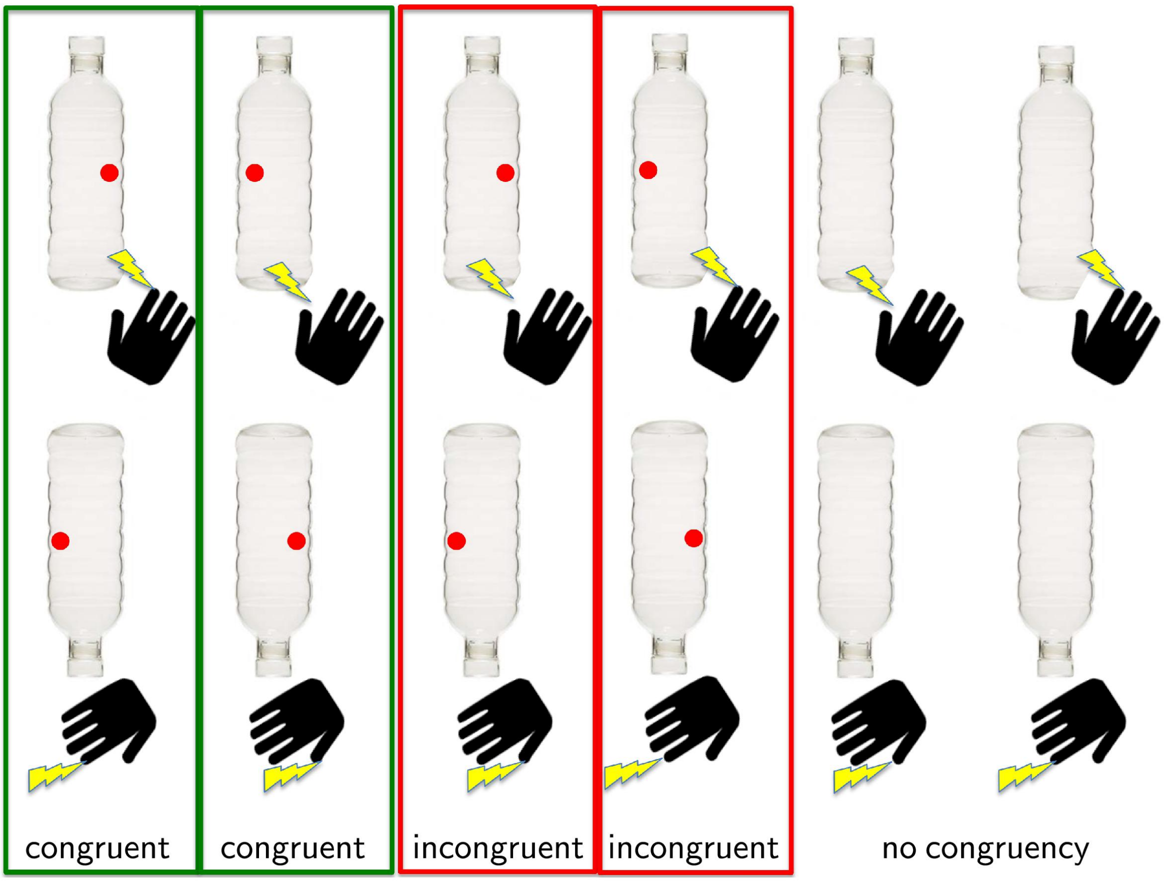


Figure 4.11.: According to the hypothesized anticipatory remapping of peripersonal space, crossmodal congruency between visual and tactile stimulation was assumed to depend on the bottle orientation. Green frames indicate the different congruent conditions, incongruent conditions are marked with a red frame. Control conditions without visual distractor were not considered in terms of congruency. This figure was created by Dr. Belardinelli and is used here with her permission.

4.4.2. Results

The predictions were largely confirmed in two experiments. With respect to response times, a significant crossmodal congruency effect was observed already at 200 ms after target onset. Furthermore, the congruency depended on the bottle orientation, indicated by a significant three-way interaction between bottle orientation, side of the visual distractor and stimulated finger. The according results are presented in Fig. 4.12. In case of upright bottles, a visual stimulation on the right side yielded faster responses if the index finger was stimulated, visual distractors on the left side yielded

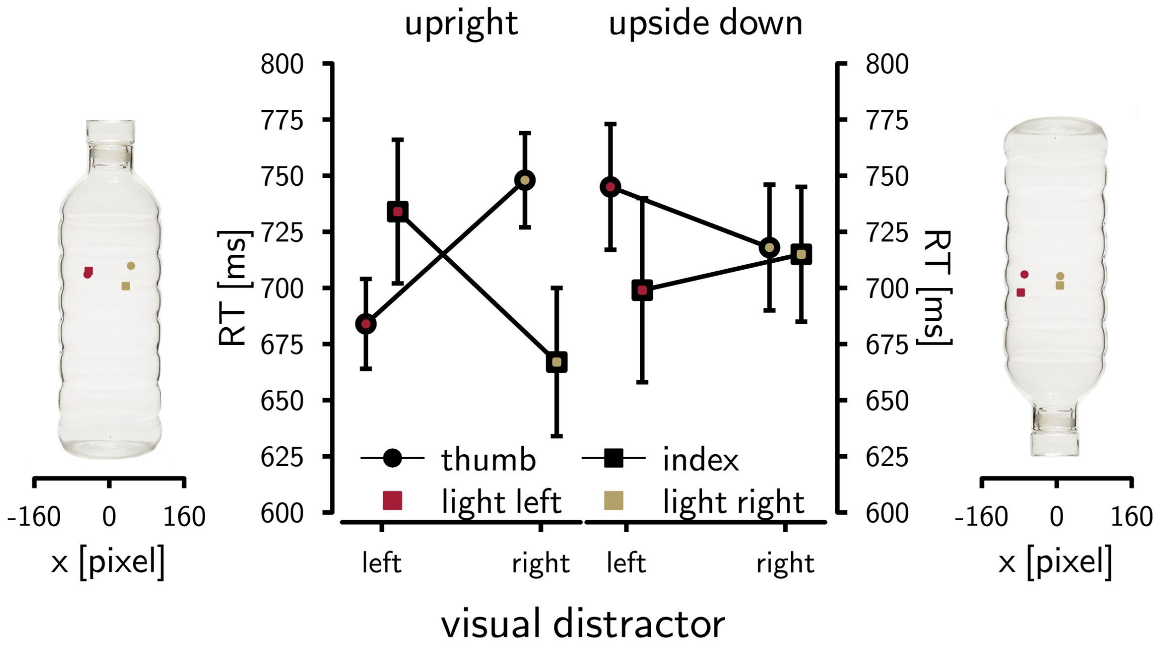


Figure 4.12.: Three-way interaction between bottle orientation, side of the visual distractor and stimulated finger with respect to verbal response times. Error bars indicate the standard error of the mean. Participants responded faster if the visual distractor location matched the future location of the stimulated finger. This congruency effect was more pronounced for upright bottles. The plots on the left and right side show the gaze centroid within the time frame of 300 to 600 ms after stimulation onset. The gaze is strongly biased towards the visual distractor (color coding), effects of spatial compatibility with the stimulated finger (different markers) are more subtle.

faster responses in case of stimulations of the thumb. This pattern was reversed for bottles presented upsight down. Hence, the crossmodal congruency occurred with respect to the future grasp position - even before the participants actually initiated the movement or rotated their hand.

Analysis of the gaze data revealed a small but significant bias towards the future location of the stimulated finger. Especially the results from trials without visual distractor show that this effect reflects an effect of the vibrotactile stimulation and cannot be accounted for solely by visual capture induced by the distractor. The respective results are shown in Fig. 4.13. Effects on oculomotor behavior occur as early as the crossmodal congruency, implying that remapping of peripersonal space and spatial

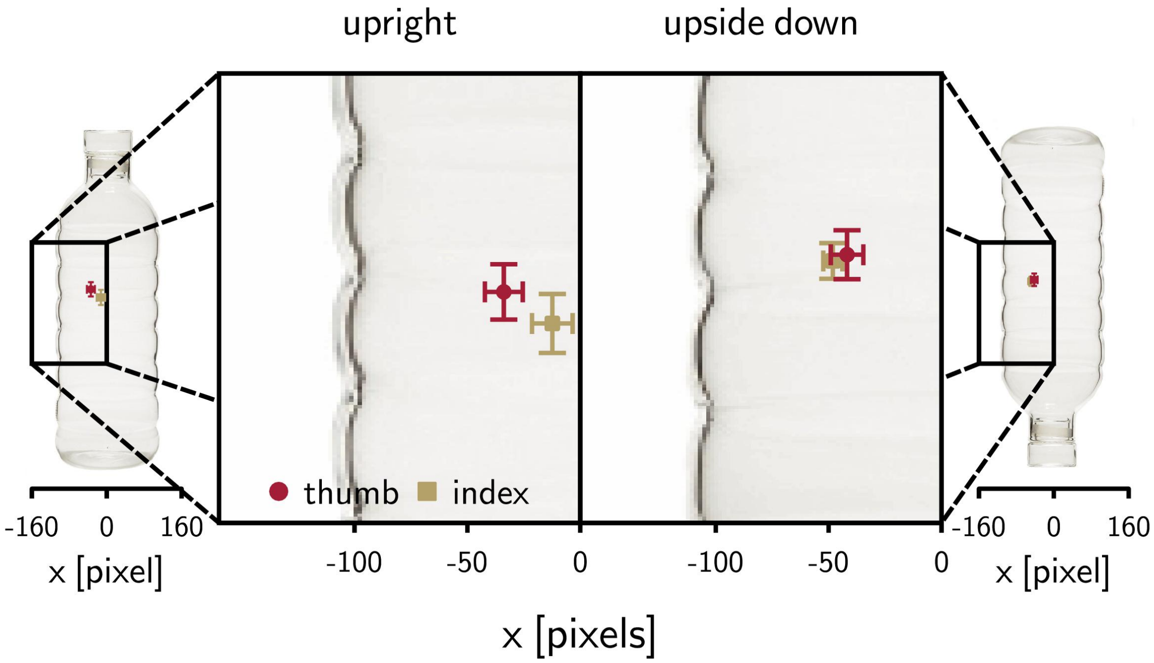


Figure 4.13.: Gaze centroids within the time frame of 300 to 600 ms after stimulation onset in trials without visual distractor. Error bars indicate the standard error of the mean. The centroids show a slight but significant bias towards the future location of the stimulated finger.

predictions in eye-centered frames of reference occur at the same time.

The second experiment mostly confirmed these results, however, the time-course of the crossmodal congruency effect showed a notable difference. While in the first experiment crossmodal congruency was observed for early stimulations, crossmodal congruency in the second experiment was more pronounced for late stimulations. Apparently, remapping of peripersonal space was modulated by the temporal predictability of the vibrotactile stimulation.

4.4.3. Summary

The results show how peripersonal space is dynamically remapped in anticipation of the future hand position in an object interaction task. The orientation-dependent crossmodal congruency shows how sensory processing is directed towards the future, yielding compatibility between vision and touch at the upcoming hand position.

This remapping can occur very early, 200 ms after the presentation of the target stimulus. However, results from the second experiment show that the remapping not necessarily occur before movement onset. Apparently, the initiation of the remapping depends on task parameters, like the expected time point of the vibrotactile stimulation.

With respect to the proposed model, the results show how the ISC can be used to remap peripersonal space into the future. Since the results imply crossmodal congruency with the future hand location, even if the actual hand did not even start to move or to rotate, the prediction seems to focus on the next event boundary, that is, the grasp on the target object. Eye-tracking results show how this remapping goes along with spatial predictions in an eye-centered frame of reference. The gaze shifted towards the future finger positions, hence the future finger position was also converted into a positional estimate relative to the eyes. Together, these results show that anticipatory behavior control is indeed realized in terms of a simulation of the to be expected sensations. However, it remains open whether this simulation exclusively focus on the event boundary, or if it is realized in terms of a continuous simulation of the hand trajectory. Modulations by the SOA can be interpreted as indicators for a continuous simulation, but a throughout investigation requires a comparison of crossmodal congruency effects for the actual and future hand position.

Different theories on anticipatory behavior control propose that the cognitive system is essentially predictive - anticipating desired goal states and employing these predictions in movement planning (e.g. Hoffmann, 1993; Hommel et al., 2001a; Prinz, 1990). In order to so, a common code is necessary, which relates current and predicted sensory stimulation with motor commands (for instance the common code proposed by TEC, Hommel et al., 2001a, 2001b). The presented results show how such a common code can be realized in terms of a simulation within the proposed ISC. As it is shown in Fig. 2.1, the proposed ISC relates different frames of reference and accordingly sensory modalities. During movement planning, spatial predictions in different frames of reference, for instance eye- and hand-centered ones, are derived. These predictions can be considered as a sensorimotor simulation, activating the sensory systems which are also used to process the respective stimuli. Hence, predictions are not limited to a single frame of reference and due to the activation within the sensory system, stimulus processing is biased towards the expected location.

According to theories of embodied cognition, simulation is a core form of neural computation (Barsalou, 2008). Sharing a common representational format, these simulation processes are assumed to be involved in various cognitive abilities. The presented results imply that simulation yields sensory activation across frames of reference and modalities. If this is a general feature of the simulation mechanism, it should not be restricted to movement control. In the study presented in the next section, this assumption was tested with respect to mental rotation.

4.4.4. Personal Contribution

Dr. Belardinelli, Prof. Butz and I derived the paradigm together, based on the setup applied by Brozzoli et al. (2010). Data collection was done by Stephanie Blumenschein and Gina Hermann as part of their bachelor theses. Together with Dr. Belardinelli, I implemented the setup, particularly the speech recognition, the vibrotactile stimulation and the respective integration into the Matlab code of the experiment. I assisted Dr. Belardinelli and Prof. Butz during the preparation of the manuscript, especially with respect to the supplementary material and the figures.

4.5. Mental Rotation as Spatial Simulation

Results presented in this section have been published in:

Lohmann, J., Rolke, B., & Butz, M. V. (2017): **In touch with mental rotation: Interactions between mental and tactile rotations and motor responses**. *Experimental Brain Research*, 235(4), 1063-1079. doi: 10.1007/s00221-016-4861-8.

As it was argued in the introduction, the ISC proposed in Figure 2.1, is rooted in sensorimotor experience. After sufficient experience, the ISC can be used to simulate event transitions by activating modal codes, that is, in terms of a sensorimotor simulation. If this is indeed the case, concurrent modal stimulation with similar spatiotemporal dynamics should interfere with the ongoing simulation. Furthermore, the simulation should also activate motor areas, facilitating responses which are congruent with the dynamics of the simulation. There is indeed already some evidence for the sensorimotor grounding of cognitive functions. For instance with respect to mental rotation - the ability to mentally rotate an observed or imagined object - the involvement of the motor system has been shown in various studies (Gardony, Taylor, & Brunyé, 2014; Wexler, Kosslyn, & Berthoz, 1998; Wohlschläger & Wohlschläger, 1998). Hence, mental rotation lends itself as a suitable task to investigate possible interactions between modal codes with similar dynamics during the execution of a cognitive task, as well as possible motor activation induced by the task. To investigate these hypotheses I conducted two experiments, combining a mental rotation task with concurrent modal stimulation.

In their seminal study, Shepard and Metzler (1971) let their participants compare the parity of two three-dimensional block figures, which were presented in different orientations. Shepard and Metzler observed a linear relationship between response times and angular disparity of the stimuli. This so-called disparity effect occurs in many different mental rotation tasks, applying different stimuli, including alphanumeric characters (Cooper & Shepard, 1973), images of natural objects (McMullen & Jolicoeur, 1990), abstract geometric figures (Liesefeld & Zimmer, 2013), pictures of body parts (Ionta et al., 2007; Parsons, 1987a, 1987b), or tactile forms in case of blindfolded or blind participants (Marmor & Zaback, 1976). The disparity effect was interpreted as the footprint of a mental rotation process that simulates an actual rotation, which takes longer in case of larger angular disparities. This possible overlap between mental and physical rotations was investigated in several studies. Even if the results are mixed, the overlap could be reliably shown in tasks that can be accomplished either with the help of physical rotations, or purely mentally (Gardony et al., 2014; Wexler et al., 1998; Wohlschläger & Wohlschläger, 1998). Further evidence for an involvement of the motor system in mental rotation comes from dual-task interference paradigms (Wexler et al., 1998), and studies on the effects of action effects on mental rotation (Schwartz & Holton, 2000). These results are further supported by neuroimaging data. It has been shown that the same parietal brain areas that encode spatial real world transformations and partially also the same motor areas that produce according motor activities are engaged in mental rotation (see Zacks, 2008, for an overview). Together, these findings on mental rotation imply a sensorimotor simulation process which realizes the rotation. The behavioral as well as the neuroimaging results further imply a strong spatial component with respect to the underlying representation. Both assumptions are in line with the proposed ISC.

If mental rotation is indeed realized via a spatial simulation that draws on sensorimotor resources, different modalities should be able to activate, access, and interact with it, as long as they match the activation pattern. For instance a tactile, rotational stimulation should affect an ongoing mental rotation. Such an interference should not occur if mental rotation is realized within a single modality, or if the spatial simulation does not engage modal activations. To provide further evidence for the realization of mental rotation within the ISC, I investigated mental rotation performance during concurrent, rotational tactile stimulation in two experiments. Since the assumed ISC relies on sensory and motor codes, the model predicts the activation of motor codes which correspond with the simulated rotation. Hence, I also investigated a possible congruency effect between mental rotation direction and response side.

4.5.1. Method

The applied dual-task interference paradigm is shown in Fig. 4.14. To investigate selective interference between tactile and mental rotation, I combined a primary mental rotation task with a secondary tactile change detection task. In the mental rotation task, participants had to judge the parity of rotated letters. In the first experiment, participants responded to mirrored letters with their right index finger, while they responded to letters in their canonical orientation with their right middle finger. This response mapping was varied within participants in the second experiment. In the secondary task, participants had to detect a change within the rotation direction of a tactile stimulation applied to their palm. The tactile stimulation device is shown in Fig. 4.14 (bottom). If mental and tactile rotation are indeed processed in - at least partially - overlapping codes, there should be selective interference between tactile and mental rotation. Butz, Thomaschke, Linhardt, and Herbort (2010) could show that visual and tactile information are integrated within a head-centered frame of reference. I assumed faster response times if mental and tactile rotation direction matched within in this frame of reference, while they should be elevated in case of mismatch. Besides this compatibility between visual and tactile rotation, the hypothesized spatial code should also yield activation in motor areas during mental rotation. If this is indeed the case, the directional information conveyed by the mental rotation should facilitate matching response directions. Clockwise mental rotations should facilitate responses to the right (with the middle finger), while counterclockwise mental rotations should facilitate responses to the left (with the index finger). To sum up, based on these hypotheses an interaction between mental and tactile rotation direction, as well as an interaction between mental rotation direction and response side was predicted.

4.5.2. Results

The hypotheses were largely confirmed in both experiments. Noteworthy, in both experiments the sought interaction between mental rotation direction and tactile rotation direction only occured in trials that involved a change in the tactile rotation direction.

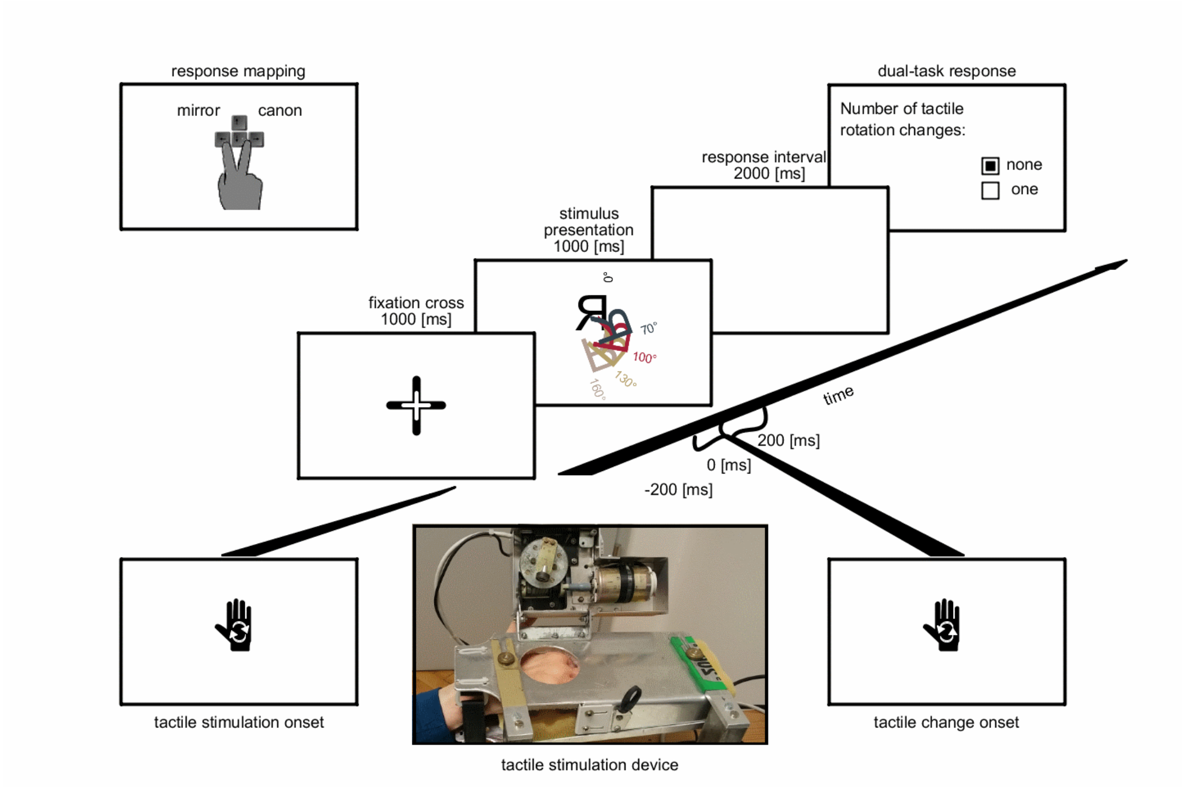


Figure 4.14.: Schedule of a single trial in the applied dual task setup. Tactile stimulation started with the presentation of the fixation cross. The target letter appeared one second later. Participants had to perform a parity judgment. Letters always appeared at the center of the screen, the slight offset between the shown disparities was only included for the sake of visibility. After completing the parity judgment, participants had to indicate whether the direction of the tactile stimulation changed during the trial. The tactile stimulation device is shown in the lower center. Tactile stimulation was always applied to the left hand of the participants and responses were given with the right hand. The response mapping for the parity judgment is shown in the upper left. In the second experiment, this mapping varied between blocks.

In trials where the tactile rotation direction stayed the same, no significant interaction was found. The results from both experiments regarding the two relevant interactions between mental and tactile rotation direction, as well as between mental rotation direction and response side are shown in Fig. 4.15. To verify that the interaction between mental rotation direction and parity indeed reflects a compatibility between mental rotation direction and response side, a direct variation of the response mapping was required. Otherwise a habitual response preference, or semantic effects like markedness (a spatial-semantic association between parity and response side, e.g.

Nuerk, Iversen, & Willmes, 2004), could not be ruled out as explanation. In the second experiment, I varied the response mapping within participants. Here, participants had to complete two sessions of the dual-task setup. In one session the response mapping was the same as in the first experiment, in the second one, the response mapping was reversed. If mental rotation indeed facilitates directional responses, the results should yield a significant three-way interaction between response mapping, response side and mental rotation direction. This three-way interaction was indeed found in the second experiment, both in trials without a change in tactile rotation direction, as well as in trials involving a change in the tactile rotation direction. However, compared to the first experiment, the effect was more pronounced in the latter case. The interactions between mental rotation direction and response side, as well as the interactions between response mapping, mental rotation direction and response side were significant with respect to error rates as well in both experiments. With respect to the compatibility between response side and mental rotation direction, this was true for trials without a change in tactile rotation direction, as well as trial involving a change in the tactile rotation direction.

4.5.3. Summary

The two reported experiments show an overlap between mental rotation and concurrent tactile rotational stimulation. In trials where the rotation direction changed, a clear interaction between the mental and the new tactile rotation direction was visible: Responses were faster and more accurate if the rotation directions matched, compared to trials were they did not match. Since this interaction was not observed in trials without changes in the tactile rotation direction, it seems that the tactile rotation direction is not integrated automatically in the mental rotation process. Apparently, the tactile rotation direction had to be salient to effectively interact with the ongoing mental rotation.

Besides the compatibility between mental and tactile rotation direction, the results showed a compatibility effect between mental rotation direction and response side. Participants responded faster if they had to mentally rotate the stimuli towards side

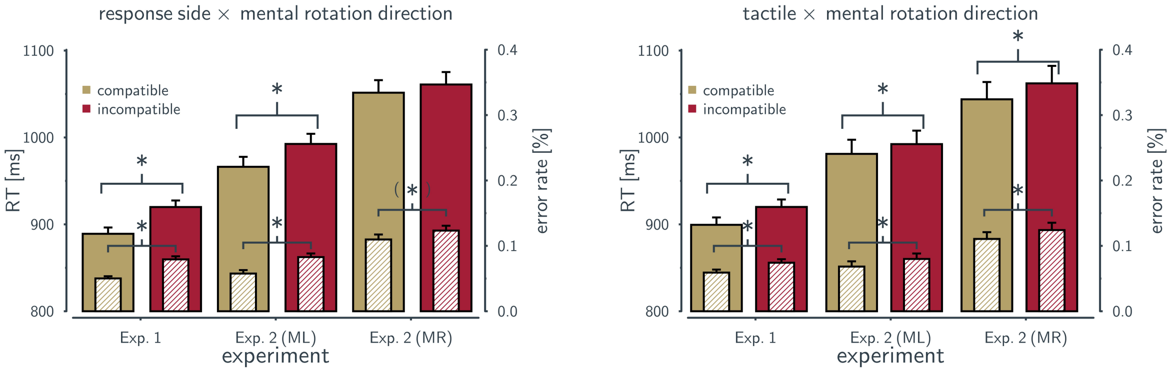


Figure 4.15.: Compatibility effects between mental rotation direction and response side (left panel) and between mental and tactile rotation direction (right panel) with respect to response times and error rates for both experiments. Right key responses were considered to be compatible with clockwise mental rotations, while left key responses were considered compatible with counterclockwise mental rotations. The response mapping varied in the second experiment. In one session participants responded with the left key to mirrored stimuli (mapping ML, same as in the first experiment). In the other session, this mapping was inverted (mapping MR). Response times are shown as solid bars, shaded bars refer to error rates. Error bars indicate the standard error of the mean. Significant differences with $\alpha = .05$ are indicated by an asterisk. Numerically, the compatibility effects are comparable across experiments, the overall increased response time in the second experiment is due to the more extreme disparities. Please note that data in the left panel combines change as well as no-change trials, while for the data in the right panel only trials where the tactile stimulation direction changed were considered.

compatible with the parity response. This effect was visible in both change and no-change trials. Furthermore, the variation of the response mapping in the second experiment, and the according three-way interaction between response side, mental rotation direction and response mapping showed that this effect is not due to a mere semantic mapping, for instance between mirrored and left - even though the mapping of left and mirrored and right and canonical seems more natural (Casasanto, 2009; Natale, Gur, & Gur, 1983) and yielded stronger compatibility effects than the inverse mapping. The observed compatibility implies a bidirectional mapping between mental rotation and motor activation. This finding complements earlier studies that have shown that motor activation can facilitate compatible mental rotations (e.g. Janczyk, Pfister, Crognale, & Kunde, 2012; Wang, Proctor, & Pick, 2003).

The selective interference between mental and tactile rotation and the facilitation of spatially compatible motor responses through mental rotation are in line the proposed ISC. These findings are difficult to integrate into models assuming a unimodal, for instance visual, basis for mental rotation, or which assume mental rotation to be completely abstract. From an embodied perspective, the observed overlap can be accounted for by considering the development of the ability to mentally rotate objects. Rotating a physical objects yields various sensory effects, including visual, tactile and proprioceptive changes, which are associated with certain motor commands. The consistencies within these modal activations are integrated through the ISC. With increasing practice, the activation of the ISC by means of an intended rotation, can initiate the simulation of the outcome, without engaging in overt motor activity. Even if this simulation is more abstract than an actual physical rotation, it recruits the same sensorimotor systems which are involved in physical rotations that is the simulation remains grounded, yielding compatibility effects between rotational stimulations in different sensory modalities, and compatibility effects with respect to motor commands. Seeing the apparent abstraction of the rotation model compared to an actual physical rotation, which allows the simulation of arbitrary rotations, the ISC appears to convey conceptual knowledge regarding rotations. If this is indeed the case, mental rotation can serve as one example for the grounding of a higher cognitive function within the ISC.

4.5.4. Personal Contribution

Prof. Butz, Prof. Rolke and I derived the paradigm together. I implemented the experimental procedure as well as a low-level interface for the tactile stimulation device. Data collection was done by Simone Kurek. I performed the data analysis. Prof. Butz and Prof. Rolke helped me during the preparation and revision of the manuscript.

5. Discussion

According to the model presented in Fig. 2.1, a constantly operating inference mechanism yields predictive models, relating sensory events and motor activation in an ISC. The four reported studies aimed at evaluating certain predictions derived from the outlined model.

Model identification is assumed to focus on modalities which are behaviorally most relevant in a certain context. This focus should result in a top-down bias, favoring the respective modality in case of multisensory conflict. This prediction was tested with respect to multisensory integration, which is usually described in terms of a maximum likelihood integrator, combining sensory signals based on the reliability of the individual sources, that is, in a bottom-up fashion (Ernst & Banks, 2002). In the presented study, participants had to perform a grasp and carry task whose success depended on an accurate grip aperture. The results yielded a clear bias toward task relevant proprioceptive information, since unimodal visual estimates of the object size were shifted in the same direction as proprioceptive estimates. Hence, the results confirmed the expected, yet previously not shown top-down bias in multisensory integration.

Besides the top-down modulation due to task requirements, the results from the multisensory integration study also showed that the participants could not separately retrieve visual and proprioceptive estimates regarding the object size. Apparently, the cognitive system tried to maintain a consistent representation of the target object across modalities. According to the proposed model, this striving for consistency is a general property of the model identification and inference mechanism and should not be limited to a single representation, or a single frame of reference. This prediction was tested in a second study, investigating the effects of multisensory conflict across frames of reference. Spatial body representations covering the peripersonal space are closely coupled to enable fast and reliable interactions (McGuire & Sabes, 2009). In order to do so, a consistent mapping between these representations is required (Butz et al., 2007; Ehrenfeld, Herbort, & Butz, 2013; Glenberg et al., 2013). To maintain consistency, conflicts in a single representation should yield adaptation across frames of reference, otherwise estimates in the different frames of reference would not correspond. Indeed, previous studies have shown that conflict in hand space yields adaptation in elbow angles in a manner consistent with the shift in hand space (Butz et al., 2014). In the reported study, this adaptation was further investigated, contrasting the adaptation of external and self-representations as well as the saliency of the manipulated information. Participants performed an object interaction during which proprioceptive and visual hand position were dissociated. Before and after the interaction they had to located themselves or an external reference, either with their hands visible or invisible. If there is indeed a striving for consistency operating across frames of reference, then the conflict in hand space should yield adaptation for the tested positional estimates, especially for external reference which should rely more on the manipulated visual information than self-localization (Longo & Lourenco, 2006). The results confirmed this hypothesis: shifts in positional estimates were observed for both self- and external localization, whereby the effect was more pronounced in the case of visible hands. Hence, the data is in line with the assumed striving for consistency, which apparently links spatial representations across frames of reference.

Top-down biases towards task relevant information and a striving for consistency are assumed to be fundamental for the identification and maintenance of the proposed ISC. The proposed model also provides predictions regarding model inference, that is, how the acquired ISC is used in anticipatory behavior control. The ISC is assumed to predict salient changes in the sensory input, so-called event-boundaries (Zacks & Tversky, 2001; Zacks & Swallow, 2007), and the motor activity necessary to realize them. The prediction is realized in terms of a simulation regarding the upcoming sensory signals. Since this simulation is assumed to be grounded within the representations that also process the actual sensory signals, it should yield at least a partial remapping of the respective representation at the location where the event is expected to unfold. This anticipatory remapping was investigated in a study combining a crossmodal congruency task (Spence et al., 2004) with pantomimic object interactions. The results showed a clear crossmodal congruency effect for the future hand position at the target object. Since this effect was clearly visible already at movement onset, the results are in line with the assumption of an anticipatory remapping of spatial body representations as a part of an event-oriented prediction.

Motor planning seems to be realized in terms of an event-oriented prediction involving pre-activation of upcoming sensory events in the respective representation. According to the proposed model, the prediction of the event involves a simulation of the sensorimotor dynamics necessary to generate it. This mechanism should not be restricted to motor planning, but should also be active during the internal simulation of an effect, without overt motor activation. Such internal simulations are assumed to be involved in higher cognitive functions, like mental imagery (Jeannerod, 1995; Barsalou, 1999), or mental rotation (Wexler et al., 1998). According to the proposed model, this simulation is realized by the ISC by activating sensorimotor codes. If this is indeed the case, modal activities – which are not necessarily associated with performing the intended action, but which exhibit similar dynamics as the simulation – should affect the internal simulation. This assumption was tested and confirmed in a dual-task paradigm combining mental rotation and concurrent tactile change detection for a rotating tactile stimulus. Responses were faster when the direction of both rotations matched, as compared to cases were they mismatch. This integration was not automatic. Reliable congruency effects were only observed when the tactile rotation direction changed. Apparently, tactile rotation had to be salient in order to interact with the ongoing mental rotation. Besides this congruency between tactile and mental rotation, the results yielded a strong spatial compatibility effect regarding the response. Participants were faster to perform a right-sided response after clockwise mental rotation. Similarly, left-sided responses were faster in case of counterclockwise rotations. These results are in line with the proposed ISC, since the simulation is assumed to activate spatially compatible motor codes as well.

In sum, the referred experimental results confirmed central predictions regarding the identification, maintenance and application of the ISC and its possible involvement in higher cognitive functions. More precisely, the results show how (i) a top-down bias for task relevant modalities affect multisensory integration, (ii) a striving for consistency influences the integration of sensory information across frames of reference, (iii) the ISC integrates various modalities via a spatial code, (iv) the ISC supports anticipatory motor control by providing event-oriented predictions, and (v) the ISC is engaged in internal simulations, supporting higher cognitive functions like mental rotation. The proposed model provides an account how a rather abstract, essentially spatial, code evolves from sensorimotor interactions. Seeing the possible contribution to higher cognitive functions, the ISC might be considered as a precursor of symbolic representations. However, even if the presented results are in line with the outlined model, several open questions remain. First, I argued that the ISC preserves propositional spatial information, which can support higher cognitive functions. So far, mental rotation remains the only higher cognitive function investigated in this respect. Hence, it remains open whether spatial models indeed play a causal role in other forms of higher cognition. Second, the presented results only allow a qualitative evaluation of the outlined model. To show that a spatial format is indeed suitable for an abstraction from sensorimotor activations, a quantitative implementation of the proposed ISC is desirable. These two issues are shortly discussed in the next two section, an outlook concludes this thesis.

5.1. Spatial Grounding of Cognitive Functions

According to embodied cognition, simulations are a central form of neural computation (Barsalou, 2008). The outlined model highlights the role of the ISC in these simulations. This assumption is in line with recent theories that consider the role of spatial models for instance with respect to deductive reasoning (Knauff, Mulack, Kassubek, Salih, & Greenlee, 2002; Ragni & Knauff, 2013). These approaches highlight the role of relational, spatial knowledge for the transformation of premises into a model of the relationships between the referred entities. Relations like “left of”, “in front”, or “behind” provide a structure for the referred entities and in turn allow to evaluate propositions regarding them. Even non-spatial relations like “earlier - later”, or "better - worse" seem to be mapped onto spatial relations (Gattis & Holyoak, 1996). Together, these findings imply a central role of spatial models in propositional inference in particular and spatial reasoning in general. According to the outlined model, these models are identified through sensorimotor interaction. A direct test of this assumption, however, requires the manipulation of the sensorimotor contingencies that convey the spatial relations.

5.2. An Artificial Spatial Problem Solver

According to the proposed model, sensorimotor activations in different frames of reference are mapped into a common, spatial code, the ISC. Learning is assumed to focus on event boundaries and the sensorimotor dynamics necessary to establish them. During inference, the ISC provides a prediction regarding the necessary dynamics to realize a certain state and guide goal-directed behavior. The ISC is assumed to preserve spatial relations, which might be reused in other cognitive task than behavior control. Accordingly, an implementation of the ISC would require three core components: (i) a common spatial code, providing a mapping between different modalities across frames of reference, (ii) an event detector, which segments activations within this common code, and (iii) a simulation mechanism, which enables the simulation of sensorimotor dynamics necessary to reach a certain goal state. Different quantitative models are available to realize these components, however, so far they have not been used in an integrated architecture.

With respect to the common spatial code, gain fields have been proposed as a possible implementation. Considering the population level, gain fields provide a multiplicative modulation of the baseline response of a single neuron. This yields a modulation of the relative contribution of a neuron to the population output, which is read out by the next neural layer. Different types of neural signals produce gain fields (see Blohm & Crawford, 2009, for an overview), which implies that they are a general computational mechanism in sensorimotor processing. Gain field-like structure might provide a biologically plausible implementation of a common spatial code, however, without a learning mechanism they can only reflect current spatial relations. Predictive, Bayesain models have been discussed as suitable learning architectures in this respect (e.g. Friston, 2005). The *modular modality frame* (MMF) model (Ehrenfeld & Butz, 2013; Ehrenfeld et al., 2013) is one example of such an architecture. It models the maintenance of a postural and visual representation of an arm in terms of neural population codes. Besides providing a predictive spatial model which can account for movement planning, MMF implements a striving for consistency that maintains a consistent structure of the mappings in case of sensor errors or noise. Hence, in line with the reported results on consistency in spatial mappings across frames of reference, the MMF architecture provides a computational framework to represent a common spatial code, its maintenance, and its application in movement planning.

Predictive Bayesian models have also been discussed as possible implementations of an event detection system. Such models provide a description for the processing and prediction of the sensorimotor flow. In case of mismatch between model and current information, an error signal arises. According to different models (Reynolds, Zacks, & Braver, 2007; Zacks, Kurby, Eisenberg, & Haroutunian, 2011), this error signal might be suitable to detect event boundaries. From a formal point of view, these models propose that event boundaries are characterized by non-uniform transition probabilities. While observations within a certain event are highly predictable, the predictability is greatly reduced at the beginning of a new event. This approach provides a sound description for the inference process, since it can account for the detection of event boundaries given a set of learned predictive models. Hence, predictive Bayesian models offers an account for the identification of the ISC as well as its application in goal-directed inference.

As mentioned above, the ability to generate internal spatial models seems a core component of imagery as well as reasoning (Ragni & Knauff, 2013; Winter, Marghetis, & Matlock, 2015). The quantitative implementation of the ISC preserves spatial relations and might be the basis for a general spatial simulator, when it is combined with an inference mechanism. So far, there are only few implementations of probabilistic inference mechanisms. One example from the visual domain is the so-called *Bayesian program learning* (BPL, Lake, Salakhutdinov, & Tenenbaum, 2015). BPL learns stochastic models to represent visual concepts in a compositional manner, based on parts, subparts, and their spatial relations. Besides the ability to generate visual models, BPL is also able to identify likely generative models for a certain stimulus. Since BPL can be applied to any compositional structure, it seems well-suited to model the envisioned spatial inference mechanism. Here, the compositional structure would consist of the hierarchical event-structure realized within the ISC. Visual parts and subparts can be replaced with goals (e.g. consume an object) and subgoals (reach for an object). These components should allow to generate sequential actions towards a certain goal state. This can be considered as a first step towards an artificial spatial reasoner, even if it remains open how this spatial model generation generalizes to other tasks than behavior control.

To sum up, a combination of the MMF model with an event detector could provide a quantitative implementation of the proposed ISC. Combined with BPL, this architecture might allow to model basic spatial inference. However, the evaluation of such a model would require a sufficiently complex environment as well as an agent that is able to interact with it.

5.3. Outlook

According to embodied cognition, cognitive functions can be described as mental simulations, rather than the manipulation of a set of symbols. Understanding how these simulations operate, how the underlying models are acquired, and how meaning is conveyed by simulations remain central questions within the context of embodied cognition. The presented model and the according experiments highlight the role of spatial models. While evidence for the formation, maintenance, and application of an ISC was provided, its possible role in higher cognitive functions remains illusive.

If understanding spatial relations is indeed fundamental for deductive reasoning and propositional inference, a manipulation of the perception of these relations should affect these cognitive functions. Manipulations of these relations is hard to achieve in real world setups, since they reflect environmental properties. The derived VR setups provide a possible solution, because they allow to manipulate action-effect contingencies as well as environmental properties in a systematic manner. If exploration of such altered environments indeed affects reasoning, this would be direct evidence for the causal role of a sensiromotor grounded ISC in higher cognitive functions.

Besides behavioral experiments, the quantitative modeling of the ISC and the according simulation mechanism will allow a more detailed study of the proposed inference mechanism. Again, VR setups seem a suitable tool to investigate the formation and application of an ISC, embodied by an artificial agent. As it was proposed by Butz (2016), VRs can provide a complex flow of sensorimotor information and allow systematic manipulations of environmental properties.

Hence, further investigation of the ISC might provide valuable insights for the implementation of an artificial, spatial reasoner. Such a system might yield an answer on how abstract reasoning can arise from sensorimotor interaction and how a common spatial format supports different mental simulation processes.

6. Acknowledgment

I wish to thank my Ph.D. advisor Prof. Dr. Martin V. Butz for the support and advise during my promotion. I enjoyed the freedom I was given throughout my work, the room for my own interpretations, and the certainty that I could always rely on his professional advice. Furthermore, I highly appreciated the cooperative atmosphere and the interdisciplinary exchange he encouraged within the research group. I also wish to thank Prof. Dr. Bettina Rolke for providing the second assessment of this thesis and the pleasant collaboration with respect to research and teaching. The action-centred, anticipatory perspective on cognition that inspired the research questions pursued and presented here, is based on the research activities of Prof. Dr. Joachim Hoffmann and his former group at the University of Würzburg. I wish to thank him once again, not only for sparking my interest in cognitive psychology during my studies, but also for encouraging the interdisciplinary perspective that led to this work.

Over the past years, I had the opportunity and the pleasure to collaborate with many researchers. Besides my former and current colleagues at the chair of cognitive modeling, namely Dr. Anna Belardinelli, Dr. Andreas Alin, Stephan Ehrenfeld, PD Dr. Oliver Herbort, Dr. Jan Kneissler, Danil Koryakin, Sebastian Otte, Manuela Di Paolo, Fabian Schrodt, and Dr. Patrick Stalph, I also wish to thank the colleagues from the chair for evolutionary cognition and the group for neurophysiology & interventional neuropsychiatry.

During my work, I especially appreciated the opportunity to support theses and student projects. Some of these projects have been mentioned within this work and I wish to thank Jakob Gütschow, Robert Herzig, Johannes Kurz, Sara Moghimi, and Susanne Throner for their collaboration in the development and evaluation of the VR tools and setups. I also wish to thank Simone Kurek for handling the data collection in different studies.

Last but not least, I wish to thank my family and my friends for their support in stressful times, their understanding for the necessity of worked through weekends, and the reminder that there is life outside the laboratory.

7. Bibliography

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22*(4), 637–660. 69

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. 9, 59, 70

Batista, A. P., Buneo, C. A., Snyder, L. H., & Andersen, R. A. (1999). Reach plans in eye-centered coordinates. *Science*, *285*(5425), 257–260. 14

Belardinelli, A., Stepper, M. Y., & Butz, M. V. (2016). It's in the eyes: Planning precise manual actions before execution. *Journal of Vision*, *16*(1). doi: 10.1167/16.1.18 53

Blohm, G., & Crawford, J. D. (2009). Fields of gain in the brain. *Neuron*, *64*(5), 598–600. 71

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature*, *391*(6669), 756–756. 21, 43

Brotchie, P. R., Andersen, R. A., Snyder, L. H., & Goodman, S. J. (1995). Head position signals used by parietal neurons to encode locations of visual stimuli. *Nature*, *375*(6528), 232–235. 14

Brozzoli, C., Cardinali, L., Pavani, F., & Farnè, A. (2010). Action-specific remapping of peripersonal space. *Neuropsychologia*, *48*(3), 796–802. 52, 59

Brozzoli, C., Pavani, F., Urquizar, C., Cardinali, L., & Farnè, A. (2009). Grasping actions remap peripersonal space. *Neuroreport*, *20*(10), 913–917. 52

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *364*(1525), 1831–1840. 16

Butz, M. V. (2016). Toward a unified sub-symbolic computational theory of cognition. *Frontiers in Psychology*, *7*, 925. doi: 10.3389/fpsyg.2016.00925 51, 52, 74

Butz, M. V., Herbort, O., & Hoffmann, J. (2007). Exploiting redundancy for flexible behavior: Unsupervised learning in a modular sensorimotor control architecture. *Psychological Review*, *114*, 1015–1046. 43, 68

Butz, M. V., Kutter, E. F., & Lorenz, C. (2014). Rubber hand illusion affects joint angle perception. *PloS One*, *9*(3), e92854. 43, 68

Butz, M. V., Thomaschke, R., Linhardt, M. J., & Herbort, O. (2010). Remapping motion across modalities: Tactile rotations influence visual motion judgments. *Experimental Brain Research*, *207*, 1–11. 62

Canzoneri, E., Ubaldi, S., Rastelli, V., Finisguerra, A., Bassolino, M., & Serino, A. (2013). Tool-use reshapes the boundaries of body and peripersonal space representations. *Experimental Brain Research*, *228*(1), 25–42. 13, 43

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right-and left-handers. *Journal of Experimental Psychology: General*, *138*(3), 351–367. 65

Cohen, Y. E., & Andersen, R. A. (2000). Reaches to sounds encoded in an eye-centered reference frame. *Neuron*, *27*(3), 647–652. 14

Cohen, Y. E., & Andersen, R. A. (2002). A common reference frame for movement plans in the posterior parietal cortex. *Nature Reviews Neuroscience*, *3*(7), 553–562. 13, 14

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual information processing* (pp. 72–121). New York: Academic Press. 60

Davidson, E. J., & Wright, P. (2002). Selective processing of weight-and shape-related words in bulimia nervosa: Use of a computerised Stroop test. *Eating Behaviors*, *3*(3), 261–273. 31

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396. 16

Driver, J., & Spence, C. (1998). Cross–modal links in spatial attention. *Philosophical*

Transactions of the Royal Society B: Biological Sciences, *353*(1373), 1319–1331. 52

Ehrenfeld, S., & Butz, M. V. (2013). The modular modality frame model: Continuous body state estimation and plausibility-weighted information fusion. *Biological Cybernetics*, *107*(1), 61–82. 72

Ehrenfeld, S., Herbort, O., & Butz, M. V. (2013). Modular neuron-based body estimation: Maintaining consistency over different limbs, modalities, and frames of reference. *Frontiers in Computational Neuroscience*, *7*, 148. doi: 10.3389/fncom.2013.00148 68, 72

Ehrsson, H. H. (2007). The experimental induction of out-of-body experiences. *Science*, *317*(5841), 1048–1048. 21

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. 21, 34, 67

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169. 35

Fogassi, L., Gallese, V., Fadiga, L., Luppino, G., Matelli, M., & Rizzolatti, G. (1996). Coding of peripersonal space in inferior premotor cortex (area F4). *Journal of Neurophysiology*, *76*(1), 141–157. 13

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., & Slater, M. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine*, 1–8. 32

Freeman, J., Dale, R., & Farmer, T. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, *2*, 59. doi: 10.3389/fpsyg.2011.00059 31

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*(1456), 815–836. 72

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. 11, 12, 18

Gamberini, L., Seraglia, B., & Priftis, K. (2008). Processing of peripersonal and extrapersonal space using tools: Evidence from visual line bisection in real and virtual environments. *Neuropsychologia*, *46*(5), 1298–1304. 21

Gardony, A. L., Taylor, H. A., & Brunyé, T. T. (2014). What does physical rotation reveal about mental rotation? *Psychological Science*, *25*(2), 605–612. doi: 10.1177/0956797613503174 60, 61

Gattis, M., & Holyoak, K. J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 231–239. 71

Glenberg, A. M. (1997). What memory is for: Creating meaning in the service of action. *Behavioral and Brain Sciences*, *20*(1), 41–50. 7

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(4), 586–596. 9

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565. 9

Glenberg, A. M., & Robertson, D. A. (2000). Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning. *Journal of Memory and Language*, *43*(3), 379–401. 9

Glenberg, A. M., Witt, J. K., & Metcalfe, J. (2013). From the revolution to embodiment: 25 years of cognitive psychology. *Perspectives on Psychological Science*, *8*(5), 573–585. 9, 68

Goodale, M., Jakobson, L., & Keillor, J. (1994). Differences in the visual control of pantomimed and natural grasping movements. *Neuropsychologia*, *32*(10), 1159–1178. 32

Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, *42*(1-3), 335–346. 9

Herbort, O., & Butz, M. V. (2011). Habitual and goal-directed factors in (everyday) object handling. *Experimental Brain Research*, *213*(4), 371–382. 54

Hoffmann, J. (1993). *Vorhersage und Erkenntnis: Die Funktion von Antizipationen in der menschlichen Verhaltenssteuerung und Wahrnehmung*. Verlag für Psychologie, Hogrefe. 7, 58

Holmes, N. P., Calvert, G. A., & Spence, C. (2007). Tool use changes multisensory interactions in seconds: Evidence from the crossmodal congruency task. *Experimental Brain Research*, *183*(4), 465–476. 52

Holmes, N. P., & Spence, C. (2004). The body schema and multisensory representation(s) of peripersonal space. *Cognitive Processing*, *5*(2), 94–105. 13, 43, 49

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001a). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*, 849–937. 51, 58

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001b). Codes and their vicissitudes. *Behavioral and brain sciences*, *24*(5), 910–926. 51, 58

Ionta, S., Fourkas, A. D., Fiorio, M., & Aglioti, S. M. (2007). The influence of hands posture on mental rotation of hands and feet. *Experimental Brain Research*, *183*(1), 1–7. 15, 61

Iriki, A., Tanaka, M., & Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurones. *Neuroreport*, *7*(14), 2325–2330. 13

Janczyk, M., Pfister, R., Crognale, M. A., & Kunde, W. (2012). Effective rotations: Action effects determine the interplay of mental and manual rotations. *Journal of Experimental Psychology: General*, *141*, 489-501. doi: 10.1037/a0026997 65

Jeannerod, M. (1986). The formation of finger grip during prehension. a cortically mediated visuomotor pattern. *Behavioural Brain Research*, *19*(2), 99–116. 29, 30

Jeannerod, M. (1995). Mental imagery in the motor context. *Neuropsychologia*, *33*(11), 1419–1432. 69

Jeannerod, M., Arbib, M. A., Rizzolatti, G., & Sakata, H. (1995). Grasping objects: The cortical mechanisms of visuomotor transformation. *Trends in Neurosciences*, *18*(7), 314–320. 29

Kaschak, M. P., & Glenberg, A. M. (2000). Constructing meaning: The role of affordances and grammatical constructions in sentence comprehension. *Journal of Memory and Language*, *43*(3), 508–529. 9

Knauff, M., Mulack, T., Kassubek, J., Salih, H. R., & Greenlee, M. W. (2002). Spatial imagery in deductive reasoning: A functional MRI study. *Cognitive Brain Research*, *13*(2), 203–212. 70

Lake, B. M., Salakhutdinov, R., & Tenenbaum, J. B. (2015). Human-level concept learning through probabilistic program induction. *Science*, *350*(6266), 1332–1338. 72

Lakoff, G., & Johnson, M. (1980a). Conceptual metaphor in everyday language. *The Journal of Philosophy*, *77*(8), 453–486. 9

Lakoff, G., & Johnson, M. (1980b). The metaphorical structure of the human conceptual system. *Cognitive Science*, *4*(2), 195–208. 9

Liesefeld, H. R., & Zimmer, H. D. (2013). Think spatial: The representation in mental rotation is nonvisual. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(1), 167–182. 60

Linkenauger, S. A., Bülthoff, H. H., & Mohler, B. J. (2015). Virtual arm's reach influences perceived distances but only after experience reaching. *Neuropsychologia*, *70*, 393–401. 21

Loeber, S., Grosshans, M., Korucuoglu, O., Vollmert, C., Vollstädt-Klein, S., Schneider, S., . . . Kiefer, F. (2012). Impairment of inhibitory control in response to food-associated cues and attentional bias of obese participants and normal-weight controls. *International Journal of Obesity*, *36*(10), 1334–1339. 31

Longo, M. R., & Lourenco, S. F. (2006). On the nature of near space: Effects of tool use and the transition to far space. *Neuropsychologia*, *44*(6), 977–981. 43, 50, 68

Longo, M. R., & Lourenco, S. F. (2007). Space perception and body morphology: Extent of near space scales with arm length. *Experimental Brain Research*, *177*(2), 285–290. 50

Macaluso, E., & Maravita, A. (2010). The representation of space near the body through touch and vision. *Neuropsychologia*, *48*(3), 782–795. 13

Marmor, G. S., & Zaback, L. A. (1976). Mental rotation by the blind: Does mental rotation depend on visual imagery? *Journal of Experimental Psychology: Human Perception and Performance*, *2*(4), 515–521. 61

McGuire, L. M., & Sabes, P. N. (2009). Sensory transformations and the use of multiple reference frames for reach planning. *Nature Neuroscience*, *12*(8), 1056–1061. 42, 68

McMullen, P. A., & Jolicoeur, P. (1990). The spatial frame of reference in object naming and discrimination of left-right reflections. *Memory & Cognition*, *18*(1), 99–115. 60

Morey, C. C., & Cowan, N. (2005). When do visual and verbal memories conflict? the importance of working-memory load and retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 703–713. 36

Natale, M., Gur, R. E., & Gur, R. C. (1983). Hemispheric asymmetries in processing emotional expressions. *Neuropsychologia*, *21*(5), 555–565. 65

Neisser, U. (2014). *Cognitive Psychology: Classic Edition*. Psychology Press. 7

Nuerk, H.-C., Iversen, W., & Willmes, K. (2004). Notational modulation of the snarc and the marc (linguistic markedness of response codes) effect. *Quarterly Journal of Experimental Psychology Section A*, *57*(5), 835–863. 64

Parsons, L. M. (1987a). Imagined spatial transformation of one's body. *Journal of Experimental Psychology: General*, *116*(2), 172–191. 61

Parsons, L. M. (1987b). Imagined spatial transformations of one's hands and feet. *Cognitive Psychology*, *19*(2), 178–241. 61

Petkova, V. I., & Ehrsson, H. H. (2008). If i were you: Perceptual illusion of body swapping. *PloS one*, *3*(12), e3832. 21

Prinz, W. (1990). *A Common Coding Approach to Perception and Action*. Springer. 7, 58

Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in Human Neuroscience*, *9*, 197. doi: 10.3389/fnhum.2015.00197 36, 41

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review*, *120*(3), 561–588. 70, 72

Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science*, *31*(4), 613–643. 72

Riva, G. (2005). Virtual reality in psychotherapy: Review. *Cyberpsychology & Behavior*, *8*(3), 220–230. 32

Rosenbaum, D. A., Marchak, F., Barnes, H. J., Vaughan, J., Slotta, J. D., & Jorgensen, M. J. (1990). Constraints for action selection: Overhand versus underhand grips. In M. Jeannerod (Ed.), *Attention and Performance 13: Motor Representation and Control* (pp. 321–342). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc. 54

Schwartz, D. L., & Holton, D. L. (2000). Tool use and the effect of action on the imagination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1655–1665. 61

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703. doi: 10.1126/science.171.3972.701 60

Speer, N. K., Zacks, J. M., & Reynolds, J. R. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science*, *18*(5), 449–455. 12

Spence, C., Pavani, F., Maravita, A., & Holmes, N. (2004). Multisensory contributions to the 3-d representation of visuotactile peripersonal space in humans: Evidence from the crossmodal congruency task. *Journal of Physiology*, *98*(1), 171–189. 52, 69

Sridharan, D., Levitin, D. J., Chafe, C. H., Berger, J., & Menon, V. (2007). Neural dynamics of event segmentation in music: Converging evidence for dissociable ventral and dorsal networks. *Neuron*, *55*(3), 521–532. 12

Stricanne, B., Andersen, R. A., & Mazzoni, P. (1996). Eye-centered, head-centered, and intermediate coding of remembered sound locations in area LIP. *Journal of Neurophysiology*, *76*(3), 2071–2076. 14

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences*, *14*(9), 400–410. 35

Talsma, D., & Woldorff, M. G. (2005). Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, *17*(7), 1098–1114. 35

Thelen, A., Talsma, D., & Murray, M. M. (2015). Single-trial multisensory memories affect later auditory and visual object discrimination. *Cognition*, *138*, 148–160. 36

van Ee, R., van Boxtel, J. J., Parker, A. L., & Alais, D. (2009). Multisensory congruency as a mechanism for attentional control over perceptual selection. *Journal of Neuroscience*, *29*(37), 11641–11649. 35

Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, *7*(11), 483–488. 15, 16

Wang, D.-Y. D., Proctor, R. W., & Pick, D. F. (2003). The simon effect with wheel-rotation responses. *Journal of Motor Behavior*, *35*(3), 261–273. 65

Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition*, *68*(1), 77–94. 60, 61, 69

Winter, B., Marghetis, T., & Matlock, T. (2015). Of magnitudes and metaphors: Explaining cognitive interactions between space, time, and number. *Cortex*, *64*, 209–224. 72

Wohlschläger, A., & Wohlschläger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 397–412. 60, 61

Yokum, S., Ng, J., & Stice, E. (2011). Attentional bias to food images associated with elevated weight and future weight gain: An fMRI study. *Obesity*, *19*(9), 1775–1783. 31

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of Cognitive Neuroscience*, *20*(1), 1–19. 15, 61

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience*, *23*(12), 4057–4066. 72

Zacks, J. M., & Michelon, P. (2005). Transformations of visuospatial images. *Behavioral and Cognitive Neuroscience Reviews*, *4*(2), 96–118. 14, 15

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, *133*(2), 273–293. 12

Zacks, J. M., & Swallow, K. M. (2007). Event segmentation. *Current Directions in Psychological Science*, *16*(2), 80–84. 18, 68

Zacks, J. M., Swallow, K. M., Vettel, J. M., & McAvoy, M. P. (2006). Visual motion and the neural correlates of event perception. *Brain Research*, *1076*(1), 150–162.

12

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21. 12, 18, 68

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, *6*(5), 292–297. 12

Zwaan, R. A., & Taylor, L. J. (2006). Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General*, *135*(1), 1–11. 9

A. Behavioral bias for food reflected in hand movements: A preliminary study with healthy subjects

The manuscript has been published in *Cyberpsychology, Behavior, and Social Networking*. The final publication is available from Mary Ann Liebert, Inc., publishers https://doi.org/10.1089/cyber.2015.0311 and can be found by the following reference:

Schroeder, P. A., Lohmann, J., Butz, M. V., & Plewnia, C. (2015): **Behavioral bias for food reflected in hand movements: A preliminary study with healty subjects**. *Cyberpsychology, Behavior, and Social Networking*. 19(2), 120-126.

Behavioral bias for food reflected in hand movements: A preliminary study with healthy subjects

Philipp A Schroeder,[1,2] Johannes Lohmann,[3] Martin V Butz,[3] & Christian Plewnia[1,4]

1 Department of Psychiatry and Psychotherapy, Neurophysiology & Interventional Neuropsychiatry, University Tübingen, Germany

2 Department of Psychology, Diagnostics and Cognitive Neuropsychology, University Tübingen, Germany

3 Department of Computer Science, Cognitive Modeling, University Tübingen, Germany

4 Werner Reichardt Centre for Integrative Neuroscience (CIN), Tübingen, Germany

IN PRESS, CYBERPSYCHOLOGY, BEHAVIOR, AND SOCIAL NETWORKING

Abstract

Palatable food induces general approach tendencies when compared to non-food stimuli. For eating disorders, the modification of an attention bias towards food was proposed as a treatment option. Similar approaches have been efficient for other psychiatric conditions and recently successfully incorporated approach motivation. The direct impact of attentional biases on spontaneous natural behavior has hardly been investigated so far, although actions may serve as an intervention target, especially seeing the recent advances in the field of embodied cognition. With this study, we addressed the interplay of motor action execution and cognition when interacting with food objects. In a Virtual Reality setting, healthy participants repeatedly grasped or warded high-calorie food or hand-affordant ball objects using their own dominant hand. This novel experimental paradigm revealed an attention-like bias in hand-based actions: 3D objects of food were collected faster than ball objects and this difference correlated positively with both individual body-mass index and diet-related attitudes. The behavioral bias for food in hand movements complements several recent experimental and neurophysiological findings. Implications for the use of Virtual Reality in the treatment of eating-related health problems are discussed.

Keywords: Virtual Reality, Attention bias, Hand movement, Embodied cognition

Introduction

Automaticity and dysfunctional cognitive control are regarded as fundamental mechanisms for the generation, maintenance, and treatment of hazardous and risky behavior.[1] However, despite of the widespread acceptance of a tight link[2,3], the correspondence between cognitive processes and overt behavior is not fully understood. Cognitive processes and overt behavior are often intertwined in interactive and reciprocally influential ways, both being significantly affected by involuntary biases based on memory and attentional processes.[4–6] An illustrative example for the complex and often challenging interaction between thinking and acting is the individual behavior towards food and eating. Here, derailed cognitive and/or behavioral control of food selection and intake can frequently result in impaired health conditions, with highly adverse individual and economic consequences.[7]

Current research on biased automaticity predominantly highlights attentional processes. For instance, palatable food as compared to non-food pictures elicit faster key presses[8] and provoke prolonged gaze responses.[8–10] Obviously, a strong attentional bias towards food, possibly undermined by deficient inhibitory control, contributes to actual behavior, i.e., increased snack food intake.[11] However, it is largely unclear if and how precisely these biases, predominantly assessed in laboratory environments, are reflected in everyday behavior.

In clinical settings, attention bias modification has been proposed as a promising therapeutic tool for eating disorders.[12] By repeatedly directing attention towards or away from illness-specific stimuli, relevant automatic cognitive processes are modifiable. Similar approaches were extensively employed in the treatment of anxiety and mood disorders, produced small-to-medium effect sizes in meta-analyses[13,14,see also: 15] and can improve depression symptomatology.[16] Interestingly, modification paradigms recently successfully incorporated approach and avoidance motivation by use of implicit associations tasks[17,18] or directed peripherals, i.e., in form of the zooming joystick approach-avoidance task.[11,19,20] However, at least in the eating domain, approach-avoidance training triggered complex responses and has not been effective necessarily.[21,22]

Besides an emphasis on approach-avoidance motivation,[23] an attention bias for food should be related to actual behavioral tendencies, ultimately supporting excessive or deficient energy intake. Further, considering increasing interest in grounded and embodied cognition,[2,24] attention bias modification paradigms could profit from incorporating active sensorimotor interactions to a more explicit degree. In this case, it can be expected that actions will be influenced by intrinsic states of the subjects (such as current volitional and long-term goals), thus integrating properties of the self, current intentions, and bodily activity.[25] Vice versa, seeing that cultural experience can specifically alter embodied cognitive processes,[26] behavioral training of particular object interactions might be an ecologically valid tool in modifying dysfunctional cognition and in providing personal experiences to reshape object-related tendencies.

Recent studies outlined that biased cognitive processing was also reflected in motor activity and even action decisions.[5,27,28] For instance, postural asymmetries were triggered by food preferences, documenting approach and avoidance by postural sway towards (away from) highly preferred (non-preferred) palatable items.[6] In food-deprived individuals, food pictures automatically elicited stronger activity in the musculus zygomaticus, possibly attenuating subsequent anticipated food intake.[29]

With this study, we aimed at investigating the behavioral bias for food at the level of natural hand movements. More specifically, by introducing a novel Virtual Reality setup, we allow for a shift of focus from cognitive to behavioral aspects of biased processing. In an experimental scenario, participants were asked to interact with virtual high-calorie food objects and with non-food control objects using their own dominant hand. Interactions with ball objects were employed as a non-food reference category with clear hand interactions as well as memory representations.[30,31] We investigated an attentional bias for food, which can be considered a rather common cognitive phenomenon,[32,33] in the course of hand grasp and ward movements, assessing the timing of movement initiation, object contact, and object collection. In line with behavioral and neuroimaging results,[34] we hypothesized that a behavioral bias for food would increase in correlation with participant's body-mass index (BMI) and that it should also be related to eating attitudes.

Materials and methods

Participants

Twenty-three healthy and right-handed student participants (8 male; mean age: 23.9 years, range: 19-33; LI > 65[35]) were invited to participate in a Virtual Reality (VR) experiment. All participants appeared satiated and were served unflavored popcorn at the beginning of the experiment to counteract possible adverse effects of VR (i.e., simulation sickness) and to ensure that they were equally fed (i.e., because automatic attention allocation is enhanced in conditions of hunger).[8] No history of severe mental disease was reported by any participant. The experimental procedure was approved by the ethical commission of the University Hospital Tübingen and informed consent was collected prior to the beginning of the experiment. The overall procedure lasted 80-90 minutes and a monetary compensation of 10€ was offered.

Virtual Reality

Participants sat comfortably in a dimly lit room and were equipped with an Oculus Rift © DK2 stereoscopic head-mounted display (Oculus VR LLC, Menlo Park, California). Hand movements were tracked using the Leap Motion © near-infrared sensor (Leap Motion Inc, San Francisco, California), placed in front of participants on a small cupboard of 45cm height (Figure 1A).

The task was implemented using the Virtual Reality equipment low-level API and custom code in Unity3D 4.5/C#. Food and ball interaction objects were obtained from the Unity3D asset store and modified to an identical size. The VR scenario placed participants inside of a baseball arena and the interaction hand was displayed by a simplified stylized virtual hand (see Figure 1) to avoid body-ownership mismatches.[i.e., 36]

In a practice block, participants were introduced to grasping and warding interactions using neutral, non-textured sphere objects. Trials were initiated by placing the dominant hand at a predefined starting position in space and transparent color-switching cues indicated the correct position. After the HMD was oriented centrally towards the fixation cross without head movements for 1000 ms, a visual color cue appeared for 400 ms, indicating both the required interaction and the onset location of the target object. All grasping interactions afforded that objects were collected and placed within a close virtual box (Figure 1B).

After participants were accustomed to the task, all food and ball objects were rated for likeability using a virtual visual analog scale. In the following experimental scene, participants repeatedly grasped or warded food or ball objects that appeared at a constant distance of 20 cm in front of them after trial initiation. The front-parallel positions of the objects were randomized with maximal 10/5 cm deviation permitted in the x-/y-axis. In each of 10 blocks, 24 randomly selected objects were presented and feedback was given using a progress bar in the virtual sky and the increasingly filled box that continuously displayed the last six collected objects. Each of the different objects (balls: volleyball, basketball, baseball, tennis ball; food: slice of pizza, hamburger, piece of chocolate pie, donut) appeared equally often with a warding or grasping instruction.

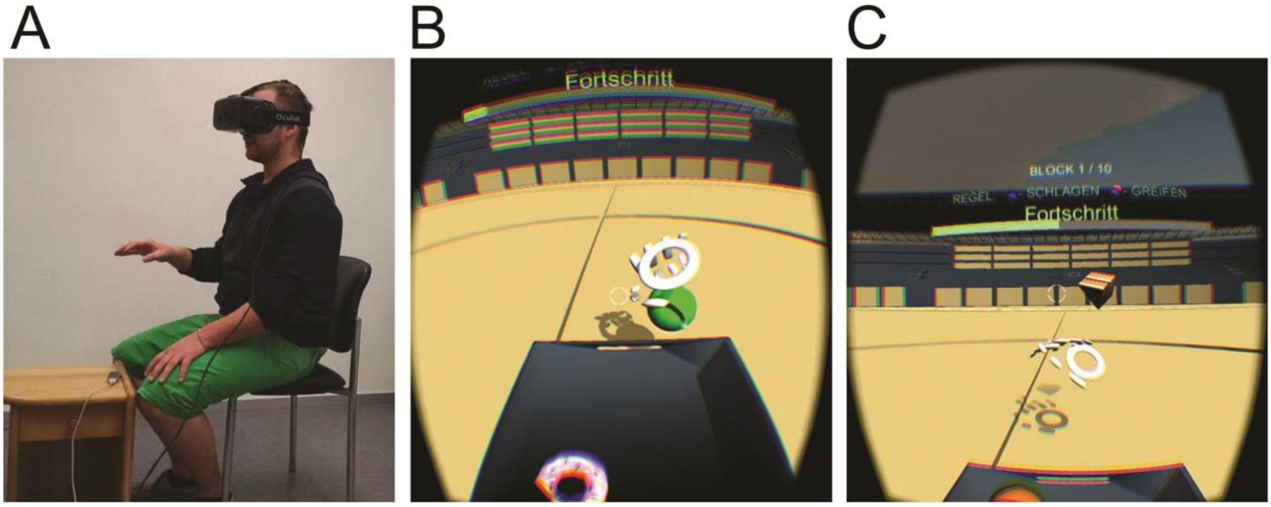


Figure 1. ***Virtual Reality setting.*** *Panel A depicts a participant equipped with the Oculus Rift © DK2 HMD and interacting with the Leap Motion © Recorder. Panel B depicts the grasping and collection interaction, Panel C depicts the warding interaction and both B+C display a participant's left eye view only.*

Questionnaires

Mood was assessed before and after the experiment, using the positive and negative affect scale.[37,German translation by: 38] After the experiment, we assessed presence using the IPQ[39,40] and simulation sickness using the SSQS.[41]

Eating attitudes and body-mass index (BMI) were assessed using the EAT-26 questionnaire[42] in its German translation.[43] The questionnaire assesses eating-related thoughts and attitudes and discriminates (1) Dieting, (2) Bulimia and Food Preoccupation, and (3) Oral Control.

Data analysis and data treatment

Response times were collected at different trial stages (hand movement initiation, object contact, and object collection) using C#/.NET code embedded into the VR application. Kolmogorov-Smirnov tests for normality of median response times (RTs) signaled normal distribution of the data (*z*s < .98, *p*s > .29) and RTs for movement onset and object contact were subjected to separate 2 x 2 repeated measures ANOVAs comprising the factors $\text{OBJECT}_{\text{food,ball}}$ and $\text{INTERACTION}_{\text{grasp,ward}}$. Median collection times, available for grasp trials only, were subjected to a paired t-test. The significance level was set at $\alpha = 0.05$. Questionnaires were analyzed accordingly and critical measures (BMI, eating attitude scores) were correlated with collection time differences for food vs. ball objects (Pearson coefficient for continuous variables).

Results

Virtual Reality

One male participant was not comfortable with the hand tracking procedure and repeatedly failed in positioning the starting hand. Two female participant's data were lost due to technical problems. Data are reported for the remaining 20 participants.

Presence ratings as obtained from the IPQ were high (male participants: *M*=49.4, *SE*=6.2; female participants: *M*=46.0, *SE*=3.4), indicating that participants were sufficiently immersed in the VR scenario. There were no significant gender differences in presence ratings, *t*(18)=0.530, *p*=.60. Despite the long exposure to the VR system, simulation sickness was unincisive (mean score on the SSQS: 5.75, *SE*=0.88), with largest strain related to Eye Pressure (mean rating on a 1-to-4 scale: 2.0, *SE*= 0.86), followed by Fatigue (*M*=1.95, *SE*=0.89), Fullness of Head (*M*=1.80, *SE*=1.11) and Difficulty Focusing (*M*=1.65, *SE*=0.86). There were no significant effects of the VR exposure on participant's mood, *p*s>.18.

Movement Onset

Hand movements were initiated by leaving the predefined starting position after an interaction cue was replaced by a target object. Median response times after object onset were submitted to the 2 x 2 repeated measures ANOVA comprising the factors OBJECT$_{\text{ball,food}}$ and INTERACTION$_{\text{grasp,ward}}$. There was a significant OBJECT$_{\text{ball,food}}$ main effect, *F*(1,19)=4.32, *p*=.05, ηp^2=0.19, reflecting earlier movement initiations towards Food objects (*M*=1361 ms, *SE*=112 ms) as compared to Ball objects (*M*=1417 ms, *SE*=121 ms). Neither the main effect INTERACTION$_{\text{grasp,ward}}$, *F*(1,19)=2.80, *p*=.11, nor the two-way interaction was significant, *F*(1,19)=0.13, *p*=.72.

Object Contact

Ward movements were executed faster than grasp movements (Mean difference: 218.72 ms, *SE*=60.88 ms) as reflected by a significant main effect of INTERACTION$_{\text{grasp,ward}}$, *F*(1,19)=12.90, *p*=.002, ηp^2=0.41. Again, a timely advantage emerged for Food objects, *F*(1,19)=4.50, *p*=.05, ηp^2=0.19, but there was no significant two-way interaction effect, *F*(1,19)=1.19, *p*=.29.

Collection Times

Finally, in grasp trials, we tested whether Food objects were generally faster collected than Ball objects (or vice versa), as reflected by the time from movement onset until an object was placed inside the box underneath the participant. In line with the preceding analyses, we indeed found that Food objects were faster gathered than Ball objects, and the difference was significant, *t*(19)=3.692, *p*=.002, d=0.16 (Figure 2). Compared with the initial bias observed in movement onset times, the behavioral bias for food apparently increased by 39.5% (Δ=44.33ms, *SE*=24.81ms), and we noted a statistical trend for significance, *t*(19)=1.79, *p* =.09, d=0.28. This increase was specific for return dynamics and not found in contact times already (Δ=5.73ms, *SE*=44.23ms), *t*(19)=.01, *p*=.90, and a descriptive reduction of the initial bias in ward movements was not significant (Δ=-18.83ms, *SE*=22.97ms), *t*(19)=.82, *p*=.29.

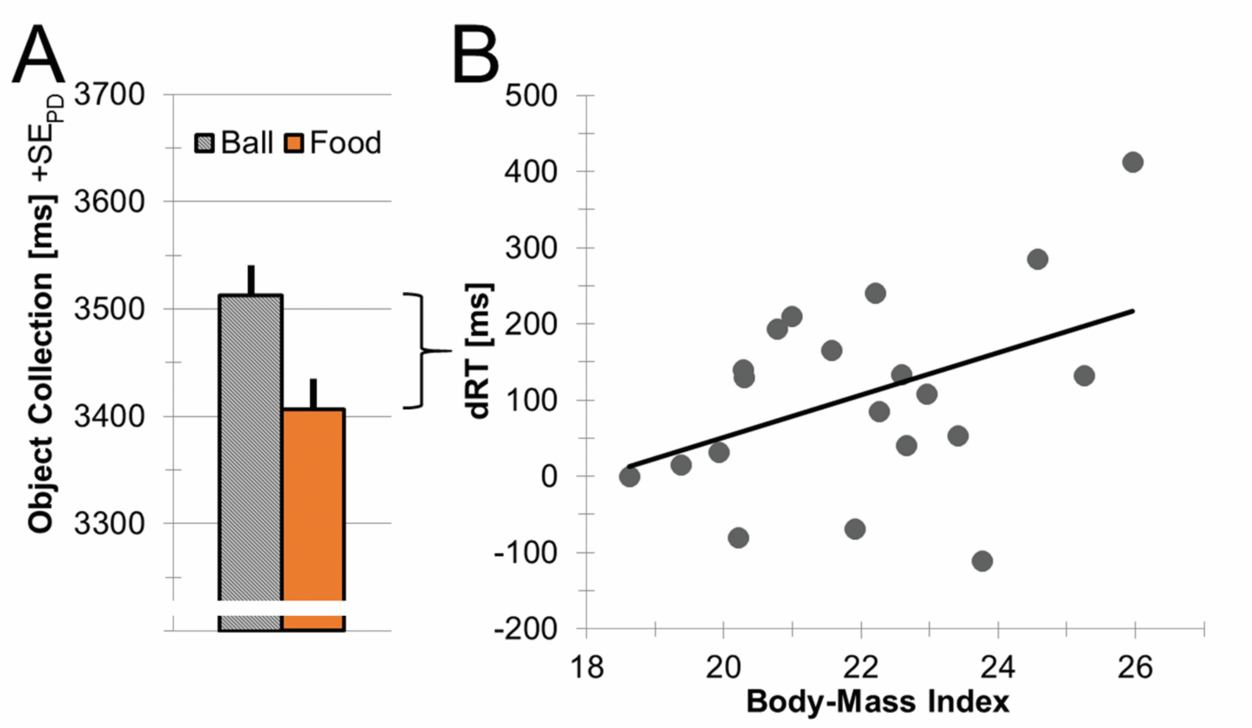


Figure 2. ***Behavioral bias for food in collection times (A) and correlation with BMI (B).*** *Participants collected food objects (orange bar) significantly faster than Ball objects (grey bar). The individual difference (dRT=Ball-Food [Collection RT]) correlated positively with the Body-Mass Index, r=.428, p=.030. Error bars reflect standard errors of paired difference.*[42]

BMI and Eating Attitudes

Although we recruited a healthy student population, there was considerable variance in both BMI (mean: 22, range: 19-26) and eating attitude scores from the EAT-26 questionnaire (mean: 4.4, range: 0-24). To estimate whether the behavioral bias for food as assessed in this VR experiment was related to the individual bodily consequences of, and attitudes towards eating, we correlated individual BMI and the EAT-26 questionnaire (with subscores for Dieting, Bulimia and Food Preoccupation, and Oral Control) with the collection-time difference between Food and Ball objects (dRT=Ball-Food; Figure 3B). Following the literature, larger collection time advantages (i.e., favoring food-related actions) should emerge with greater BMI,[i.e., 34] and one-tailed Pearson's correlation coefficients were tested in this direction. With increasing BMI, the collection time advantage in collecting food was indeed larger, r=.428, p=.030. For eating attitudes, again, a positive correlation emerged, r=.388, p=.046, which was driven by the Dieting subscale, r=.503, p=.012. There were no significant correlations with the Bulimia and Food Preoccupation and Oral Control subscales. While together these results are suggestive of a food-specific effect, they do not consider possible differential low-level effects of stimulus features (i.e., object complexity or color) or involvement of further moderator variables.

Control analysis: Object valence or edibility?

A further analysis was conducted to differentiate whether edibility or likeability of the employed food objects accounted for the observed effects. Recent studies indicated that valence as abstract concept might be associated with space in an embodied way.[45] Likewise, food preferences modulated approach posture[6] and it was possible that any behavioral bias in our experiment was driven by the individual valence and not the possible edibility of any visual object.

As expected, there was a trend for edible objects to be rated more positively than ball objects, though the biserial correlation of object type and valence ratings was rather small, r=.15, p=.059. Nevertheless, by ranking individual object ratings, we determined both relative pleasant and unpleasant Food and Ball objects for each participant and included this further factor of $\text{V}\small{\text{ALENCE}}_{\text{pleasant,unpleasant}}$ in our analyses. Although we could not rule out low-level feature differences, this approach would allow us to understand whether any difference was attributable to edibility itself, to an object's valence or to an interaction of the two factors (i.e., a strong preference for specific food might account for biased grasping).

However, we neither found a main effect of Valence nor any interaction with this additional factor in all three ANOVAs, all *p*s > .26. The previously reported results were essentially replicated, suggesting that our findings were indeed food-specific.

Discussion

With this study, we tested a new Virtual Reality (VR) setup for measuring hand-tracked movements in simulated interaction with 3D high-calorie food and non-food ball objects. A simple scenario was set up using commercially available hard- and software modified for experimental control. Participants were comfortable with the VR, reported high presence and little-to-none adverse effects, and successfully solved the task of repeatedly warding and grasping different 3D objects.

Complementing previous findings, food objects were significantly faster grasped and collected than non-food objects. This result replicates an attentional bias for food in movement onset times and additionally demonstrates that biased attentional processes can effectively channel the execution of actions. Thereby, employing current technological advances in the reign of VR, we revealed a motor bias for food that was expressed through actual hand movements. Given the early appearance of this bias in movement initiations, well-known attentional processes are likely involved, possibly corroborating subsequent motor execution differences. The significance of our results is that the attentional bias for food was reflected in motor actions. Furthermore, the bias was stable and apparently even increased for the return dynamics of grasp movements. In this study, an overall consequence of biased attention and motor responses towards food resulted in a faster collection of food items compared to hand-affordant ball objects.

Attentional and behavioral bias for food

Of course, for now, the external validity of our findings might be restricted to the exact objects examined. However, given that the observed behavioral bias related to individual body-mass indices, similar to results from highly-controlled settings,[34] it is very likely that edibility critically accounts for the bias. Further, hand movements were neither valence-specific nor modulated by individual preferences, which supports the food-specificity of the bias. Our results also resemble the existing literature on attentional biases towards food: When afforded to respond to food words as compared to object words in a go/nogo task, faster key presses were documented in both obese patients and healthy controls.[46] In contrast to neutral cues, both appetizing and unappetizing food cues elicited faster responses with increasing BMI in a dot-probe paradigm.[34] In general, attentional biases towards food had been investigated exhaustively using the Food Stroop task: Here, participants color-judge food-related words and control words by pressing a key as fast as possible.[47] In the task, food words interfere with color naming particularly in clinical populations, constituting selective attentional processes elicited by psychopathology-relevant stimuli. For healthy controls, small effect sizes were found in a recent meta-analysis (d=.21)[48] and the bias was moderately increased for eating-disordered females (d=.49), i.e., bulimic patients.[48,49] Thus, it can be concluded from the presented literature that attentional biases for food stimuli are a robust phenomenon. In our paradigm, we first succeeded at replicating the bias in a setup employing natural hand responses.

Our results however go beyond this replication in an ecologically more valid VR scenario, and this is because the bias was stable for natural collection movements. Whereas the Food Stroop task and related paradigms are mainly cognitive visual tests, our VR setup incorporated visuo-proprioceptive hand tracking and assessed actual interactions with virtual food objects.

Clinical implications

Body perception, i.e., body image distortion, is a core symptom of anorexia nervosa, but the concept includes nonvisual and multisensory impairments involving tactile and proprioceptive sensory components.[50] Following embodied cognition theories, an allocentric lock (i.e., physical appearance to others is bolstered at the cost of real-time bodily experience) into the negatively perceived and objectified body can produce dysfunctional eating behavior with body shame priming food/body related experience.[51] Thus, full body tracking approaches potentially allow both researchers and practitioners to address the bodily interrelations in parallel to the cognitive dysfunctions of eating disorders. In this vein, VR has been repeatedly used as a tool to assess body-image disturbances in eating disorders.[52,53] Given that body-ownership was readily achieved even for over- und underweight virtual bodies,[54] possible multidimensional interventions could be achieved using perceptual, but also action simulation. Full-body tracking and multimodal simulation might be powerful tools to promote a functional bodily self-awareness. Likewise, it might be possible to reinstate an altered reward-system[55] and, at the same time, channel healthy action execution in case of excess food intake.

Additional implications for clinical applications might be considered: For instance, in Anorexia Nervosa, an initial attentional orientation towards food was observed, but gaze duration for food pictures decreased as compared to healthy controls, suggesting attentional avoidance in later phases of information processing.[56] Similar to recent approach-avoidance attention bias modification paradigms,[21] grasping healthy food as a continuous action can be promoted by the use of VR training scenarios.

On the other hand, Binge-Eating Disorder was characterized by food-related impulsivity[57] and particularly supported by an early locus in stimulus processing with increasing symptom severity.[58] Here, VR might offer an ecologically valid approach to practice inhibitory control, possibly augmenting current efforts for pervasive gesture-based training.[59,60] In future settings, it will be interesting to also examine movement speed and trajectories more exhaustively which might allow further insights into the precise mechanisms of biased action and inhibitory control.[61]

Limitations

Several limitations of this study have to be considered: First and foremost, we studied a small and healthy population of student volunteers in an experimental setting, limiting the generalizability and possible clinical applicability of the data. Second, it needs further clarification whether low-level stimulus features such as object complexity or color are involved in the behavioral bias. Although our control analysis ranked objects individually and thereby effectively permutated low-level features of the objects used in this study, the results obtained here and in the correlation analysis are suggestive, but not conclusive of a food-specific effect. Third, at this point, the precise interplay of cognitive and behavioral biases in the development, maintenance, and relapse of psychiatric disorders is unclear. Although computational frameworks are available for short-term action effects,[62] subtle long-term effects of overeating might deter cognitive processing. Finally, the developmental process of dysfunctional habits or actions that can evolve into a distinct behavioral problem (or vice versa) needs further investigations.

Gender differences in Virtual Reality

Different previous studies suggested gender differences in the use of virtual reality and presence and it was documented that men achieved higher feelings of presence in virtual environments.[63,64] In our sample, such gender differences were not detected. Obviously, our setup differed from previous VR studies by including a natural user interface only: Participants did not have to adapt to artificial interactions, i.e., by using a joystick or mouse. In contrast, participants' own hands were projected into the virtual environment and used to solve the task. Given our unbalanced and small sample, we can only speculate about the underlying mechanisms for a lack of gender differences in presence ratings. For instance, the novel input mechanism (natural hand tracking) might have been equally difficult to master for both genders. Possibly, use of the Leap Motion input device in this study alleviated possible advances from previous encounters with video games in male populations.[64] Alternatively, while most VR scenarios afford some kind of spatial navigation and gender differences were particularly large in spatial processing,[65] the lack of this particular element might have mitigated possible effects of gender. However, to address any of these possibilities, it will be necessary to further investigate gender differences in the use of virtual reality with natural and artificial user interfaces.

Conclusion

Our experimental data provide preliminary evidence from healthy subjects for a food-related manual bias as reflected in the dynamics of hand movements. Faster motor activity towards edible objects and the correlation of the bias with BMI and diet attitudes point towards a significant functional relevance for eating behavior and potentially eating-related pathology. Simple VR setups are feasible for the assessment of hand movements in relation to a controlled presentation of ecologically valid stimuli and allow for the investigation of behavioral biases in extension to well-known attentional biases for food. Therefore, VR may usefully complement other behavioral, eye-tracking[9] and neurophysiological approaches in studying eating-related pathology.[66–68] Moreover, particularly in combination with training techniques, VR intervention programs might constitute a promising tool for the modification of pathophysiologically relevant behavioral biases in various psychiatric disorders.

References

1 Friese M, Hofmann W, Wiers RW. On taming horses and strengthening riders: Recent developments in research on interventions to improve self-control in health behaviors. *Self and Identity* 2011; **10**:336–51.

2 Barsalou LW. Grounded cognition. *Annual review of psychology* 2008; **59**:617–45.

3 Fuchs T, Schlimme JE. Embodiment and psychopathology: a phenomenological perspective. *Current Opinion in Psychiatry* 2009; **22**:570–5.

4 Werthmann J, Field M, Roefs A, Nederkoorn C, Jansen A. Attention bias for chocolate increases chocolate consumption--an attention bias modification study. *Journal of behavior therapy and experimental psychiatry* 2014; **45**:136–43.

5 Schroeder PA, Pfister R. Arbitrary numbers counter fair decisions: trails of markedness in card distribution. *Frontiers in Psychology* 2015; **6**.

6 Brunyé TT, Hayes JF, Mahoney CR, Gardony AL, Taylor HA, Kanarek RB. Get in my belly: food preferences trigger approach and avoidant postural asymmetries. *PloS one* 2013; **8**:e72432.

7 World Health Organization. (2000) Obesity: preventing and managing the global epidemic. No. 894. Geneva: W.H.O.

8 Nijs IMT, Muris P, Euser AS, Franken IHA. Differences in attention to food and food intake between overweight/obese and normal-weight females under conditions of hunger and satiety. *Appetite* 2010; **54**:243–54.

9 Werthmann J, Roefs A, Nederkoorn C, Mogg K, Bradley BP, Jansen A. Attention bias for food is independent of restraint in healthy weight individuals-an eye tracking study. *Eating behaviors* 2013; **14**:397–400.

10 Castellanos EH, Charboneau E, Dietrich MS, *et al.* Obese adults have visual attention bias for food cue images: evidence for altered reward system function. *International Journal of Obesity* 2009; **33**:1063–73.

11 Kakoschke N, Kemps E, Tiggemann M. Combined effects of cognitive bias for food cues and poor inhibitory control on unhealthy food intake. *Appetite* 2015; **87**:358–64.

12 Renwick B, Campbell IC, Schmidt U. Review of attentional bias modification: A brain-directed treatment for eating disorders. *European Eating Disorders Review* 2013; **21**:464–74.

13 Browning M, Holmes EA, Harmer CJ. The modification of attentional bias to emotional information: A review of the techniques, mechanisms, and relevance to emotional disorders. *Cognitive, Affective, & Behavioral Neuroscience* 2010; **10**:8–20.

14 Mogoaşe C, David D, Koster EHW. Clinical efficacy of attentional bias modification procedures: An updated meta-analysis. *Journal of Clinical Psychology* 2014; **70**:1133–57.

15 Plewnia C, Schroeder PA, Wolkenstein L. Targeting the biased brain: non-invasive brain stimulation to ameliorate cognitive control. *The Lancet Psychiatry* 2015; **2**:351–6.

16 Baert S, De Raedt R, Schacht R, Koster EHW. Attentional bias training in depression: therapeutic effects depend on depression severity. *Journal of behavior therapy and experimental psychiatry* 2010; **41**:265–74.

17 Kemps E, Tiggemann M, Martin R, Elliott M. Implicit approach-avoidance associations for craved food cues. *Journal of Experimental Psychology: Applied* 2013; **19**:30–8.

18 Haynes A, Kemps E, Moffitt R. Inhibitory self-control moderates the effect of changed implicit food evaluations on snack food consumption. *Appetite* 2015; **90**:114–22.

19 Wiers RW, Eberl C, Rinck M, Becker ES, Lindenmeyer J. Retraining automatic action tendencies changes alcoholic patients' approach bias for alcohol and improves treatment outcome. *Psychological Science* 2011; **22**:490–7.

20 Rinck M, Becker ES. Approach and avoidance in fear of spiders. *Journal of Behavior Therapy and Experimental Psychiatry* 2007; **38**:105–20.

21 Becker D, Jostmann NB, Wiers RW, Holland RW. Approach avoidance training in the eating domain: testing the effectiveness across three single session studies. *Appetite* 2014; **85**:58–65.

22 Neimeijer RAM, de Jong PJ, Roefs A. Automatic approach/avoidance tendencies towards food and the course of anorexia nervosa. *Appetite* 2015; **91**:28–34.

23 Elliot A. Approach and avoidance motivation and achievement goals. *Educational psychologist* 1999; **34**:169–89.

24 Butz MV. How and Why the Brain Lays the Foundations for a Conscious Self. *Constructivist Foundations* 2008; **4**:1–42.

25 Riva G, Waterworth JA, Waterworth EL, Mantovani F. From intention to action: The role of presence. *New Ideas in Psychology* 2011; **29**:24–37.

26 De la Fuente J, Casasanto D, Román A, Santiago J. Can Culture Influence Body-Specific Associations Between Space and Valence? *Cognitive science* 2015; **39**:821–3.

27 Fuchs T. Embodied cognitive neuroscience and its consequences for psychiatry. *Poiesis und Praxis* 2009; **6**:219–33.

28 Brunyé TT, Gardony A, Mahoney CR, Taylor HA. Body-specific representations of spatial location. *Cognition* 2012; **123**:229–39.

29 Hoefling A, Likowski KU, Deutsch R, *et al.* When hunger finds no fault with moldy corn: Food deprivation reduces food-related disgust. *Emotion* 2009; **9**:50–8.

30 Bub DN, Masson MEJ. Grasping beer mugs: on the dynamics of alignment effects induced by handled objects. *Journal of Experimental Psychology: Human Perception and Performance* 2010; **36**:341–58.

31 Taylor LJ, Zwaan RA. Grasping spheres, not planets. *Cognition* 2010; **115**:39–45.

32 Dobson, K.S., Dozois, D.J.A. (2004). Attentional biases in eating disorders: A meta-analytic review of Stroop performance. Clinical Psychology Review, 23, 1001-1022.

33 Johansson, L., Ghaderi, A., Andersson, G. (2005). Stroop interference for food- and body-related words: a meta-analysis. Eating Behaviors, 6, 271-281.

34 Yokum S, Ng J, Stice E. Attentional bias to food images associated with elevated weight and future weight gain: an FMRI study. *Obesity* 2011; **19**:1775–83.

35 Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 1971; **9**:97–113.

36 Lugrin J, Latt J, Latoschik ME. (2015) Avatar Anthropomorphism and Illusion of Body Ownership in VR. In: Proceedings of the IEEE VR. Arles, France: in press.

37 Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of personality and social psychology* 1988; **54**:1063–70.

38 Krohne HW, Egloff B, Kohlmann C-W, Tausch A. Untersuchungen mit einer deutschen Version der ‘Positive and Negative Affect Schedule’ (PANAS). / Investigations with a German version of the Positive and Negative Affect Schedule (PANAS). *Diagnostica* 1996; **42**:139–56.

39 Regenbrecht H, Schubert T. Real and Illusory Interactions Enhance Presence in Virtual Environments. *Presence: Teleoperators and Virtual Environments* 2002; **11**:425–34.

40 Schubert T, Friedmann F, Regenbrecht H. The experience of presence: Factor analytic insights. *Presence* 2001; **10**:266–81.

41 Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 1993; **3**:203–20.

42 Garner DM, Olmsted MP, Bohr Y, Garfinkel PE. The Eating Attitudes Test: Psychometric features and clinical correlates. *Psychological medicine* 1982; **12**:871–8.

43 Meermann R, Vandereycken W. (1987) Therapie der Magersucht und Bulimia nervosa. Berlin: De Gryter.

44 Pfister R, Janczyk M. Confidence intervals for two sample means: Calculation, interpretation, and a few simple rules. *Advances in Cognitive Psychology* 2013; **9**:74–80.

45 Casasanto D. Embodiment of abstract concepts: good and bad in right- and left-handers. *Journal of Experimental Psychology: General* 2009; **138**:351–67.

46 Loeber S, Grosshans M, Korucuoglu O, *et al*. Impairment of inhibitory control in response to food-associated cues and attentional bias of obese participants and normal-weight controls. *International Journal of Obesity* 2012; **36(10)**:1334-1339.

47 Davidson EJ, Wright P. Selective processing of weight- and shape-related words in bulimia nervosa: Use of a computerised Stroop test. *Eating Behaviors* 2002, **3**:261-273.

48 Johansson L, Ghaderi A, Andersson G. Stroop interference for food- and body-related words: a meta-analysis. *Eating Behaviors* 2005, **6**:271-281.

49 Dobson KS, Dozois DJA. Attentional biases in eating disorders: A meta-analytic review of Stroop performance. *Clinical Psychology Review* 2004, **23**:1001-1022.

50 Gaudio S, Brooks SJ, Riva G. Nonvisual Multisensory Impairment of Body Perception in Anorexia Nervosa: A Systematic Review of Neuropsychological Studies. *PLoS ONE* 2014; **9**:e110087.

51 Riva G. Out of my real body: cognitive neuroscience meets eating disorders. *Frontiers in human neuroscience* 2014; 8:236.

52 Riva G, Bacchetta M, Baruffi M, Molinari E. Virtual Reality-based multidimensional therapy for the treatment of body image disturbances in binge eating disorder: A

preliminary controlled study. *IEEE Transactions on Information Technology in Biomedicine* 2002; **6**:1–11.

53 Riva G, Bacchetta M, Baruffi M, Molinari E. Virtual reality-based multidimensional therapy for the treatment of body image disturbances in obesity: a controlled study. *Cyberpsychology & Behavior* 2001; **4**:511–26.

54 Piryankova I V, Wong HY, Linkenauger SA, *et al.* Owning an overweight or underweight body: distinguishing the physical, experienced and virtual body. *PloS one* 2014; **9**:e103428.

55 Volkow ND, Wang GJ, Fowler JS, Telang F. Overlapping neuronal circuits in addiction and obesity: evidence of systems pathology. *Philosophical Transactions of the Royal Society of London B: Biological sciences* 2008; **363**:3191–200.

56 Giel KE, Friederich HC, Teufel M, Hautzinger M, Enck P, Zipfel S. Attentional processing of food pictures in individuals with anorexia nervosa - An eye-tracking study. *Biological Psychiatry* 2011; **69**:661–7.

57 Schag K, Schönleber J, Teufel M, Zipfel S, Giel KE. Food-related impulsivity in obesity and binge eating disorder--a systematic review. *Obesity reviews* 2013; **14**:477–95.

58 Schmitz F, Naumann E, Trentowska M, Svaldi J. Attentional bias for food cues in binge eating disorder. *Appetite* 2014; **80**:70–80.

59 Scott-Brown KC, van der Pol M, Moncrieffe C, *et al.* (2012) Service-Please: an interactive healthy eating serious game application for tablet computer. In:

Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers. Swinton: British Computer Society: 381–5.

60 Orji R, Mandryk RL, Vassileva J, Gerling KM. (2013) Tailoring persuasive health games to gamer type. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI ’13. New York: ACM Press: 2467.

61 Dignath D, Pfister R, Eder AB, Kiesel A, Kunde W. Something in the way she moves—movement trajectories reveal dynamics of self-control. Psychonomic Bulletin & Review 2014; 21:809–16.

62 Herbort O, Butz M V. Too good to be true? Ideomotor theory from a computational perspective. *Frontiers in Psychology* 2012; **3**:1–17.

63 Felnhofer A, Kothgassner OD, Beutl L, Hlavacs H, Kryspin-exner I. (2012) Is Virtual Reality made for Men only ? Exploring Gender Differences in the Sense of Presence of Psychology , Working Group Clinical and Health Psychology , University of Vienna. In: Proceedings of the International Society on Presence Research, Philadelphia, USA. .

64 Lachlan K, Krcmar M. Experiencing Presence in Video Games: The Role of Presence Tendencies, Game Experience, Gender, and Time Spent in Play. *Communication Research Reports* 2011; **28**:27–31.

65 Viaud-Delmon I, Ivanenko YP, Berthoz A, Jouvent R. Sex, lies and virtual reality. *Nature neuroscience* 1998; **1**:15–6.

66 Nijs IMT, Franken IHA, Muris P. Food cue-elicited brain potentials in obese and healthy-weight individuals. *Eating behaviors* 2008; **9**:462–70.

67 Svaldi J, Tuschen-Caffier B, Peyk P, Blechert J. Information processing of food pictures in binge eating disorder. *Appetite* 2010; **55**:685–94.

68 Kullmann S, Schweizer F, Veit R, Fritsche A, Preissl H. Compromised white matter integrity in obesity. *Obesity Reviews* 2015; **16**:273–81.

B. Grasping Multisensory Integration: Proprioceptive Capture after Virtual Object Interactions

The manuscript has been submitted to the 39th Annual Meeting of the Cognitive Science Society and was accepted for publication in the conference proceedings after revision. The proceedings are not published yet (20th of July 2017), after publication, the article will be available as:

Lohmann, J., Gütschow, J., & Butz, M. V. (2017): **Grasping Multisensory Integration: Proprioceptive Capture after Virtual Object Interactions**. Proceedings of the 39th Annual Meeting of the Cognitive Science Society. London, UK: Cognitive Science Society.

Grasping Multisensory Integration: Proprioceptive Capture after Virtual Object Interactions

Johannes Lohmann (johannes.lohmann@uni-tuebingen.de)
Department of Computer Science, Cognitive Modeling, Sand 14
Tübingen, 72076 Germany

Jakob Gütschow (jakob.guetschow@student.uni-tuebingen.de)
Department of Computer Science, Cognitive Modeling, Sand 14
Tübingen, 72076 Germany

Martin V. Butz (martin.butz@uni-tuebingen.de)
Department of Computer Science, Cognitive Modeling, Sand 14
Tübingen, 72076 Germany

Abstract

According to most recent theories of multisensory integration, weighting of different modalities depends on the reliability of the involved sensory estimates. Top-down modulations have been studied to a lesser degree. Furthermore, it is still debated whether working memory maintains multisensory information in a distributed modal fashion, or in terms of an integrated representation. To investigate whether multisensory integration is modulated by task relevance and to probe the nature of the working memory encodings, we combined an object interaction task with a size estimation task in an immersive virtual reality. During the object interaction, we induced multisensory conflict between seen and felt grip aperture. Both, visual and proprioceptive size estimation showed a clear modulation by the experimental manipulation. Thus, the results suggest that multisensory integration is not only driven by reliability, but is also biased by task demands. Furthermore, multisensory information seems to be represented by means of interactive modal representations.

Keywords: Multisensory Integration; Multisensory Conflict; Object Interaction; Virtual Reality

Introduction

Adaptive interaction with the environment requires the combination of various sensory signals. According to theories of predictive coding, this integration is driven by a desire for consistency between internal models and the external world (Friston, 2010), as well as by a desire for consistency across different internal models (Butz, Kutter, & Lorenz, 2014; Ehrenfeld, Herbort, & Butz, 2013). Research on the mechanism of multisensory integration has shown that this consistency is achieved in terms of a maximum likelihood integration which combines different sensory signals based on their respective reliability estimates, resulting in a Bayesian estimate about the state of the external world (Ernst & Banks, 2002; Ernst & Bülthoff, 2004). It is still debated, however, whether this estimate is represented by means of an integrated representation (Cowan, 2001) or by means of separate, modality specific representations which are integrated on demand (Baddeley & Hitch, 1974). Experimental results show strong interactions between modalities in the internal representation, for instance between visual and auditory working memory (Morey & Cowan, 2005). Furthermore, unimodal retrieval from a multisensory representation is affected by previous modal encodings (Thelen, Talsma, & Murray, 2015). Quak, London, and Talsma (2015) suggest that task requirements typically determine whether a unimodal or a complex, multisensory representation is formed.

Our aim in the present study was two-fold. First, we wanted to investigate whether multisensory integration is modulated by task relevance. Second, we wanted to probe the nature of the stored representations. To investigate these questions, we combined an object interaction task involving multisensory conflict with a size estimation task. We let participants perform a grasp-and-carry task in an immersive virtual reality, by tracking the hands of the participants. Conflict was introduced in terms of a visual offset, either expanding or shrinking the visual grip aperture, thereby dissociating vision and proprioception. Moreover, we augmented the object interaction with vibrotactile feedback, which signaled when the relevant object was grasped. After the object interaction, we let participants judge the size of the object they interacted with either visually or based on the grip aperture. If vision and proprioception are integrated, visual estimates should be biased in the same way as proprioceptive estimates. On the other hand, if there was no bias in visual estimates, this would imply an independent storage of modal information.

Method

Participants

Twenty students from the University of Tübingen participated in the study (seven males). Their age ranged from 18 to 34 years (M = 22.1, SD = 3.9). All participants were right-handed and had normal or corrected-to-normal vision. Participants provided informed consent and received either course credit or a monetary compensation for their participation. Three participants could not complete the experiment due to problems with the motion capture system, only the data of the remaining 17 participants was considered in the data analysis.

Apparatus

Participants were equipped with an Oculus Rift© DK2 stereoscopic head-mounted display (Oculus VR LLC, Menlo

Park, California). Motion capture was realized by the combination of a Synertial IGS-150 upper-body suit and an IGS Glove for the right hand (Synertial UK Ltd., South Brighton, United Kingdom). Rotational data from the suit's and glove's inertial measurement units was streamed to the computer controlling the experiment via a Wifi connection. The data was then used to animate a simplistic hand model in a virtual reality. Since the IGS system only provides rotation data, we used a Leap Motion© near-infrared sensor (Leap Motion Inc, San Francisco, California, SDK version 2.3.1) to initially scale the virtual hand model according to the size of the participants' hands. To allow participants to confirm their size estimates without manual interactions, participants were equipped with a headset. Speech recognition was implemented by means of the Microsoft Speech API 5.4. The whole experiment was implemented with the Unity® engine 5.0.1 using the C# interface provided by the API. During the experiment, the scene was rendered in parallel on the Oculus Rift and a computer screen, such that the experimenter could observe and assist the participants.

To provide the participants with vibrotactile feedback during object interactions, we used two small, shaftless vibration motors attached to the tip of the thumb and the index finger of the participants. The diameter of the motors was 10 mm, the height was 3.4 mm. The motors were controlled via an Arduino Uno microcontroller (Arduino S.R.L., Scarmagno, Italy) running custom C software. The microcontroller was connected to the computer via a USB port which could be accessed by the Unity® program. If a collision between the virtual hand model and an object was registered in the VR, the respective motor was enabled with an initial current of 2.0 V. The deeper the hand moved into the object, the higher the applied current (up to 3.0 V) and the according vibration. At a current of 3.0 V, the motors produced a vibration with 200 rotations per second, the resulting vibration amplitude was 0.75 g. The wiring diagram as well as additional information regarding the components are available online. [1]

Virtual Reality Setup

The VR scenario put participants in a small clearing covered with a grasslike texture, surrounded by a ring of hills and various trees. A stylized container was placed in the center of the scene and served as target for the transportation task (see Fig. 1, left panel). The to-be-grasped and carried object was a cube rendered with a marble texture. The size of the cube varied from trial to trial but the cube always appeared at the same position in the scene. Textual information, like trial instructions and error feedback were presented on different text-fields aligned at eyeheight in the background of the scene.

Centered at the participants' hip[2], the task space covered 60 cm from left to right and 55 cm in depth. Corresponding to the data generated by the IGS suit an upper body rig was placed in the scene. It was positioned about 45 cm in front of the spawning position of the cube, slightly behind the the container. Hence, participants could reach both the container as well as the cube comfortably with their right arm. The rig itself was not rendered, only the right hand of the participants appeared in the scene visually.

The multisensory conflict between visual and proprioceptive grip aperture was realized in terms of a visual angular offset on the root joints of the thumb and index finger. They could be rotated either 10° towards each other, or away from each other. To maintain the same aperture, this visual offset had to be compensated by an adjustment of the actual aperture in the opposite direction. To compensate for a visual offset shrinking the grip aperture, the grip aperture had to be wider, while a visual offset extending the grip aperture required a closer grip aperture. In one third of the trials, no manipulation was applied (the different offset conditions are shown in Fig. 1, right panel).

Procedure

Participants received a verbal instruction at the beginning of the experiment regarding the use and function of the applied VR equipment. Then, they were equipped with the inertial motion capture system, consisting of the suit and the glove. If necessary, the finger sensors of the glove were fixated with rubber bands. After aligning the sensors and enabling the data streaming, the vibration motors were fastened underneath the thumb and index finger tip with rubber bands. Participants were then seated comfortably on an arm chair.

After this, participants were asked to hold their right hand over the Leap sensor to scale the virtual hand size according to their actual hand size. The control was then switched to the IGS system and participants put on the HMD to start the training phase. Participants could practice the grasping and carrying of the cube until they felt comfortable with the task. They had to complete at least 15 successful repetitions of the task before they were allowed to proceed. The grasp and carry task is described in detail in the next section.

After completing the training, the experimenter switched manually to the main experiment. The experiment consisted of eight blocks, each composed of 15 trials. The multisensory conflict between seen and felt grip aperture was introduced during the intertrial interval while the screen was blacked out.[3] In each trial participants had to grasp a cube and put it into the target container. After the object interaction, the scene faded out and one of two possible reproduction scenes

---

[1] http://www.wsi.uni-tuebingen.de/lehrstuehle/cognitive-modeling/staff/staff/johannes-lohmann.html

[2] Based on the inertial data from the IGS suit, it is possible to calculate a kinematic chain with the hips as root. Hence, the position of the hip joint in the virtual scene is the reference point for all body movements.

[3] While most participants remained unaware to the manipulation and attributed the variance in their grip aperture to inaccuracies of the tracking equipment, two participants reported to be aware of the manipulation after the experiment. Seeing that conscious awareness was not critical in this experiment, we did not perform a behavioral manipulation check in terms of a signal detection task to determine whether participants were able to consciously detect the manipulation of the visual grip aperture.

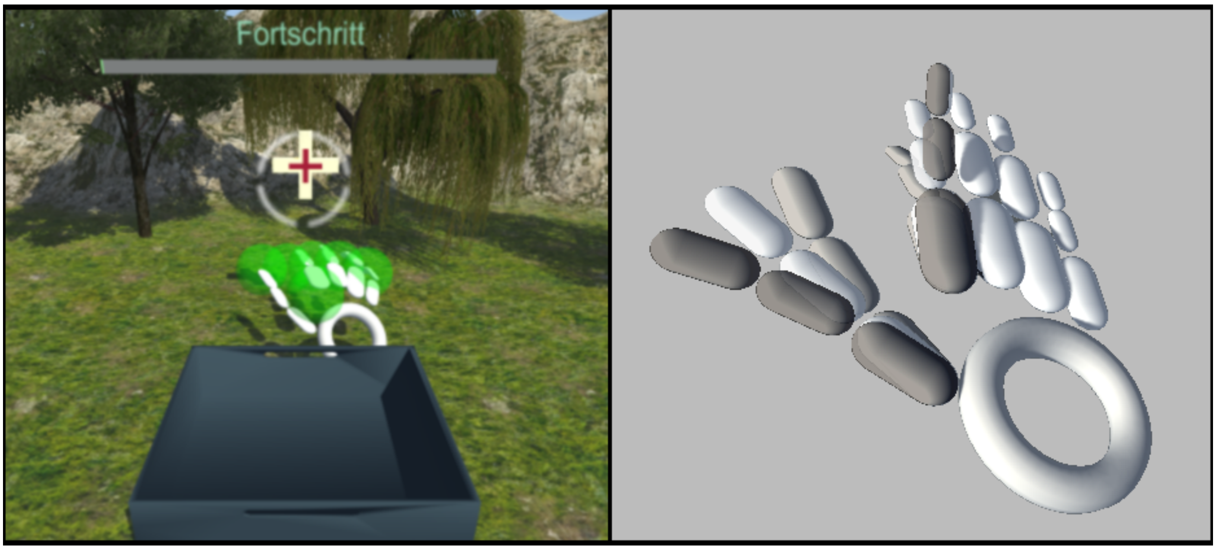


Figure 1: The left panel shows the VR scene and the initial position and fixation checks before the presentation of the target cube. Participants had to maintain a stable fixation on the fixation cross, the green spheres represent the starting position. The right panel shows the different offset conditions. Inward offsets are indicated by the light gray joints, dark gray joints indicate the outward offset condition.

appeared. This was independent of the success in the object interaction, the reproduction scene was also shown in case of error trials. In these scenes participants had to reproduce the size of the cube they interacted with either visually or by indicating the size in terms of a grip aperture. After each block, there was a break of at least ten seconds, after the fourth block, a longer break of at least two minutes was administered. Participants were allowed to put off the HMD during the breaks. After the experiment, participants were asked to complete a presence questionnaire (IPQ, Schubert, Friedmann, & Regenbrecht, 2001). The whole procedure took 90 to 120 minutes, including the preparation and the practice trials.

Grasp and Transportation Task At the beginning of each trial, participants had to move their right hand into a designated starting position, consisting of red, transparent spheres indicating the required positions of the fingers and the palm. The spheres turned green when the respective joints were in position. Furthermore, participants had to maintain a stable looking direction on a fixation cross (see Fig. 1, left panel). When both requirements were met, the fixation cross as well as the visible markers of the initial position disappeared and the target cube appeared. Participants were instructed to grasp the cube with a pinch grasp and to move it into the target container. A successful pinch required the tips of the thumb and the index finger to be placed on opposite sites of the cube and to maintain a stable grip aperture. Participants received vibrotactile feedback whenever touching the cube. The feedback scaled with the depth of penetration, becoming more intense the deeper the fingers were moved into the cube. The task was successfully completed by placing or dropping the cube into the container. Success was indicated by the cube bursting into an explosion of smaller green cubes. Interactions were canceled if the cube was penetrated overly strongly, dropped outside the container, moved outside the reachable space (e.g. by throwing it), or in case the interaction took more than 20 seconds. If one of the conditions was met, participants received error feedback and the trial progressed with the reproduction task.

After completing or failing the interaction, the markers for the initial position reappeared and participants had to move their hands back into the initial position. Then a visual mask was applied, accompanied by random vibrations on the finger tips. The visual and tactile masking commenced for one second. After the masking the scene faded to black and after one second, one of the two reproduction scenes appeared. The offset manipulation was removed during the blank interval.

Size Estimation In both versions of the size estimation task, participants had to reproduce the cube size. For the visual reproduction, the scene was similar to the one in which the interaction took place. However, the ground textures were replaced and different tree models were used to avoid possible comparisons between the cube size and external landmarks. A cube was placed at the center of the scene, at the same position where the cube during the interaction phase appeared. Above the cube, a slider was displayed, which allowed the participants to scale the cube by dragging the slider button with their fingertips. The slider spanned approximately 20 cm from left to right. The initial position of the slider button and thus the initial size of the visual reference cube was determined by the cube size during the interaction phase. For the smaller three sizes the slider started out at 10% and for the two larger sizes it started out at 90% of the sliding range.

For the proprioceptive reproduction, all visuals were deactivated (including the hand model), only the horizon as well as small white sparks in the center of the scene remained active to remind the participants that the experiment was still running. Participants were instructed to indicate the size of the cube they interacted with by means of the grip aperture between thumb and index finger. To confirm their estimate, participants were requested to say the German word for "continue" or "done" ("weiter" or "fertig"). The voice control identified these commands and ended the trial, recording either the slider position - indicating the visual edge length of the cube - or the grip aperture as the size estimate.

Factors

We varied three factors across trials. First, the edge length of the cube, which had to be interacted with and which size had to be estimated, was either 7 cm, 7.35 cm, 7.7 cm, 8.05 cm, or 8.4 cm. Second, the visual grip aperture was either shrunk, or extended by 10°, or corresponded with the felt grip aperture. In the following, we will refer to visual offsets shrinking the aperture as inward offsets, conversely, we will refer to offsets extending the aperture as outward offsets. Third, we varied the reproduction modality, which could either be visual or proprioceptive. Hence, the experiment followed a 5 × 3 × 2 within-subject design. Each of the 30 conditions was repeated four times, resulting in 120 trials. The trial order was randomized.

Dependent Measures

Besides the size estimates in the two different reproduction conditions, we obtained several time measures. Movement onset was determined as the time between the end of the fixation until leaving the starting position. Contact time refers to the time between movement onset and successful grasp. Interaction time refers to the time interval between the grasp and reaching the container.

Results

Data was aggregated according to the 5 × 3 × 2 within-subject design. Seeing that the size estimation had to be performed after error trials as well, there are no missing data with respect to the size estimates. For the duration measures, only correct trials were considered. The overall error rate was high (nearly 30%), due to the task complexity. In case of missing time data, the respective cell mean was interpolated within participants by the mean over all conditions with the same offset type. For all dependent measures, values differing more than two times of the standard deviation from the mean were excluded, which was the case for 2% of all data points.[4]

Size estimates, time measures, and error rates were analyzed with repeated measures ANOVAs using R (R Core Team, 2016) and the *ez* package (Lawrence, 2015). All post-hoc t-tests were adjusted for multiple comparisons by the method proposed by Holm (Holm, 1979). Results from the presence questionnaire were compared with the reference data from the online database.[5] There were no significant differences.

Table 1: ANOVA table for the analysis of the **size estimates**. The assumption of sphericity was violated for the cube size factor and the interaction between offset and reproduction condition, the according p-values were subjected to a Greenhouse-Geisser adjustment.

factor	df	F	p	η_p^2
size	4	34.84	$< .001^*$	.69
offset	2	17.55	$< .001^*$	.52
repro. type	1	0.48	.50	.03
size× repro. type	4	2.94	$.027^*$	.16
offset × repro. type	2	3.95	$.045^*$	.20
size × offset	8	1.03	.42	.06
size × offset × repro. type	8	2.35	$.022^*$	.13

Size Estimates

Data were analyzed with a 5 (cube size) × 3 (offset) × 2 (reproduction type) factors repeated measures ANOVA. Results are shown in Tab. 1. The analysis yielded significant main effects for cube size and offset. The main effect for cube size matches the actual cube size: larger cubes were estimated larger and smaller cubes were estimated smaller. To check if the estimates were veridical, we tested whether the estimated cube sizes differed from the actual cube sizes. None of the respective comparisons yielded significant results.

With respect to the main effect of offset, participants overestimated the cube size in case of inward offsets, compared to conditions with no offset ($t(16) = 3.45$, $p = .007$). For outward offsets participants underestimated the cube size, compared to conditions with no offset ($t(16) = 2.98$, $p = .009$). Finally participants provided larger estimates in case of inward, compared to outward offsets ($t(16) = 5.23$, $p < .001$).

Both, cube size and offset interacted with the reproduction condition. The interaction between cube size and reproduction type is due to a systematic overestimation of the larger cubes in case of the visual reproduction. In both cases, the estimates are significantly larger than the actual sizes of 8.05 cm ($t(16) = 4.26$, $p = .003$), and 8.4 cm ($t(16) = 3.21$, $p = .022$), respectively.[6]

The interaction between reproduction condition and offset was further analyzed with post-hoc t-tests. Estimates in case of outward offsets were significantly smaller than in case of

---

[4] Please note that the data pattern remains nearly unaffected if the data is not filtered. Removing the size estimates from error trials only reduces the effect size of the three-way interaction.

[5] Available at http://www.igroup.org/pq/ipq/index.php

[6] The considerable overestimation might be partially due to the initial slider position in the visual reproduction, starting at 90% of the sliding range for larger cubes.

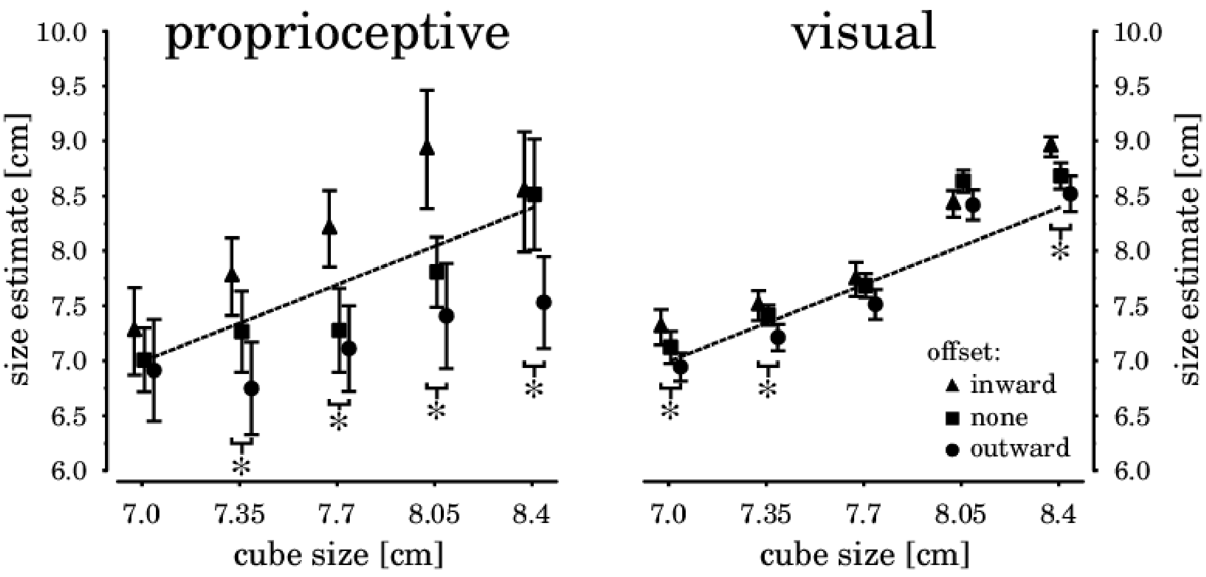


Figure 2: Three-way interaction between reproduction condition, cube size and offset. Significant differences with $p < .05$ between estimates in case of inward and outward offsets are indicated by an asterisk. The respective t-tests were one-sided (inward > outward) and were adjusted for multiple comparisons. The dashed line indicates the actual cube size.

inward offsets, both, for visual ($t(16)$ = -2.21, $p = .021$), as well as for proprioceptive ($t(16)$ = -5.48, $p = .002$) reproduction. However, the differences between the offset conditions were much more pronounced in case of proprioceptive reproduction, resulting in the observed two-way interaction.

This pattern of results was modified by a three-way interaction between cube size, offset and reproduction condition. Separate ANOVAs for the different cube sizes showed that the interaction between reproduction condition and offset was only present for cubes of intermediate (7.7 cm) and large size (8.05 cm). For these two conditions, there were no significant differences between the offset conditions in case of visual reproduction. The differences for proprioceptive reproduction remained significant. The main effect of offset, however, remained significant for all of these separate analyses.

With respect to our hypotheses, the difference between inward and outward offsets is most relevant. To check whether inward offsets always result in larger estimates than outward offsets, we checked whether the respective difference is significant for the five different cube sizes, separately for the two reproduction conditions. In case of proprioceptive reproduction, the difference is significant for all cube sizes, except the smallest one of 7 cm. For visual reproduction the differences reached significance for all cube sizes, except the intermediate (7.7 cm) and large size (8.05 cm). The results are shown in Fig. 2.

Time Measures

Data were analyzed with a 5 (cube size) × 3 (offset) factors repeated measures ANOVA. No significant effects were found for the movement onset times. The analysis of object contact times yielded a significant main effect for offset ($F(2,32)$ = 76.57, $p < .001$, η_p^2 = .83). Slowest contact times were observed for outward offsets, while inward offsets yielded the fastest response times. All of the respective pairwise comparisons yielded significant results. The analysis of the interaction times yielded a significant main effect for offset as well ($F(2,32)$ = 4.90, $p < .014$, η_p^2 = .23). Participants were slower in transporting the cube in case of outward offsets. Post-hoc t-tests showed that the interaction times were significantly elevated in case of outward offsets, both compared to inward offsets ($t(16)$ = 2.39, $p = .042$), as well as to trials without offset ($t(16)$ = 2.42, $p = .042$).

Error Rates

The analysis of the error rates yielded significant main effects for cube size ($F(4,64)$ = 4.27, p = .004, η_p^2 = .21) and offset ($F(2,32)$ = 12.22, $p < .001$, η_p^2 = .43). In general, participants made fewer errors during interactions with larger cubes. Furthermore, error rates were higher in case of inward offsets. Post-hoc t-tests showed that error rates increased for inward offsets, when compared to both outward offsets ($t(16)$ = -3.67, $p = .004$), and no offsets ($t(16)$ = -4.56, $p < .001$).

General Discussion

Previous studies on multisensory integration have shown a dominance of visual information in the perception of object size (e.g. Ernst & Banks, 2002). To investigate whether task demands, which require to focus on another modality, can reduce this dominance, we let participants perform a grasp-and-carry task under multisensory conflict between vision and proprioception. In order to do so, we manipulated the mapping between seen and felt grip aperture. After the object interaction we let participants estimate the size of the object they interacted with – either visually or by providing a proprioceptive estimate via grip aperture. Our results show a systematic bias in the size estimates due to the introduced offset between seen and felt grip aperture. A wider grip aperture resulted in object size overestimations, while a smaller aperture yielded underestimations. This was true for both, visual and proprioceptive size estimates. Hence, the adaptation of the size estimation followed the proprioceptive adaptation, which was necessary to compensate for the visual offset.

While the offset manipulation led to different actual grip apertures for cubes of the same size, the visual impression of both the cube size and the grasp of the virtual hand remained the same. Thus, if the size estimate was dominated by the visual impression, there should have been no effect of the offset condition in the visual reproduction trials. In contrast, our results show a clear influence of proprioceptive information on the size estimates in both modalities. However, this influence was much more pronounced in the case of the proprioceptive reproduction. Apparently, proprioceptive information dominated the resulting percept, even if proprioception was much noisier than vision, indicated by the comparatively large variance in the proprioceptive size estimates.

The combination of VR with motion capturing enabled us to dissociate vision and proprioception in an interactive setup. Compared to previous studies, which investigated the effects of mismatching sensory information regarding an object, the applied setup allows to manipulate the own body perception without affecting the visual impression of the external, virtual world. Some issues with respect to the experimental setup remain. The high error rates imply that even with the vibrotactile augmentation, the object interaction remained difficult for the participants. Especially in case of outward offsets, participants took quite long to grasp and carry the cube. The error rates were elevated for inward offsets, which were associated with the fastest grasping and interaction times, implying a speed accuracy trade-off. Furthermore, our setup did not comprise a control condition without grasping. Including trials which only require touching the object will clarify whether the mere presence of a graspable object yields a bias towards proprioceptive information, or if performing the actual interaction is necessary to induce the bias.

Despite these issues, the results allow us to draw the following two conclusions. First, visual and proprioceptive information regarding the object size seem to be stored separately, but are able to affect each other. If there was only a single percept reflecting the cube size across modalities, then the reproduced size should be independent of the reproduction modality. This is clearly not the case, given the huge difference in the variance of the visual and proprioceptive estimates and the stronger bias in proprioceptive compared to visual reproduction. This conclusion dovetails with results reported by (Ernst & Banks, 2002), who showed that sensory data are stored separately, when they originate from different modalities. Second, the integration process that produces a visual or a proprioceptive estimate is influenced by the type of reproduction. The considerable difference between the effect sizes implies a different weighting of the modality-specific encodings in the two reproduction conditions.

References

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation*, *8*, 47–89.

Butz, M. V., Kutter, E. F., & Lorenz, C. (2014). Rubber hand illusion affects joint angle perception. *PloS One*, *9*(3), e92854.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and brain sciences*, *24*(1), 87–114.

Ehrenfeld, S., Herbort, O., & Butz, M. V. (2013). Modular neuron-based body estimation: maintaining consistency over different limbs, modalities, and frames of reference. *Frontiers in Computational Neuroscience*, *7*(Article UNSP 148).

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*(4), 162–169.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 65–70.

Lawrence, M. A. (2015). ez: Easy analysis and visualization of factorial experiments [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=ez` (R package version 4.3)

Morey, C. C., & Cowan, N. (2005). When do visual and verbal memories conflict? the importance of working-memory load and retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 703.

Quak, M., London, R. E., & Talsma, D. (2015). A multisensory perspective of working memory. *Frontiers in human neuroscience*, *9*.

R Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence*, *10*(3), 266–281.

Thelen, A., Talsma, D., & Murray, M. M. (2015). Single-trial multisensory memories affect later auditory and visual object discrimination. *Cognition*, *138*, 148–160.

C. Lost in Space: Multisensory Conflict yields Adaptation in Spatial Representations across Frames of Reference

The manuscript has been published in *Cognitive Processing*. The final publication is available at Springer via http://dx.doi.org/10.1007/s10339-017-0798-5 and can be retrieved by the following reference:

Lohmann, J., & Butz, M. V. (2017): **Lost in Space: Multisensory Conflict yields Adaptation in Spatial Representations across Frames of Reference**. *Cognitive Processing*, 1-18. doi: 10.1007/s10339-017-0798-5.

Noname manuscript No.
(will be inserted by the editor)

Lost in Space: Multisensory Conflict yields Adaptation in Spatial Representations across Frames of Reference

Johannes Lohmann · Martin V. Butz

Received: date / Accepted: date

Abstract According to embodied cognition, bodily interactions with our environment shape the perception and representation of our body and the surrounding space, that is, peripersonal space. To investigate the adaptive nature of these spatial representations, we introduced a multisensory conflict between vision and proprioception in an immersive virtual reality. During individual bimanual interaction trials, we gradually shifted the visual hand representation. As a result, participants unknowingly shifted their actual hands to compensate for the visual shift. We then measured the adaptation to the invoked multisensory conflict by means of a self-localization and an external localization task. While effects of the conflict were observed in both tasks, the effects systematically interacted with the type of localization task and the available visual information while performing the localization task (i.e. the visibility of the virtual hands). The results imply that the localization of one's own hands is based on a multisensory integration process, which is modulated by the saliency of the currently most relevant sensory modality and the involved frame of reference. Moreover, the results suggest that our brain strives for consistency between its body and spatial estimates, thereby adapting multiple, related frames of reference, and the spatial estimates within, due to a sensory conflict in one of them.

Keywords Spatial Perception · Peripersonal Space · Multisensory Integration · Multisensory Conflict · Virtual Reality

---

J. Lohmann
Department of Computer Science, Cognitive Modeling, Sand 14 72076 Tübingen
Tel.: +49 7071 - 29 70481
Fax: +49 7071 - 29 5719
E-mail: johannes.lohmann@uni-tuebingen.de

M. V. Butz
Department of Computer Science, Cognitive Modeling, Sand 14 72076 Tübingen
Tel.: +49 7071 - 29 70429
Fax: +49 7071 - 29 5719
E-mail: martin.butz@uni-tuebingen.de

1 Introduction

Adaptive interaction with the environment requires the combination of various sensory signals and motor commands. To reach for an object, for example, it is necessary to transform a visually-grounded location estimate into a motor command. During this process, visually-based spatial information needs to be transferred into postural, that is, proprioceptive information, which corresponds to an inverse kinematic frame of reference transformation. The required computations are complex because the frames of reference are not only grounded in different sensory modalities, but they also differ in the origin of the respective coordinate system [10, 36]. Various studies and computational models indicate that transformations between these frames of reference are learned by sensorimotor interactions and are used both to predict action outcomes and to generate goal-directed behavior [2, 5, 11, 18].

Many experiments have shown that encodings of body-relative spatial representations integrate multisensory information and are highly adaptive. Research on the space surrounding the body – the so-called *peripersonal spaces*, including spatial representations surrounding body parts and reachable space – has revealed close interactions between motor codes, vision, proprioception, and kinesthetics [see 21, 35, for overviews]. Single cell recordings in monkeys, but also multisensory interference paradigms in humans, have shown that peripersonal encodings in premotor and parietal cortical areas are responsive to nearby stimuli relative to a particular body surface, anticipate approaching stimuli, and partially relocate their receptive fields during tool usage [7, 15, 21, 24].

Other experiments have shown that multisensory conflict between vision and proprioception can bias reaching movements toward the manipulated visual hand position. Usual methods to generate multisensory conflict involve virtual reality [17], prism goggles [40], or mirror reflections of a hand [22, 23]. Another common approach to induce multisensory conflict is the rubber hand illusion [RHI, 4]. In this case, participants watch a rubber hand being stroked, while their own unseen hand is stroked synchronously. This procedure induces a relocation of the perceived position of the participants' own hand toward the rubber hand. More recent results imply that this adaptation is not restricted to a hand-centered frame of reference but also affects other frames of reference, such as the elbow angle [6]. These results imply that a postural body schema [21] is involved, which is used to project the false visual information about the location and orientation of ones hand into other frames of reference, and the encodings of related current body state estimations within these frames of reference – such as postural joint angle estimations.

Besides sensory information, action possibilities and action plans also affect the representation and perception of peripersonal space. Distance perceptions depend – to a certain extent – on the effort necessary to perform a certain action. For example, the perceived distance increases with the required amplitude of a hand movement [27, 26]. In a similar vein, it has been shown that perceived size of graspable objects scales with the currently perceived hand size [30]. Furthermore, the perceived reachability of an object depends not only on actual distance, but also on motor capabilities: In [8], participants received a TMS stimulation on the left premotor and motor cortex while judging whether a visual target was graspable or not. Stimulation over the motor areas interfered with the reachability judgment when the target was close to the boundary of peripersonal space. For very close and very far targets, however, this interference was absent. This implies a direct contribution of the motor system in visual reachability judgments, at least when the decision process is complex. The available results thus suggest that representations of peripersonal space also interact with motor encodings.

Interestingly, sensory information sources are weighted in dependence of the stimulus distance from one's own body. It has been shown that the visual component dominates the representation of extrapersonal space, while the representation of peripersonal space relies more on proprioception [25, 32]. This dissociation between peripersonal and extrapersonal space has been studied, for example, based on the phenomenon of pseudo-neglect [see 25, for an overview]. Participants show a systematic leftward bias for the midpoint in line bisection tasks when the lines are presented within peripersonal space. This bias decreases with increasing distance and is reversed for lines in extrapersonal space. Moreover, several studies suggest that the extent of the currently perceived peripersonal space scales with the currently perceived own arm length [33, 34]. It thus appears that peripersonal and extrapersonal spatial representations depend on one's current body morphology and body representation. Within these spatial representations, the weighting – or saliency – of visual information increases with the distance from one's own body.

In sum, research has shown that the currently active peripersonal spatial encodings integrate multisensory information sources as well as motor information in a highly interactive and versatile, adaptive manner. Furthermore, results regarding the RHI [6] as well as the adaptive extension and contraction of peripersonal space [21, 34] imply that currently active spatial encodings do not only adapt within the manipulated spatial representation, but also other related spatial encodings adapt accordingly to certain degrees.

In the presented study, we investigated in further detail to which extent other spatial encodings adapt to a multisensory conflict. We intended to measure spatial encoding adaptations without impairing one of the conflicting modalities. Classic approaches like the RHI allow to introduce multisensory conflict in visually-grounded hand space. However, possible aftereffects can only be studied when the manipulated hand remains invisible. Thus, it cannot be studied how participants would use the manipulated hand in interaction or pointing movements under visual control. In contrast, experiments applying prism goggles allow this kind of test, because the hands remain visible during multisensory conflict. However, prism goggles introduce a general inconsistency between vision and proprioception throughout many different frames of reference, because the whole visual field is shifted. Virtual reality (VR) in combination with online motion capturing allows to combine the advantages of the RHI and prism goggles. In such a setup, the visual hand position can be manipulated selectively without affecting other objects in the visual field. The artificial hands can be used under visual control without disrupting the multisensory conflict. Moreover, it is possible to investigate the saliency of the manipulated visual information by hiding the hands in some trials and showing them in others. Previous studies imply the general validity of such VR setups. For example, it has been shown that an altered body morphology in VR can affect spatial representations [31] and that the distinction between peri- and extrapersonal space remains valid [16].

We thus manipulated the mapping between visual and proprioceptive hand position, producing a multisensory conflict regarding the hand position. In order to do so, we combined an immersive virtual reality with online motion capturing. Participants had to perform a complex, bimanual task during which the visual hand representations were increasingly shifted, resulting in an according, increasing correction of the actual hands to maintain the target position in the VR. Based on the results of Butz et al [6], we expected that the conflict between vision and proprioception will result in adaptations of the spatial representations throughout the kinematic chain.

We thus investigated to which extent this conflict yields adaptation effects not only in dominantly visual but also in dominantly proprioceptive spatial representations and the respective frames of reference. To do so, we had participants localize both, an external object shown in the VR and their own body, before and after introducing the visual conflict during the bimanual task. Participants simply had to point toward the believed locations. To explore the role of visual saliency during the pointing tasks, we hid the virtual hand models during the localization task in half of the trials. We expected larger errors during external object localizations in contrast to self-localizations, because the weighting of the manipulated visual information should be stronger seeing that the target location is also visually encoded. Moreover, we expected stronger effects in the case of visible hands, because the shifted visual information should continue to strongly influence the veridical proprioceptive information. Nonetheless, we still expected to measure significant effects of the introduced conflict even with invisible hands, since the shifted visual information about the hand is never explicitly falsified. Even in the case of self-localization with invisible hands we expected to measure significant location adaptations – albeit to a lesser degree – because the visual conflict of the hand's location should extend to the proprioceptive modality along the body's kinematic chain [in accordance with 6].

2 Method

2.1 Participants

33 students from the University of Tübingen participated in the study (22 males). Their age ranged from 18 to 30 years (M = 21.7, SD = 2.5). All but one participants were right-handed and all had normal or corrected-to-normal vision. Participants provided informed consent and received course credit for

their participation. Before the experiment the participants were briefed with a cover story, stating that the purpose of the study was the evaluation of different hand-models in virtual object interactions. After the experiment, participants were debriefed and the manipulation was explained to them. They were then offered to withdraw their data. No participant withdrew the data.

2.2 Virtual Reality Setup

To immerse participants in the virtual reality, they were equipped with an Oculus Rift © DK2 stereoscopic head-mounted display (Oculus VR LLC, Menlo Park, California). To allow object interaction, hand motions were captured with a Leap Motion © near-infrared sensor (Leap Motion Inc, San Francisco, California, SDK version 2.3.1), placed 30 cm in front of the participants on a table. Participants were seated comfortably in an arm-chair, the height was adjusted so the participants could put their hands in the center of the sensors' tracking range with their elbows resting on the chair. The room was dimly lit to avoid interference between the Leap sensor and external infrared light.

So far, there are only few systematic evaluations of the tracking accuracy of the Leap Motion sensor. According to Guna et al. [19], the standard deviation of positional measurements is about 0.5 mm in the center of the tracking area. The frame rate of the sensor is not constant and cannot be adjusted manually. With all power saving options disabled, the target frame rate is in the range of 115 Hz. In our preliminary tests, the sampling rate was always in the range of 100 Hz. We never observed rates as low as reported by Guna et al. (average around 40 Hz). The whole experiment was implemented with the Unity ® engine 5.2.2 using the C# interface provided by the API. During the experiment, the scene was rendered in parallel on the Oculus Rift and a computer screen, such that the experimenter could observe and assist the participants.

The VR scenario put participants in a static mountain scenery, with a basket at the outer right corner of their reachable task space. During the experiment a flower spawned at the center of the scene and participants had to pick the petals and put them into the basket (see Fig. 1). For this interaction, their hands were rendered with three different hand models, obtained from the Leap VR assets. During the training a robotic hand model was used (the one displayed in the figures). The alternative hand model was either a stylized, minimal hand model, or another robot hand model (see the middle and right hand model in Fig. 2). The order in which block which hand model was applied, varied randomly and was balanced between participants. Hand models only affected the visual appearance of joints and bones, the underlying kinematics remained the same. The flower model and the scripts necessary to simulate its growth were obtained from the Leap Motion assets.

Centered at the sensor, the task space covered an area of 70 cm from left to right and 50 cm in depth. The flower spawned at the sensor position and

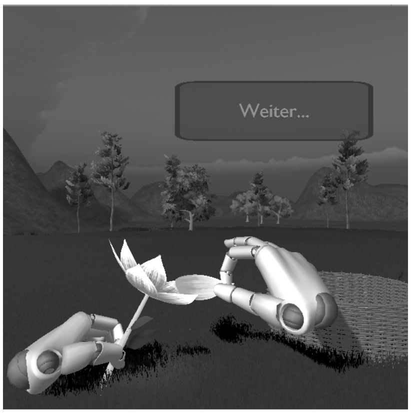


Fig. 1: Object interaction in the VR. In each trial, the participants' task was to collect as many petals as possible within 40 seconds. The petals could only be picked, when the stem of the flower was grasped. During the object interaction an offset was introduced, dissociating the hand visualization and the actual hand position above the sensor. The picture was obtained from the training trials, the visible continue button was not displayed during the main part of the experiment.

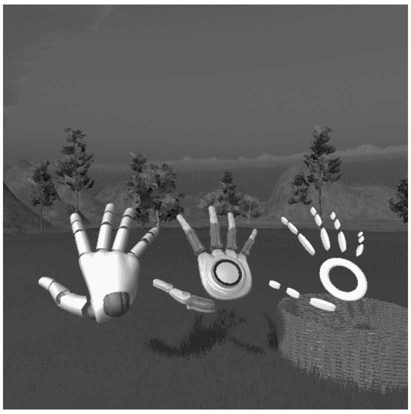


Fig. 2: The different hand models applied in the experiment. The leftmost hand model was used during training. The other two models were used in the two blocks of the experiment. The hand model remained the same within one block. The order of applied models was balanced across participants.

the basket was 40 cm away from the sensor position. The extent of the task space is depicted in Fig. 3.

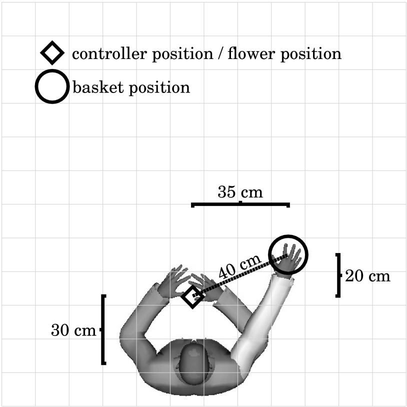


Fig. 3: Outline of the task space. Participants were seated 30 cm in front of the sensor. Centered around the sensor, the task space covered an area of about 50 cm in depth and 70 cm from left to right. The distance between the center of the task space - where the flower spawned - and the basket was about 40 cm.

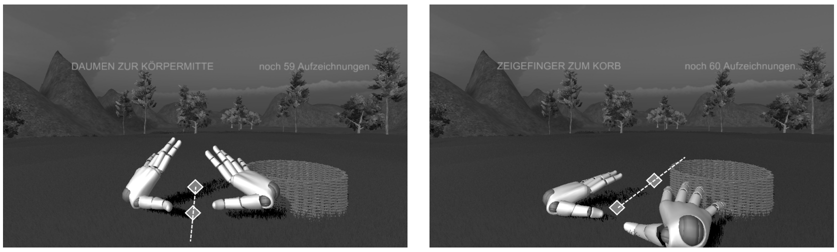
2.3 Procedure

Each trial consisted of three stages. At the beginning, participants had to perform a localization task. Next, they were asked to pick as much petals from the shown flower as possible. During this object interaction phase, the experimental manipulation of the visual hand model was done. Finally, participants had to repeat the localization task. The two tasks are described in detail below. The visual offset was reset at the end of the trial, hence the initial localization always commenced with veridical visual information regarding the hand position. The experiment consisted of two blocks, each consisting of 12 trials. Before the actual experiment, participants were given time to train the localization and the object interaction tasks. Participants were given at least 15 minutes to train both tasks, most of the participants spent about 10 minutes training the localization task. The whole experiment took about one hour. At the end of the experiment, participants were debriefed and were asked to complete a presence questionnaire.

2.3.1 Localization

At the beginning and at the end of each trial, participants had to locate themselves and an external reference within the scene. The order of the tasks varied from trial to trial and was balanced within the experiment. The localization was realized by pointing to the reference with both hands.

For the self-localization, participants were instructed to point with the tip of their thumbs to themselves, while keeping both hands stable and in the same height above the sensor. In case of the external reference, participants were



(a) Self-localization. (b) External localization.

Fig. 4: In the localization task, participants had to locate themselves within the scene (left image), or they had to point to the external reference, i.e. the basket (right image). During the localization, the position of the different hand joints was stored. Based on these data we obtained positional estimates for the palm, the index finger, and the thumb. Based on the averaged centroids, different dependent measures were calculated. To estimate the accuracy of the self-localization, a line was derived from the palms' and thumbs' centroids. Similarly, to estimate the accuracy of the external localization, the centroids of the palms and index fingers were used. The constructed lines are indicated by the dashed lines in the two images. The origins of the two lines are the palms' centroid.

instructed to point to the basket with their index fingers, again while keeping both hands stable and in the same height above the sensor. Both tasks are shown in Fig. 4a (self-localization) and Fig. 4b (external localization). Since the localization tasks were the primary data source of the experiment, we implemented several checks to allow reliable data collection. To have a robust estimate of the palm and finger positions, we averaged the data of 50 successive engine updates for a single estimate. With a frame rate of 60 Hz, a successful localization thus took at least 800 ms.

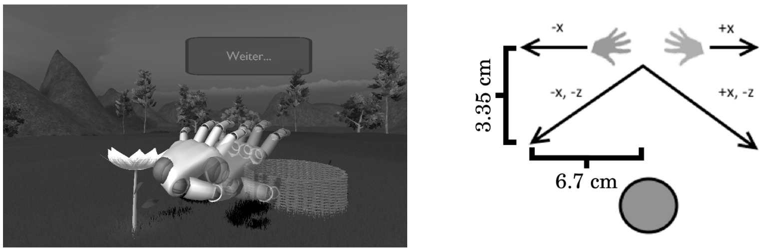
During the localization, participants received instructions and feedback via two text displays located at the upper left and the upper right of their visual field (see Fig. 4a and Fig. 4b). In the left display, the required localization was instructed, that is, either pointing to one's own body or pointing towards the basket. The number of remaining data samples was presented in the upper right in terms of a counter starting at 50. If the data collection was canceled, error feedback was given in a central, otherwise hidden text field.

To increase robustness and standardization of the recorded data, data collection was canceled and restarted when (a) participants moved their hands too much, (b) the hands were not parallel, (c) the fingers were not stretched, or (d) the fingers were bent upward or downward. Hand movements were registered when the positions of the palms, the thumb tips, and the index finger tips changed throughout one localization procedure by more than 1.7 cm compared to the initial position. Hands were considered to be parallel in the horizontal plane as long as they differed in height by less than 1.7 cm. Fingers were considered to be stretched as long as the bending angle between proximal finger joints and the respective finger tip was below 30°. To check whether a finger pointed upward or downward overly strongly, the angular distance between the pointing direction of the respective finger and the global vertical axis (y axis in Unity ®) was compared. When the angular distance became larger than 20°, the recording was canceled. When the data collection was canceled due to one these reasons, participants received feedback about the error and the localization procedure was restarted.

Besides these postural constraints, we also checked the data quality directly. The Leap sensor provides an internal confidence estimate for each hand measurement. When the confidence of a data point dropped below 60%, it was not collected.

The experiment was divided into two blocks. In one block, the hand model was shown during the localization tasks, while it was hidden in the other block. Seeing that the induced visual offset persisted throughout the end of the trial, participants performed the localization with the visual conflict present in the block where the had model remained visible. The block order varied randomly and was balanced across participants [1].

Due to the error checks, the localization task was quite challenging for the participants – especially with invisible hands. As a result, participants were given time to train both localization tasks, both with visible and invisible hands, until they felt comfortable with the procedure.



(a) Maximum horizontal offset.

(b) Offset conditions.

Fig. 5: The different offsets applied during the object interaction. The left image displays the maximum horizontal offset, the solid hand indicates the veridical position, while the transparent hand indicates the hand after the shift. The overall magnitude of the offset was rather small (6.7 cm on the x-axis, 3.35 cm on the z-axis), in the range of one hand's width. The right image displays the four different offset conditions. The filled circle at the bottom indicates the participant's position within the scene. All offsets were applied in the same direction to both hands. A positive offset on the x-axis shifted both hands to the right from the participant's point of view, while a negative offset on the z-axis shifted both hands toward the participant.

---

[1] Due to the uneven number of participants, there is one more data sample for the visible - invisible condition.

2.3.2 Object Interaction

After the initial localization in each trial was accomplished, a flower bloomed in the center of the scene. Participants were instructed to pick as many petals as possible and to put them into the basket. In order to do so, they had to grab the stem with the left hand and to pick the petals with the right hand (see Fig. 1 for an example). Whenever all petals of a flower were picked, a new flower bloomed at the center of the scene. According to the cover story, participants were made to belief that this was the main task of the experiment. Actually, the task was used to introduce the offset between visual and felt hand position. A visual offset to the left required a positional offset of the actual hands to the right to compensate and vice versa. In the following, we refer to the visual offset whenever we use the term offset.

Four different offset conditions were used, which varied from trial to trial and which were repeated three times per block. The offset was introduced gradually and only while the hands were moving. The maximum amplitude of the offset was 6.7 cm in the horizontal plane and 3.35 cm in depth.

More precisely, the offset was only increased when (a) at least one hand was grasping something – either the stem, a petal, or both, and (b) at least one hand was moving. The criteria for a hand to be considered moving were the same as in the localization task. The positions of the palms, the thumb tips, and the index finger tips were stored. A movement was registered when the distance between the current position and one of the stored reference positions changed by more than 1.7 cm. In this case, the new positions were stored as reference.

If both conditions were met, the offset was increased by a fraction of $\frac{1}{320}$, that is 0.02 cm in the horizontal plane, or 0.01 cm in depth. Hence, to introduce the full offset, it required 320 frames in which the mentioned conditions were met. Given a frame rate of 60 Hz, this translates to 5.5 seconds. The conditions could not be met in every frame, for instance while grasping a petal, the movement amplitude was too small for the hand to be considered moving. It typically took about ten seconds of hand movement and object interaction to introduce the offset in its maximum amplitude. The object interaction lasted for 40 seconds.

Due to this procedure, the offset was masked as much as possible, utilizing the principles of both, visual change blindness and inattentional blindness [44, 43]. Fig. 5a shows the maximum magnitude of the offset on the horizontal axis. The transparent hand indicates the shifted hand visualization. A schematic overview of the different offset conditions is shown in Fig. 5b. All offsets were applied to both hands. For instance, a positive offset on the x-axis shifted both hands to the right from the participant's point of view. After the object interaction, participants received encouraging feedback, depending on the number of picked petals. If the number of petals was below three, participants were encouraged to try harder. In the case of three to ten petals, participants were asked to maintain the good performance. For ten petals and more, participants were commended for their excellent performance. We did not perform a behavioral manipulation check. However, none of the participants reported to be aware of the offset at the debriefing.

At the end of each trial that is after the second round of the localization task, participants were asked to move their hands laterally out of the tracking range to proceed. After the participants' hands left the tracking range and were no longer visible within the scene, the visual offset was removed. Hence, every trial started with veridical visual information regarding the hand positions.

2.3.3 Presence Questionnaire

To get an idea of the quality of our VR setup and the immersion experienced by the participants, we assessed *presence* using the igroup presence questionnaire [IPQ, 42, 39] after the experiment. The IPQ essentially allows to quantify the degree of immersion experienced by the participants within the VR. It consists of three scales. Spatial presence refers to the sense of being physically present in the VR. Involvement indicates the amount of attention directed to the VR and the subjectively experienced involvement. Finally, experienced realism quantifies the experience of realism in the VR.

2.4 Dependent Measures

The positional data from the localization task was the primary data source. In each trial participants performed the self-localization and the external-localization before and after the interaction with the flower. As noted above, participants were requested to keep the localization pose until 50 valid data samples of the relevant points (palm position, index finger, and thumb tips) were collected. Please note that we collected the actual joint positions provided by the Leap sensor, *not* the positional data from Unity ® with the applied visual offset. Based on the gathered data, three dependent measures were derived to analyze the effects of multisensory conflict between vision and proprioception on spatial representations (see also Fig. 6).

The *palm drift* is the difference between the palm positions in the pre- and post-localization. A significant palm drift would indicate aftereffects of the visual manipulation in hand space. The *angular disparity* indicates how much participants corrected the drifted position during the location task by rotating their palms differently. Varying, partial compensations would imply adaptive, weighted integrations of visual and proprioceptive information. The *positional discrepancy* is a direct measure of changes in the localization, that is, how much the positional estimate changed between pre- and post-localization. Systematic changes in the localization would imply an adaptation of the estimated position of the target. The calculation of the different measures is described in detail below.

An overview of the gathered data is provided in Tab. 1. Please note that the numeric values reflect the systematic changes between pre- and post-

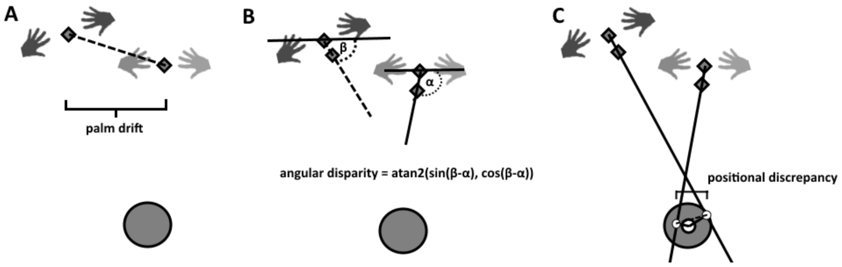


Fig. 6: Three different dependent measures were obtained from the raw data collected during the localization tasks. Darker hands indicate the post-test, while lighter hands indicate the pre-test hand positions. The filled circle at the bottom indicates the participant's position within the scene. We here show the self-localization task. In the external localization task, the calculations were essentially the same except for that the constructed line was generated by the palm and index finger centroids. **Panel A** shows the calculation of the palm drift, which is the signed, Euclidean distance between the palm centroids obtained in the pre- and the post-test. **Panel B** shows the calculation of the *angular disparity.* For the pre- and the post-test, the angle between the constructed line from palm to thumb and the horizontal axis was obtained. The angular disparity was calculated as the difference between these two angles. **Panel C** shows the calculation of the positional discrepancy. Extending the constructed line towards the target, the minimal distance to the target reference was obtained. The difference of these distances in the pre- and the post-test was used as the positional discrepancy measure.

localization. Positive values indicate a systematic change compensating for the offset, while negative values indicate a reversed adaptation effect.

2.4.1 Palm Drift

From the 50 data points per hand per localization task the centroid of the palms was obtained in terms of an [X,Z]-point, collapsing the vertical axis (y-coordinate). For the pre- and post-localization of the two references, the Euclidean distance between these two centroids was calculated (see Fig. 6, Panel A). The sign of the distance depended on the offset condition. If the post-centroid was to the left of the pre-centroid, the sign was positive in case of offsets to the right. For offsets to the left, the opposite was true.

Participants had to adjust the drift introduced during the object interaction. However, there was no need to compensate in the post-localization task, since correct localization could be achieved by rotating the hands. A significant drift in the post-localization – especially in the case of invisible hands – implies an adaptation in the spatial representation of the center of hand space.

Table 1: Means and standard errors of the primary measures in the different conditions.

vis.	ref.	offset	drift in [cm]		disparity in [deg]		discrepancy in [cm]	
			M	*SEM*	*M*	*SEM*	*M*	*SEM*
visible	self	right	2.05	0.37	2.06	0.33	0.25	0.49
		left	1.52	0.38	1.72	0.42	1.51	0.58
		toward right	1.76	0.39	2.07	0.45	1.23	0.87
		toward left	1.42	0.34	1.81	0.39	1.27	0.61
	ext.	right	2.45	0.34	3.99	0.76	1.13	0.72
		left	2.70	0.37	5.05	0.66	2.92	0.85
		toward right	1.74	0.36	3.42	0.55	1.04	0.71
		toward left	2.28	0.40	3.17	0.72	2.43	0.73
invisible	self	right	0.35	0.37	-0.09	0.39	0.41	0.70
		left	0.49	0.33	0.43	0.36	-1.13	0.62
		toward right	1.02	0.40	1.03	0.44	-1.29	0.77
		toward left	-0.01	0.33	0.35	0.42	1.04	0.75
	ext.	right	0.25	0.41	0.81	0.87	1.44	0.71
		left	0.59	0.35	1.39	0.70	0.76	0.51
		toward right	0.58	0.46	2.16	0.86	-0.35	0.57
		toward left	0.60	0.37	0.93	0.80	1.16	0.56

2.4.2 Angular Disparity

A possible drift of the hands does not necessarily yield mislocalizations, because it can be compensated for by rotating the hands. In order to point to the body center or to an external reference, location estimates have to be translated into motor commands in hand space. A compensation would imply that the invalid visual information regarding the hand position is partially compensated for by the veridical proprioceptive information. If there was no compensation at all, the participants fully relied on the manipulated visual information. Furthermore, it would imply that spatial estimates in other frames of reference, such as the own body's location, were shifted according to the visual offset in hand space. We expected a significant compensation in all cases, due to the mentioned expected interaction of proprioceptive and (manipulated) visual information.

To quantify this adaptation, we calculated the centroid of the palms and the centroids of the index fingers and thumbs, respectively. Based on these points, we obtained the rotation angles of the hands (see Fig. 6, Panel B). We refer to the pre-post difference of these angles as angular disparity. The sign of the difference depended on the offset condition. For visual offsets to the right, the sign became negative when the post-angle was greater. For visual offsets to the left, the opposite was true.

2.4.3 Positional Discrepancy

Palm drift and angular disparity quantify aftereffects and possible compensations of the multisensory conflict induced by the visual offset. To investigate whether the estimated position of one's own body or the external object was affected, we derived a direct measure of the difference between pre- and post-localization. In order to do so, we calculated the line of pointing and obtained the closest distance to the target reference (see Fig. 6, Panel C). Again, we calculated the centroid of the palms and the centroids of the index fingers and thumbs, respectively. Based on these points we obtained the following line equation in normal form (again ignoring the y-coordinate):

$$ax + bz + c = 0, \tag{1}$$

and computed the closest distance to the target reference in the following way:

$$\frac{|ax_0 + bz_0 + c|}{\sqrt{a^2 + b^2}}\ , \tag{2}$$

where x_0 and z_0 refer to the x- and z-coordinate of the reference, either self or external.

Again, we obtained the discrepancy value as a difference between the assessed closest distance during the pre- and post-localization. We thus did not compare the localization precision, but whether the data from the post-localization systematically differed from the pre-localization. Visual offsets of the hands to the right should yield compensation of the actual hand position to the left, for visual offsets to the left, the opposite is true. We signed the discrepancies accordingly. If the closest point in the post-localization was to the left of the closest point in the pre-localization, the sign was positive for visual offsets to the right. For visual offsets to the left, the sign was positive when the closest point in the post-localization was to the right of the closest point in the pre-localization. By applying this procedure, the discrepancies remain centered around 0 when there are no systematic changes in the localization. Furthermore, positive discrepancies always indicate a change in the localization in the direction of the compensation of the actual hand position. Thus, comparisons across the applied visual offsets become possible.

While the angular disparity provides information regarding a putative correction of the shifted palm position, positional discrepancy allows to quantify whether this compensation was complete. For instance, significant disparity without discrepancy would imply a complete compensation. If both measures show significant effects, this would imply a partial compensation. We expected the lowest discrepancy in the case of invisible hands and self-localization, because in this case the dominant available information is purely proprioceptive. Similarly, we expected increasingly higher discrepancies given visual information and an external, visually-presented target.

2.4.4 Data Preparation

As described above, the offset conditions were reversed with respect to the x-axis, to assure that left- and rightward shifts did not cancel out each other. To do so, we reversed the signs of the dependent measures for positive shifts on the x-axis, i.e. shifts to the right (see Fig. 5b). This preserved the sign of the respective measure and made it comparable across offset conditions. Hence, even with this adjustment, the dependent measures are centered around zero, and if there are no effects, the respective mean values should not differ from zero. An overview of the raw data in the different conditions is shown in Fig. 7.

Seeing that we analyzed the data with linear mixed effect models, no further aggregation was applied. Participants had to complete the two pre- and the two post-localizations per trial to proceed. Thus, there was no missing data.

3 Results

3.1 Data Analysis

We used R [38] and the *lme4* package [3] to perform a linear mixed effects analysis of the relationship between the respective dependent measures and the factors hand visibility (visible, or invisible), reference (self or external), and offset (positive x, negative x, positive x and negative z, and negative x and negative z). Significance of model components was estimated with the *lmerTest* package [28]. We compared models of increasing complexity with likelihood ratio tests to determine which factors were required to account for the data. In line with the approach proposed by Barr et al [1], we applied the maximal random effect structure if possible, even in the cases when the likelihood ratio tests suggested a simplified effect structure. Since all factors varied within participants and there were multiple observations per factor combination, we applied a random intercept and random slope per participant for each factor. Hence, the applied null model consisted of a global intercept and intercepts and slopes for each participant per factor level. If the null model did not converge, we used the model with the next more complex, converging random effect structure. After the identification of the null model, we added fixed effects for the experimental factors to the model as long as the likelihood ratio test between the simpler and the more complex model yielded significant results (with $\alpha = 0.05$). We always compared models differing with respect to only one factor. Models with a single fixed effect were compared with the null model, models with two fixed effects were compared with models with one fixed effect and so on.

For all measures, the four mean values of the visibility × reference interaction were tested against zero via one-sample t-tests. The significance level was set to $\alpha = 0.05$ in all cases. The four resulting p-values were adjusted for multiple comparisons by the method of Holm [20]. The raw data for this interaction is shown in Fig. 8. IPQ questionnaires were analyzed using the reference scores provided by the igroup consortium [2]. The respective detailed analyses can be found in Appendix A. To check for possible learning effects and effects of the different offset conditions, we also analyzed the number of

---

[2] Data is available at the igroup website: http://www.igroup.org/pq/ipq/data.php

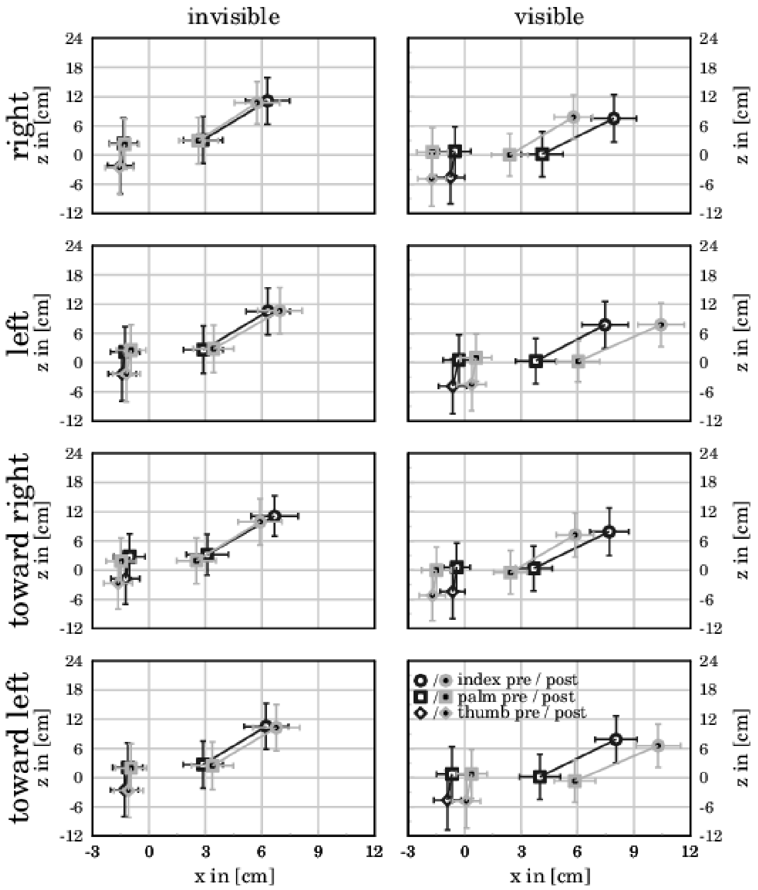


Fig. 7: Aggregated raw data for the different conditions. Only the horizontal (x-axis) and the depth axes (z-axis) are considered. The error bars indicate the standard error of the mean in the respective condition. The left column shows data obtained with invisible hands, the right column data with visible hands. Square markers indicate palm positions, diamonds represent thumb positions, and circles refer to index finger positions. Dark markers indicate data from the pre-localization, while light markers indicate data from the post-localization. Rows indicate the different *visual* offset conditions. The horizontal shift between pre- and post-localization data indicates the *opposed compensation* of the actual hand position. Each plot contains data from both localization tasks (self- and external localization). The respective palm and finger tip positions are connected with lines. Please note that only the centroids for finger and palm positions are displayed, e.g. the black diamond indicating the average thumb position in the pre-test is obtained in terms of the centroid of the two thumb tip positions.

petals picked during the interaction task. The respective analysis can be found in Appendix B.

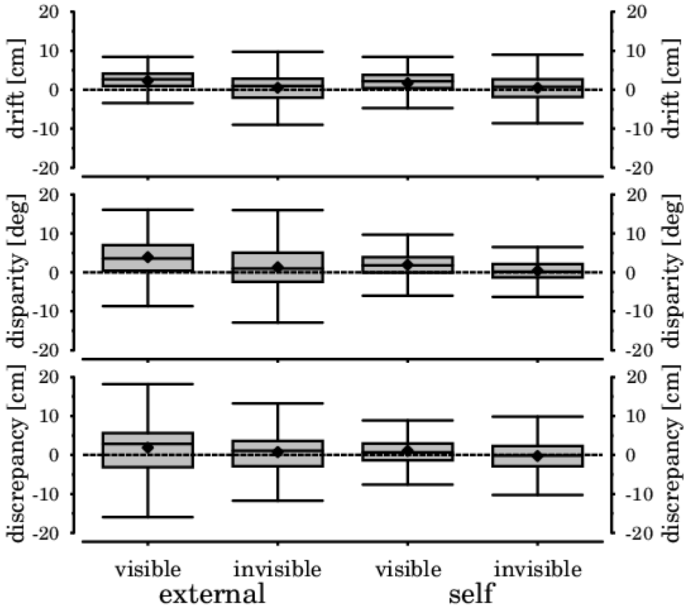


Fig. 8: Interaction between hand visibility and localization reference. Horizontal lines within the boxes indicate the median, the diamond marker represents the mean. Dashed lines indicate a value of zero. The according t-test results are shown in Tab. 5. The scale for angular disparity indicates angles in degrees. For the two other measures, the y-axis represents centimeters. Diamonds indicate the mean, while the horizontal line within the boxplots indicates the median. The lower hinge indicates the end of the first, the upper hinge the end of the third quartile. The notches cover ±1.5 of the interquartile range.

3.2 Palm Drift

The null model with the maximal random effect structure did not converge. A comparison of null models with a simpler random effect structure yielded best results for a random slope, random intercept model for the factors hand visibility and offset. Fixed effect models of increasing complexity were compared with the null model. The addition of hand visibility as fixed effect increased the fit significantly ($\chi^2(1)$=25.77, $p < 0.001$, $df = 16$, AIC = 8668). According to the parameter estimates, visible hands increased the drift by 1.5 cm $\pm$ 0.24 (standard error) compared to invisible hands. A complete description of the

Table 2: Description of the best fitting linear mixed effect model for the *palm drift*. Only a fixed effect for hand visibility improved the fit significantly. More complex models, assuming a fixed effect of reference or the respective interaction between visibility and reference failed the significance criterion ($\alpha = .05$). Significance of random effects was analyzed in terms of likelihood ratio tests. Fixed effects were analyzed with t-tests. The respective p-values were calculated based on the Satterthwaite approximation.

effect	variance	χ^2	estimate	std. error	t	*df*	p
random effects							
participants							
visibility		13.5				3	.004*
intercept	0.04						
slope	0.83						
offset		14.1				10	.169
intercept	0.29						
slope$_{\text{toward left}}$	0.18						
slope$_{\text{right}}$	1.04						
slope$_{\text{toward right}}$	1.22						
residual	13.08						
fixed effects							
intercept			0.48	0.14	3.14	33.13	.002*
visibility			1.51	0.24	6.25	33.18	$<$ .001*

model is given in Tab. 2. The addition of reference as a fixed factor did not improve the model fit significantly, even though a tendency was observable ($\chi^2(1)$=3.16, $p = .08$). According to the parameter estimates, the effect of reference was subtle. In the case of the external reference, the drift was increased by only .32 cm ± 0.18 (standard error). Given the non-significant main effect, a model assuming an interaction between reference and visibility did not increase the fit, either ($\chi^2(1)$=2.38, $p = .12$). Since the p-values showed a tendency toward significance, we performed paired t-tests for the interaction. The tests revealed that the drift was significantly higher for the external reference in case of visible hands ($t(32) = 2.78$, $p < .01$). For invisible hands, this difference did not reach significance ($t(32) = 0.25$, $p = .40$).

To assure that the drift differed reliably from zero, the four mean values of the visibility × reference interaction were tested against zero, the respective results are shown in Tab. 5. All mean scores differed significantly from zero.

3.3 Angular Disparity

The null model with the maximal random effect structure did converge and could be used as a baseline. The addition of visibility ($\chi^2(1)$=18.03, $p < 0.001$, *df* = 19,AIC = 10163) improved the model fit significantly. A further addition of reference as fixed effect further improved the model fit ($\chi^2(1)$=18.12, $p < 0.001$, *df* = 20, AIC = 10147). According to the parameter estimates, visible hands increased the disparity by 2.0° ± 0.41 (standard error) com-

Table 3: Description of the best fitting linear mixed effect model for the *angular disparity*. The best fit was obtained with a model assuming fixed effects for visibility and reference. A model assuming an interaction of both effects barely failed the significance criterion ($\alpha = .05$). Significance of random effects was analyzed in terms of likelihood ratio tests. Fixed effects were analyzed with t-tests. The respective p-values were calculated based on the Satterthwaite approximation.

effect	variance	χ^2	estimate	std. error	*t*	*df*	*p*
random effects							
participants							
visibility		11.6				3	.01*
intercept	0.67						
slope	2.97						
reference		1.67				3	.64
intercept	0.52						
slope	0.44						
offset		29.59				10	.001*
intercept	0.91						
slope$_{\text{toward left}}$	1.00						
slope$_{\text{right}}$	4.56						
slope$_{\text{toward right}}$	2.29						
residual	32.81						
fixed effects							
intercept			1.60	0.31	5.09	50.77	$< .001$*
visibility			2.04	0.42	4.90	33.01	$< .001$*
reference			-1.44	0.31	-4.66	74.34	$< .001$*

pared to invisible hands. If participants pointed to themselves, the disparity was reduced by $1.4° \pm 0.31$ (standard error) compared to pointing to the basket. A complete description of the model is given in Tab. 3. This model was nearly outperformed by a model assuming an interaction between reference and visibility ($\chi^2(1)$=3.66, $p = .06$). Given this tendency toward significance, we analyzed the interaction with paired t-tests. The tests revealed that the disparity was significantly higher for the external reference, both, for visible ($t(32)$ =5.01, $p < .001$) as well as for invisible hands ($t(32) = 2.90$, $p < .01$).

To assure that the disparity differed reliably from zero, the four mean values of the visibility × reference interaction were tested against zero, the respective results are shown in Tab. 5. All mean scores differed significantly from zero.

3.4 Positional Discrepancy

The null model with the maximal random effect structure did not converge. A comparison of null models with a simpler random effect structure yielded best results for a random slope, random intercept model for the factors reference and offset. The addition of hand visibility as fixed effect increased the fit significantly ($\chi^2(1)$=12.01, $p < .001$, $df = 16$, AIC = 10574). Adding reference

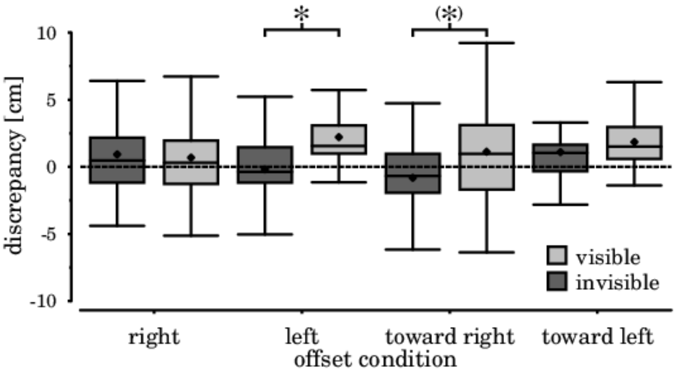


Fig. 9: Interaction between hand visibility and offset condition for the positional discrepancy data. The directions on the x-axis refer to the visual offset. The actual hands were shifted in the opposite direction to compensate for the offset. For visual shifts to the left and to the right, the discrepancy was significantly smaller in case of invisible hands. The data obtained with invisible hands in the respective conditions did not differ significantly from zero. Data obtained with visible hands did differ significantly from zero.

as an additional fixed factor significantly improved the fit compared to the model assuming hand visibility as the only fixed factor ($\chi^2(1)$=5.76, $p = .02$, $df = 17$, AIC = 10571). Adding offset as a third fixed did not improve the fit ($\chi^2(3)$=5.35, $p = .15$, $df = 20$, AIC = 10571) compared to the two-factor model. Similarly, assuming an interaction between visibility and reference did not improve the fit ($\chi^2(1)$=0.07, $p = .79$, $df = 20$, AIC = 10571). A further analysis of more complex models showed that only a model assuming an interaction between visibility and offset, as well as a fixed factor for reference, fitted the data better than the two-factor model ($\chi^2(6)$=15.05, $p = .02$, $df = 23$, AIC = 10568). A detailed description of the model is shown in Tab. 4. Regarding the parameter estimates, the discrepancy was increased by 2.4 cm ± 0.66 (standard error) in case of visible hands compared to invisible hands. The discrepancy decreased by 0.9 cm ± 0.42 (standard error) when participants pointed toward themselves compared to when they pointed toward the basket.

Given the non-significant model improvement in case of adding a fixed factor for offset alone, the numeric estimates for the fixed effect of offset alone are difficult to interpret. A further analysis of the interaction between visibility and offset is necessary. We performed post-hoc t-tests, which revealed significant and near-significant differences between visible and invisible hands in two offset conditions (p-values were adjusted for multiple comparisons by the method of Holm). For visual offsets to the left, discrepancies were significantly reduced in case of invisible hands compared to visible hands ($t(32) =$ -3.72, $p < .01$; see Fig. 9). For visual offsets toward and to the right of the participants, a similar nearly significant pattern was observed ($t(32)$ = -2.50, p = .05). No significant differences were observed for visual offsets to the right ($t(32)$ = -0.33, p = .63) and for visual offsets toward and to the left of the participants ($t(32)$ = -1.52, p = .14). These results dovetail with the model estimates for the visibility × offset interaction shown in Tab. 4. For the offset conditions, the offset to the left (leftmost boxplots in Fig. 9), which is the condition with the largest difference regarding visibility, served as baseline. Significant and near significant slopes were observed for an offset to the right of the participant, and an offset toward and to the left of the participant, which are the conditions with the smallest differences regarding visibility.

To assure that the discrepancy differed reliably from zero, the four mean values of the visibility × reference interaction were tested against zero. The respective results are shown in Tab. 5. In case of visible hands, the tests yielded insignificant results. The discrepancy for external localization with invisible hands reached only marginal significance (p = .05). For the self-localization in case of invisible hands, the discrepancy did not differ significantly from zero.

3.5 Summary

The analyses revealed a considerable effect of hand visibility for all dependent measures. With respect to the effect of reference the results are slightly mixed. While the variability in the drift can be accounted for by visibility alone, effects of reference were visible for the disparity as well as the discrepancy. In general, effects of the manipulation were more pronounced for visible compared to invisible hands. Moreover, the effect of the visual conflict was stronger during the external localization than during the self-localization. The means in the different visibility and reference conditions differed from zero in nearly all measures, except for the positional discrepancy in the case of self-localization with invisible hands. The variation of the offset factor had no effect in the case of angular disparity and palm drift. Only to account for the positional discrepancy data, the offset factor had to be included in the model. Immersion was evaluated positively. The data of the immersion questionnaire is presented in Appendix A. In general, participants complied with the picking task and their performance increased over blocks. The respective analysis is presented in Appendix B.

4 General Discussion

Our aim was to investigate possible effects of multisensory conflict regarding the position of the hands on spatial representations in different frames of reference. To do so, we measured participants' performance when localizing their own body and an external object before and after a virtual dissociation of the proprioceptive and visual hand position information. The dissociation was accomplished by means of an immersive VR setup during a bimanual

Table 4: Description of the best fitting linear mixed effect model for the *positional discrepancy*. The best fit was obtained with a model assuming fixed effects for visibility, offset, and reference, as well as an interaction between visibility and offset. The fixed effect of offset alone did not improve the model fit significantly, hence the respective estimates should be treated with caution. Significance of random effects was analyzed in terms of likelihood ratio tests, fixed effects were analyzed with t-tests, the respective p-values were calculated based on the Satterthwaite approximation.

effect	variance	χ^2	estimate	std. error	*t*	*df*	*p*
random effects							
participants							
reference		9.85				3	.02*
intercept	2.25						
slope	2.14						
offset		5.91				10	.82
intercept	0.37						
slope$_{\text{toward left}}$	0.26						
slope$_{\text{right}}$	1.87						
slope$_{\text{toward right}}$	1.65						
residual	43.45						
fixed effects							
intercept			0.26	0.57	0.46	172.4	.645
visibility			2.41	0.66	3.63	1486.4	$<$ .001*
reference			-0.91	0.42	-2.17	39.0	.036*
offset$_{\text{toward left}}$			1.29	0.67	1.93	705.3	.054
offset$_{\text{right}}$			1.12	0.70	1.59	124.6	.115
offset$_{\text{toward right}}$			-0.63	0.70	-0.90	186.7	.370
visibility × offset$_{\text{toward left}}$			-1.65	0.94	-1.76	1486.4	.078
visibility × offset$_{\text{right}}$			-2.64	0.94	-2.82	1486.4	.005*
visibility × offset$_{\text{toward right}}$			-0.45	0.94	-0.49	1486.4	.625

interaction task. To manipulate the saliency of visual and proprioceptive information, we had participants perform the localization task with either visible or invisible virtual hands. We expected stronger effects when participants saw their shifted hands, because visual dominance should continue to influence the veridical proprioceptive information. Furthermore, we expected stronger effects in the external localization task, because the spatial representation of external objects relies more on vision than on proprioception. The hypotheses were largely confirmed by means of three different dependent measures, which reflect different aspects of the adaptation to the induced multisensory conflict between vision and proprioception.

Results regarding the palm drift showed that the compensation of the visual offset persisted in the localization task, especially when the hands were visible. The drift was only slightly modulated by the target of the localization task, that is, the magnitude of the drift remained nearly the same when par-

Table 5: T-Tests against zero for the different visibility and reference conditions. The four comparisons with respect to one dependent measure were adjusted for multiple comparisons according to the procedure proposed by Holm.

visibility	reference	measure	*M*	*SEM*	df	t	p
visible	self	drift	1.69	0.19	32	8.43	< .001*
		disparity	1.92	0.20	32	7.46	< .001*
		discrepancy	1.06	0.32	32	3.83	.002*
	external	drift	2.29	0.18	32	9.73	< .001*
		disparity	3.91	0.33	32	9.78	< .001*
		discrepancy	1.88	0.38	32	3.30	.007*
invisible	self	drift	0.46	0.18	32	2.94	.012*
		disparity	0.43	0.20	32	2.20	.035*
		discrepancy	-0.24	0.36	32	-0.69	.493
	external	drift	0.51	0.20	32	2.95	.012*
		disparity	1.32	0.41	32	3.64	.002*
		discrepancy	0.75	0.30	32	2.34	$.051^{(*)}$

ticipants pointed towards themselves or the external reference. The other two measures – angular disparity and positional discrepancy – were clearly affected by the target location. When pointing to the external reference, participants showed a systematic change in the localization. The discrepancy almost vanished when participants pointed to the external reference with invisible hands. For self-localizations, participants could rely on a purely postural frame of reference, based on veridical proprioceptive information. When the hands were invisible during self-localization this seemed to be the case, seeing that participants showed a significant compensation (angular disparity) while there was no significant positional discrepancy. In case of visible hands, however, the positional discrepancy for the self-localization remained significant. Apparently, the presence of the manipulated visual information was able to bias the positional estimate of a frame of reference, that is, body posture, which is grounded in proprioceptive information.

4.1 Aftereffects in the Localization Task: Palm Drift

To succeed in the bimanual task, participants had to compensate for the visual offset by an offset of their actual hands in the opposite direction. For the localization tasks, this compensation was completely irrelevant. However, the analysis of the palm drift data indicates that the compensation persisted during the post-localization. That is, participants centered their hands during the post-localization in a way that accounted for the visual offset. Notably, this also happened when the hands were invisible during the post-localization. Moreover, the positional adaptation persisted for both the localization of one's own body as well as the localization of the external object.

This aftereffect implies that participants shifted their whole hand centered frame of reference. Since this effect was less pronounced for invisible hands (0.5 cm compared to 2.0 cm according to the model estimates) and the largest observed drift was about one third of the actual visual offset (6.7 cm in the horizontal plane), the underlying process seems to rely on a multisensory integration of visual and proprioceptive information, yielding a stronger drift if the manipulated visual information is present, that is, if the hands remained visible. Even if this drift implies an adaptation of the hand centered frame of reference – similar to classic effects like the rubber hand illusion – it does not reveal if and how other spatial representations are affected.

4.2 Partial Compensation of the Drift: Angular Disparity

While the estimated center of hand space was shifted to compensate for the visual offset, especially when the hands were visible, this drift can be compensated for by means of a palm rotation adaptation. The absence of adaptation effects would imply that (a) other spatial frames of reference, which are used to represent the position of the own body and the position of external objects, were shifted in the same way as the center of hand space and (b) proprioceptive information was not used to correct the drift in the visual modality. On the other hand, if veridical proprioceptive information is used to compensate for the drift, the respective effect should be stronger in the case of stronger discrepancies between vision and proprioception, that is, in conditions that also showed a strong palm drift.

The results show that the compensation depended on the availability of visual information and the localization task. Indeed the observed pattern of results matched the one obtained for the palm drift. Participants showed stronger adaptations in case of visible hands and external localizations. Hence, participants corrected the palm drift by means of palm rotation adaptations at least partially. To further evaluate the effect of this rotation adaptation, the positional discrepancy measure was used.

4.3 Shifts in Spatial Representations: Positional Discrepancy

To perform a localization in terms of pointing to a sensory target, it is necessary to transform the positional estimate from the sensory frame of reference into the frame of reference of the motor effectors. There is some evidence suggesting a common frame of reference for movement planning in terms of an eye-centered frame of reference [9, 36]. This common frame of reference appears to integrate positional estimates from different sensory and motor sources, and fosters exchange of information across frames of reference. The results regarding the palm drift show that the estimated center of hand space was shifted and the disparity data show that this drift was partially compensated. Since only the visual hand position was manipulated, the visual information regarding the external reference remained veridical, as well as the proprioceptive information encoding the body center.

We investigated shifts in the estimated locations of the target references by calculating the signed distance between the pre- and the post-estimates. Again the data showed a systematic change, depending on the hand visibility and the reference. In accordance with the palm drift data, the discrepancy was higher in case of visible hands. However, also the discrepancy between the external and self-reference differed significantly. Moreover, in the case of invisible hands and self-localization, no significant discrepancy could be observed, that is, the compensation via the palm rotation was complete. The significant discrepancy in the case of visible hands and self-reference shows how visual information can overwrite proprioceptive information, leading to a mislocalization of the body center. In line with results reported by Longo and Lourenco [33], the discrepancy was stronger for the more visually grounded representation of an external object, especially when the hands were visible.

In contrast to the other two measures, the type of visual offset had a significant influence and interacted with hand visibility. Visual offsets to the left and toward and to the right of the participants yielded significantly less discrepancy in the localization task when the hands were invisible compared to visible hands. This effect was not expected and is difficult to interpret. Seeing that there was a tendency that also visual shifts toward and to the left of the participants yielded less discrepancy ($p = .07$, unadjusted), it seems that only visual shifts to the right yielded the same amount of discrepancy for both visible and invisible hands. Visual shifts to the right result in a leftward shift of the actual hands, that is, in the direction of the spatial bias due to pseudo-neglect. It is tempting to assume that the compensation of the visual shift and the pseudo-neglect bias added up and resulted in the comparable discrepancy for both visible and invisible hands, but further data is necessary to verify this speculation.

4.4 Conclusion

Our results show how multisensory conflict can lead to an adaptation of the spatial representation of an external object and, to a lesser degree, to an adaption of self-localization. The data is in line with previous results that showed a different weighting of proprioceptive and visual information depending on the distance from the body center [32, 33]. Moreover, the data highlights the adaptive nature of spatial representations.

Earlier studies have shown a fast remapping of peripersonal space in the case of tool-usage [14, 24], and how this remapping affects body perception [7]. Our results extend these findings by showing remappings of self-localizations and external localizations due to adaptations in hand space given false visual hand location information. The results dovetail with results implying an adaptation of the whole body model due to changes in hand space induced by the rubber hand illusion [6]. Essentially, the results suggest a close coupling of spatial representations across different frames of reference and the crucial role of a postural body schema, which interlinks these frames of reference.

Multisensory integration of different senses seems to occur in a statistically optimal fashion [12, 13]. If senses are in disagreement, i.e. if there is multisensory conflict – induced, for example, by the rubber hand illusion or by our subliminally drifting virtual hand representation – hand location estimates will not be accurate. Seeing that participants did not always completely compensate for their shifted hand positions by according palm rotations, the induced conflict does not only influence the frame of reference in which it occurs. The remaining mislocalizations imply that other positional estimates are adapted to a certain degree as well, in order to create a consistent embedding of all sensory information available. As a result, participants always compensated for the visually-induced relocation of the hands by an adapted pointing direction. In the case of visible hands, this adaptation was never complete, implying that both, the false visual information and the veridical postural information played a significant role. Also in the case of invisible hands and the external target the adaptation was not complete, such that participants' pointing direction still missed the basket's location, suggesting a remaining influence of the visual shift. Only in the case of invisible hands and self-localization, the localization error did not differ significantly from the prior localization error, implying that in this case the focus lied nearly fully on the veridical proprioceptive perception of hand and body.

The idea of a common, multisensory-integrating frame of reference used in movement control is generally in agreement with the gathered data [9]. From a computational point of view, however, the effects can be accounted for by models that strive for the maintenance of overall consistent bodily and spatial representations across multiple different frames of reference and sensory modalities. Such representations essentially also allow task-dependent information reweighting and the computation of online validity estimations of the available sensory information [10, 11, 36].

Our setup combined the advantages of different paradigms to induce multisensory conflict, including the rubber hand illusion and prism adaptation designs. Similar to the rubber hand illusion, conflict was restricted to a single frame of reference (hand space). In accordance with prism adaptation studies, participants could perform interactions while the multisensory conflict was still active. However, our setup was limited by the tracking range of the Leap sensor. Hence, we could only apply rather subtle visual offsets. It remains an open question to which degree hand positions can be shifted until participants become aware of the manipulation. To test whether participants can reliably distinguish between different offset conditions, a behavioral manipulation check would be necessary.

In general, our study corroborates evidence that immersive VR setups are well-suited to investigate multisensory conflict and its effects on spatial representations. The application of motion capture systems with a larger tracking range will allow to investigate the effects of larger visual offsets. These investigations will shed further light on how exactly spatial representations are remapped and whether classic spatial compatibility effects, like the Simon effect or the SNARC effect [45], can be affected by remapped spatial representations. We expect that such investigations will provide an even deeper understanding of how spatial representations are rooted in our sensorimotor system, how they are acquired during sensorimotor interactions, and how they interact with other spatial and cognitive encodings.

References

1. Barr DJ, Levy R, Scheepers C, Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. J Mem Lang 68(3):255–278
2. Barsalou LW (2008) Grounded cognition. Annu Rev Psychol 59:617–645
3. Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67(1):1–48
4. Botvinick M, Cohen J (1998) Rubber hands 'feel' touch that eyes see. Nature 391(6669):756–756
5. Butz MV, Herbort O, Hoffmann J (2007) Exploiting redundancy for flexible behavior: Unsupervised learning in a modular sensorimotor control architecture. Psychol Rev 114:1015–1046
6. Butz MV, Kutter EF, Lorenz C (2014) Rubber hand illusion affects joint angle perception. PloS One 9(3):e92854
7. Canzoneri E, Ubaldi S, Rastelli V, Finisguerra A, Bassolino M, Serino A (2013) Tool-use reshapes the boundaries of body and peripersonal space representations. Exp Brain Res 228(1):25–42
8. Coello Y, Bartolo A, Amiri B, Devanne H, Houdayer E, Derambure P (2008) Perceiving what is reachable depends on motor representations: evidence from a transcranial magnetic stimulation study. PLoS One 3(8):e2862
9. Cohen YE, Andersen RA (2002) A common reference frame for movement plans in the posterior parietal cortex. Nat Rev Neurosci 3(7):553–562
10. Ehrenfeld S, Butz MV (2013) The modular modality frame model: continuous body state estimation and plausibility-weighted information fusion. Biol Cybern 107(1):61–82
11. Ehrenfeld S, Herbort O, Butz MV (2013) Modular neuron-based body estimation: maintaining consistency over different limbs, modalities, and frames of reference. Front Comput Neurosci 7(Article UNSP 148)
12. Ernst MO, Banks MS (2002) Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415(6870):429–433
13. Ernst MO, Bülthoff HH (2004) Merging the senses into a robust percept. Trends Cogn Sci 8(4):162–169
14. Farnè A, Làdavas E (2000) Dynamic size-change of hand peripersonal space following tool use. Neuroreport 11(8):1645–1649

15. Fogassi L, Gallese V, Fadiga L, Luppino G, Matelli M, Rizzolatti G (1996) Coding of peripersonal space in inferior premotor cortex (area f4). J Neurophysiol 76(1):141–157
16. Gamberini L, Seraglia B, Priftis K (2008) Processing of peripersonal and extrapersonal space using tools: Evidence from visual line bisection in real and virtual environments. Neuropsychologia 46(5):1298–1304
17. Gentilucci M, Jeannerod M, Tadary B, Decety J (1994) Dissociating visual and kinesthetic coordinates during pointing movements. Exp Brain Res 102(2):359–366
18. Glenberg AM, Witt JK, Metcalfe J (2013) From the revolution to embodiment 25 years of cognitive psychology. Perspect Psychol Sci 8(5):573–585
19. Guna J, Jakus G, Pogačnik M, Tomažič S, Sodnik J (2014) An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking. Sensors 14(2):3702–3720
20. Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat pp 65–70
21. Holmes NP, Spence C (2004) The body schema and multisensory representation(s) of peripersonal space. Cogn Process 5(2):94–105
22. Holmes NP, Spence C (2005) Visual bias of unseen hand position with a mirror: spatial and temporal factors. Exp Brain Res 166(3-4):489–497
23. Holmes NP, Snijders HJ, Spence C (2006) Reaching with alien limbs: Visual exposure to prosthetic hands in a mirror biases proprioception without accompanying illusions of ownership. Percept Psychophys 68(4):685–701
24. Iriki A, Tanaka M, Iwamura Y (1996) Coding of modified body schema during tool use by macaque postcentral neurones. Neuroreport 7(14):2325–2330
25. Jewell G, McCourt ME (2000) Pseudoneglect: a review and meta-analysis of performance factors in line bisection tasks. Neuropsychologia 38(1):93–110
26. Kirsch W, Kunde W (2013) Visual near space is scaled to parameters of current action plans. J Exp Psychol Hum Percept Perform 39(5):1313–1325
27. Kirsch W, Herbort O, Butz MV, Kunde W (2012) Influence of motor planning on distance perception within the peripersonal space. PLoS One 7(4):e34880
28. Kuznetsova A, Bruun Brockhoff P, Haubo Bojesen Christensen R (2016) lmerTest: Tests in Linear Mixed Effects Models. URL `https://CRAN.R-project.org/package=lmerTest`, r package version 2.0-32
29. Lawrence MA (2015) ez: Easy Analysis and Visualization of Factorial Experiments. URL `https://CRAN.R-project.org/package=ez`, r package version 4.3
30. Linkenauger SA, Witt JK, Proffitt DR (2011) Taking a hands-on approach: apparent grasping ability scales the perception of object size. J Exp Psychol Hum Percept Perform 37(5):1432–1441

31. Linkenauger SA, Bülthoff HH, Mohler BJ (2015) Virtual arm's reach influences perceived distances but only after experience reaching. Neuropsychologia 70:393–401
32. Longo MR, Lourenco SF (2006) On the nature of near space: Effects of tool use and the transition to far space. Neuropsychologia 44(6):977–981
33. Longo MR, Lourenco SF (2007) Space perception and body morphology: extent of near space scales with arm length. Exp Brain Res 177(2):285–290
34. Lourenco SF, Longo MR (2009) The plasticity of near space: Evidence for contraction. Cognition 112(3):451–456
35. Macaluso E, Maravita A (2010) The representation of space near the body through touch and vision. Neuropsychologia 48(3):782–795
36. McGuire LM, Sabes PN (2009) Sensory transformations and the use of multiple reference frames for reach planning. Nat Neurosci 12(8):1056–1061
37. Mohler BJ, Bülthoff HH, Thompson WB, Creem-Regehr SH (2008) A full-body avatar improves egocentric distance judgments in an immersive virtual environment. In: Proceedings of the 5th symposium on Applied perception in graphics and visualization, ACM, p 194
38. R Core Team (2016) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, URL `https://www.R-project.org/`
39. Regenbrecht H, Schubert T (2002) Real and illusory interactions enhance presence in virtual environments. Presence-Teleop Virt 11(4):425–434
40. Rossetti Y, Desmurget M, Prablanc C (1995) Vectorial coding of movement: vision, proprioception, or both? J Neurophysiol 74(1):457–463
41. Schroeder PA, Lohmann J, Butz MV, Christian P (2016) Behavioral bias for food reflected in hand movements: A preliminary study with healthy subjects. Cyberpsychology, Behavior, and Social Networking 19:120–126, DOI 10.1089/cyber.2015.0311
42. Schubert T, Friedmann F, Regenbrecht H (2001) The experience of presence: Factor analytic insights. Presence 10(3):266–281
43. Simons DJ, Chabris CF (1999) Gorillas in our midst: Sustained inattentional blindness for dynamic events. Perception 28(9):1059–1074
44. Simons DJ, Levin DT (1997) Change blindness. Trends Cogn Sci 1(7):261–267
45. Wood G, Willmes K, Nuerk HC, Fischer MH (2008) On the cognitive link between space and number: A meta-analysis of the SNARC effect. Psychol Sci Q 50(4):489–525

Appendix A IPQ Evaluation

The IPQ assesses presence in virtual realities on three different scales and allows to quantify the degree of immersion experienced by the participants within the VR. The igroup consortium provides reference data from different VR setups. We compared our data to setups that also used a head-mounted display. The reference data set comprised 24 mean values for the three scales.

Table 6: T-Test results for the different *IPQ* scales. Observed values were tested whether they exceed the reference scores. P-values and degrees of freedom were adjusted, since observed and reference scores stem from different samples with different sizes.

scale	df	t	p
spatial presence	33.4	1.76	.044*
involvement	43.0	0.04	.48
realism	42.74	0.08	.53

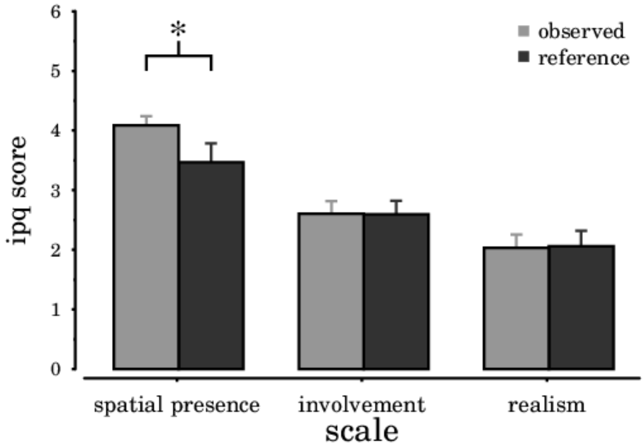


Fig. 10: Scores for the different IPQ scales in the observed (light gray) and the reference (dark gray) data. Significant differences were found for the spatial presence scale. The scores show that the experimental setup yielded comparable immersion and realism judgments as reference setups, while spatial presence was improved.

Due to a software issue, only 21 of the 33 participants completed the IPQ questionnaire in our study. We checked whether the results exceeded those of the reference data. The results of the respective t-tests are shown in Tab. 6, the data is shown in Fig. 10.

With respect to involvement and realism, the data is comparable to the reference data. In case of spatial presence, the results are significantly improved compared to the reference data ($t(33.4) = 1.76$, $p < .05$). Together, the results show a sufficient degree of immersion. Improvements with respect to spatial presence dovetail with other results that showed enhanced spatial perception in VR when participants were equipped with a body model [37], or when they could interact with the VR via bodily motion [41].

Appendix B Collected Petals

To check whether the participants complied with the manual task, the amount of petals picked was subjected to a separate analysis. In general participants complied with the task, collecting 4.5 petals on average per trial. However, there were considerable individual differences, leading to a rather high standard deviation of 1.4 petals. To further check for learning

Table 7: ANOVA table for the number of collected petals.

factor	df	F	p	η_p^2
block	1	29.15	$< .001^*$	.48
offset	1	23.03	$< .001^*$	.42
block × offset	1	0.94	.34	.03

effects and effects due to the induced drift, the number of collected petals was analyzed with a 2 × 2 repeated measure ANOVA using R [38] and the *ez* package [29]. We considered the factors block and offset condition. The experiment was divided into two blocks, the according factor had two levels. To allow a straightforward analysis of the different offset conditions, we aggregated over the variations in the depth axis (see Fig. 5b), such that the resulting factor had only two levels (visual offsets to the left or to the right).

The results of the analysis are shown in Tab. 7. There was a considerable learning effect. In the second block, participants collected significantly more petals than in the first block ($M = 3.9$ versus $M = 5.1$). Furthermore, participants collected more petals when the hands were visually shifted to the right ($M = 4.2$), than when they were shifted to the left ($M = 4.8$). This unwanted effect is most likely due to the asymmetric layout of the scene (see Fig. 3). Visual offsets to the right were compensated by moving the hands to the left. This allows a more convenient trajectory through the task space, because the hands operate in the center of the tracking range with the flower physically slightly to the left and the basket to the right of the center of the tracking range. For visual offsets to the left, the trajectory through the task space is less convenient. In this case, the shifts are compensated by placing the hands to the right of the center of the tracking range, such that the hands had to be moved even further to the right in order to reach the basket.

D. Mental Space Maps Into the Future

The manuscript has been submitted to *Cognition* and is currently under review. To avoid any issues regarding prior publication, only the abstract is included here. Currently, the manuscript can be referenced as:

Belardinelli, A., Lohmann, J., Farnè, A., & Butz, M. V.: **Mental Space Maps Into the Future**. *under review*.

Mental Space Maps into the Future

Authors: Anna Belardinelli[1], Johannes Lohmann[1*], Alessandro Farnè[2,3], Martin V. Butz[1]

Affiliations:

[1]Cognitive Modeling, Department of Computer Science and Department of Psychology, Faculty of Science, University of Tübingen, Germany.

[2] Integrative Multisensory Perception Action & Cognition Team - ImpAct, Lyon Neuroscience Research Center, INSERM U1028, CNRS U5292, Lyon, France

[3] Hospices Civils de Lyon, Neuro-immersion, Lyon, France

*Correspondence to: johannes.lohmann@uni-tuebingen.de

Abstract: It has been suggested that our mind anticipates the future to act in a goal-directed, event-oriented manner. Here we asked whether peripersonal hand space, that is, the space surrounding ones' hands, is mapped into the future while planning and executing a goal-directed object manipulation. We thus combined the crossmodal congruency paradigm (CCP), which has been used to study selective interactions between vision and touch within peripersonal space, with an object manipulation task. We expected crossmodal interactions in anticipation of the upcoming, currently planned manual object grasp. Our results confirm that visual distractors close to the future finger positions selectively influence vibrotactile perceptions. Moreover, vibrotactile stimulation influences gaze behavior in the light of the anticipated grasp. Both influences become apparent partially even before the hand starts to move. The results thus support theories of event encodings and anticipatory behavior, showing that peripersonal hand space is mapped onto the anticipated grasp.

Keywords: event segmentation theory; theory of event coding; anticipatory behavioral control; peripersonal space; cross-modal congruency paradigm

E. In touch with mental rotation: Interactions between mental and tactile rotations and motor responses

The manuscript has been published in *Experimental Brain Research*. The final publication is available at Springer via http://dx.doi.org/10.1007/s00221-016-4861-8, and can be retrieved by the following reference:

Lohmann, J., Rolke, B., & Butz, M. V. (2017): **In touch with mental rotation: Interactions between mental and tactile rotations and motor responses**. *Experimental Brain Research*, 235(4), 1063-1079. doi: 10.1007/s00221-016-4861-8.

In Touch with Mental Rotation: Interactions between Mental and Tactile Rotations and Motor Responses

Johannes Lohmann, Bettina Rolke, Martin V. Butz
Cognitive Modeling and Evolutionary Cognition, Department of
Computer Science and Department of Psychology, Faculty of
Science, University of Tübingen, Germany

Abstract

Although several process models have described the cognitive processing stages that are involved in mentally rotating objects, the exact nature of the rotation process itself remains elusive. According to embodied cognition, cognitive functions are deeply grounded in the sensorimotor system. We thus hypothesized that modal rotation perceptions should influence mental rotations. We conducted two studies in which participants had to judge if a rotated letter was visually presented canonically or mirrored. Concurrently, participants had to judge if a tactile rotation on their palm changed direction during the trial. The results show that tactile rotations can systematically influence mental rotation performance in that same rotations are favored. In addition, the results show that mental rotations produce a response compatibility effect, such that clockwise mental rotations facilitate responses to the right while counterclockwise mental rotations facilitate responses to the left. We conclude that the execution of mental rotations activates cognitive mechanisms that are also used to perceive rotations in different modalities and that are associated with directional motor control processes.

Keywords: Mental Rotation; Tactile Stimulation; Dual-Task Interference; Multimodal Interactions; Response Compatibility Effect

In Touch with Mental Rotation: Interactions between Mental and Tactile Rotations and Motor Responses

Introduction

Cognitive development is inevitably based on a combination of sensorimotor experiences and inborn developmental biases. Accordingly, theories of embodied cognition have emphasized that sensorimotor interactions shape the way in which we perceive and interact with our world (Barsalou, 1999; Butz, 2016; Hoffmann, 1993; O'Regan & Noë, 2001; Zacks & Tversky, 2001). Common, sensorimotor encodings have been proposed, which encode particular motor activities jointly with their sensory consequences (Hommel, Müsseler, Aschersleben, & Prinz, 2001; Prinz, 1990). As a result, decision making and action execution were shown to be influenced by anticipated action effects, which were implicitly or explicitly associated with the actions (Elsner & Hommel, 2001; Kunde, 2001). If common coding is indeed a general principle of cognitive development, sensorimotor encodings should be involved in many cognitive tasks. Mental rotation – the ability to mentally rotate an observed or imagined object – is one task for which the involvement of the motor system has already been shown. In the present work, we show that mental rotation performance can be affected by concurrent tactile stimulations. Moreover, we show that the mental rotation direction primes a directional motor response. These insights are in line with the assumption of one common mental rotation code, which is associated with according sensorimotor encodings but also with perceptual (in our case tactile) encodings.

Mental rotation models

In their seminal study, Shepard and Metzler (1971) let their participants compare the parity of two three-dimensional block figures, which were presented in different spatial orientations. Response times increased linearly with the angular disparity between the two figures. This disparity effect, however, is not restricted to mental rotations of three-dimensional stimuli. For instance, the same relation between response times and disparity was obtained when participants had to judge the parity of rotated alphanumeric characters (Cooper & Shepard, 1973; Heil & Rolke, 2002). The results imply that participants simulated an actual rotation to align the two figures or to rotate the character into an upright orientation.

Recent work has provided evidence for a strong overlap between mental and physical rotations – at least when the same task can be accomplished either with the help of physical rotations or purely mentally (Gardony, Taylor, & Brunyé, 2014; Wohlschläger & Wohlschläger, 1998). For example, results obtained with dual-task interference paradigms showed that motor activity selectively influences and interferes with mental rotation (Wexler, Kosslyn, & Berthoz, 1998). Besides this behavioral data, neuroimaging studies have shown that the same parietal brain areas that encode spatial real world transformations and partially also the same motor areas that produce according motor activities are engaged when participants have to perform mental rotations (see Zacks, 2008, for an overview). All in all, observations from behavioral studies as well as neuroimaging studies suggest a functional relationship between mental transformations and transformations in the real world.

Different process models of mental rotation have been proposed, which emphasize the role of spatial transformation processes in mental rotation. Just and Carpenter (1976) proposed that three stages are necessary to accomplish mental rotations in the classic Shepard-Metzler task. First, a search process detects possible corresponding parts of the two block figures. Second, a transformation and comparison process rotates and then compares the candidate segments. Finally, an additional comparison process is needed to decide if the applied rotation aligns the other parts of the figures as well. The underlying representation of the segments was assumed to be spatial, containing information like length and the absolute orientation of the perceived main axis of the target object. Compared to other models, Just and Carpenter assumed that the transformation process operates piecewise on the chosen, most informative features. The gathered data suggested furthermore that the transformation process operated in an iterative manner, mentally rotating the considered features by about 50° per step. Moreover, seeing that the typical disparity effect was also identified for tactile stimuli and in blind participants, which cannot use visual representations (Carpenter & Eisenberg, 1978), Just and Carpenter assumed the transformation process to be independent of the modality of the target stimulus. To sum up, Just and Carpenter proposed the transformation process (a) to operate in a piecewise fashion, (b) to rely on spatial instead of visual features, and (c) thus to be amodal rather than based on a specific input modality.

These assumptions are still debated and other process models stress the central role of visual imagery. Gill, O'Boyle, and Hathaway (1998) proposed a process model of mental rotation based on EEG data. Their model comprises four stages. First, the stimulus is encoded. Second, an internal mental image is generated, causing activity in the left parietal cortex. Third, this image is rotated and compared to the original, which results in activity of the left temporal cortex. Fourth, participants decide whether the rotated image matches the target stimulus, yielding activation of the right frontal cortex. This proposed cortical mapping dovetails with results from neuroimaging studies (Cohen et al., 1996). Compared to Just and Carpenter, Gill et al. emphasize the role of visual imagery and propose the generation of an internal visuo-spatial mental image, which is stored in a visual buffer and subjected to a rotation process, driven by the left temporal cortex. While the model accounts for the involvement of motor areas, the actual overlap between visual and motor rotation is not discussed in detail.

Even if both models differ with respect to the role of visual imagery, both models consider the spatial component in the actual transformation process. In a similar vein, Zacks and Michelon (2005) proposed a process model that focuses on spatial reasoning to account for mental rotation. The proposed multiple systems framework assumes that mental imagery and perception rely on the same cortical system. One central issue is the nature of the spatial representations formed in perception and mental imagery and how these spatial representations are transformed. Zacks and Michelon stress the role of reference frame transformations during mental rotation. According to their model, mental rotation involves the rotation of an object-centered frame of reference relative to an egocentric frame of reference in a continuous fashion. Compared to the model of

Just and Carpenter, Zacks and Michelon assume specific resources for different frames of reference transformations, as well as a general spatial processing resource. Furthermore, Zacks and Michelon stress the non-Euclidean nature of the geometry underlying some of the transformation processes. Especially with respect to effector-based transformations, that is, the transformation of effector-grounded frames of reference, the executed mental trajectory is not necessarily the shortest path through Euclidean space, but rather the most convenient with respect to motor execution.

Open questions

As mentioned above, the models mainly agree on the role of spatial transformations in mental rotation: According to Just and Carpenter (1976) an abstract, symbolic process is utilized that operates on spatial features in a piecewise fashion; Gill et al. (1998) argue in favor of the generation and rotation of a holistic mental image; whereas Zacks and Michelon (2005) highlight the role of spatial reference transformations, involving visual, spatial and – under certain circumstances – motor codes. In our opinion, however, two central questions remain unanswered, even though these might be closely related to one another. First, it remains open which specific process accomplishes mental rotation, i.e. *how* mental rotation is performed. Second, it is still debated on which representational codes mental rotation is based, i.e. *what* is rotated.

With respect to the *how* question, embodied cognition provides an idea that is consistent with the partial overlap of mental and physical rotations. From an embodied perspective, actual rotation experiences can be expected to develop into an internal rotation model. With practice and more experience, this knowledge will form a general rotation model, which can be used without overt activation of the respective motor codes and without actual sensory stimulation, but stays connected with its sensorimotor roots. In support of this idea, experimental studies have shown that the visual perception of ambiguous visual motion stimuli can be biased by current directional motor intentions and actual motor behavior (Wohlschläger, 2000). Moreover, even tactile rotation stimuli have been shown to be able to influence visual motion perception – at least when the tactile stimulus needs to be attended to (Butz, Thomaschke, Linhardt, & Herbort, 2010). Interestingly, in the latter study it was also shown that the mapping between tactile and visual motion depended on the orientation of the stimulated hand relative to the head, implying that multimodal interactions depend on the overlap of reference frames. This insight is in line with Zacks and Michelon's model, emphasizing that the effect of motor rotation on mental rotation depends on the alignment of the spatial axes with respect to which the rotations are performed.

The assumption of an internal rotation model that is partially grounded in motor codes also dovetails with Wexler et al. (1998)'s hypothesis that mental rotation involves the simulation of the perceptual outcomes of a rotational movement. Applying a dual-task interference paradigm, Wexler et al. showed a close coupling between mental rotation performance and motor rotations. In this experiment, participants had to perform a primary mental rotation task concurrently with a secondary motor rotation task. From the apparent overlap of the two rotations, Wexler et al. concluded that mental rotation relies on motor processes. Besides a general compatibility effect, yielding faster response times and fewer errors when motor and mental rotation direction matched, Wexler et al. showed that changes in the speed of motor rotation could slow down or speed up the mental rotation. Wexler et al. concluded that the interaction between mental and motor rotations is probably rooted in visuomotor anticipations. If our ability to mentally rotate objects is indeed grounded in the multisensory experience of actual rotations, one could expect selective interference between mental rotation and a concurrent rotation perception. Such an interference pattern would be in line with the assumption of an embodied simulation underlying mental rotation, but would conflict with models assuming an encapsulated, abstract, symbolic (Just & Carpenter, 1976) rotation process.

Even if the proposed internal rotation model provides a suitable account for the question how mental rotation is performed, it remains still open on *what* kind of representation the model operates. Recent evidence highlights the central role of informative spatial features as proposed by Just and Carpenter (1976) as well as by Zacks and Michelon (2005). To further contrast the role of visual compared to spatial stimulus features, Liesefeld and Zimmer (2013) investigated participants' mental rotation performance for different kinds of stimuli. In this study, a single stimulus was presented to the participants at the beginning of the trial, followed by a rotation cue indicating the disparity of the comparison stimulus. Subsequently, the rotated comparison stimulus appeared. Stimuli consisted of simple geometric figures, which differed with respect to orientation-dependent spatial information and visual complexity. Liesefeld and Zimmer compared the rotational speed, that is, the slopes relating response time and angular disparity, for the different types of stimuli. While an increase of orientation-dependent spatial information decreased rotational speed, variations of visual complexity did not influence rotation speed. The results suggest that the rotated representation relied primarily on orientation-dependent spatial information, instead of visual information. Thus, the results highlight the dominant role of spatial information in the representations that are manipulated during mental rotation.

In line with these results, Schwartz and Holton (2000) showed that mental spatial transformations are affected by previously obtained action knowledge. Schwartz and Holton requested participants to perform mental rotations of physical objects placed in front of them. Participants could rotate the platform on which the objects were located by pulling a spool, that is, by a non-rotational movement. Participants learned that pulling the spool could result in either a clockwise or counterclockwise rotation of the object. This induced rotation either matched the requested mental rotation direction or not. Results indicated a strong congruency effect between the learned rotation direction and the mental rotation measured by the reaction time, even when participants closed their eyes and thus did not receive visual feedback about the rotation direction. Thus, the learned mental models between motor actions, physical effects, and visual rotation effects affected mental rotation, again revealing a close relationship between mental transformation processes and transformations performed in the real world.

Based on the mentioned models and studies, we propose that mental rotation is realized by an internal spatial simulation of an actual rotation. This internal simulation capacity develops most likely from actual action experiences, forming predictive sensorimotor-grounded encodings about motor-induced object transformations. With further practice in performing object rotations and by observing others producing similar rotations, these encodings can be expected to be further generalized. In the end, a purely spatial, mental predictive encoding may be formed, which can be used to perceive, to simulate, or to execute any type of rotation (Butz, 2016). If these embodied simulation codes indeed exist, different modalities should be able to activate, access, and interact with them. For instance a tactile rotation stimulation should be – at least partially – processed by the same codes that are used in the simulation of a visual rotation. In contrast to this embodied view, if the representation underlying mental rotation is linked to only one modality, such as the visual modality, or to an encapsulated, abstract spatial process, there should be no selective interaction with rotational perceptions in other modalities. To gather further evidence in favor of an embodied spatial encoding, underlying mental rotation, we asked the question if mental rotation performance can be affected by the simultaneous presentation of a rotating stimulus in the tactile domain.

Experiment 1

The primary aim of the first experiment was to investigate selective interference between tactile and mental rotation. To do so, we combined a primary mental rotation task with a secondary tactile change detection task. In the primary task, participants had to judge the parity of rotated letters (i.e. canonical or mirrored). In the secondary task, participants had to detect changes in the rotation direction of a tactile stimulation applied to their palm (Butz et al., 2010). If tactile rotation can indeed affect mental rotation, response times should differ for conditions where both rotation directions match, compared to conditions where they do not match.

Method

Participants. Twenty students from the University of Tübingen participated in this study (six males). The mean age was 21.25 years (SD = 2.5). All participants were right-handed and had normal or corrected-to-normal vision. Participants provided informed, written consent and received either monetary compensation or course credit for their participation.

Apparatus. Visual stimuli were presented on a 22-in LCD monitor with a resolution of 1680 × 1050 pixels. Participants were seated 60 cm away from the monitor. Participants' responses were recorded via a keyboard placed in front of the monitor. Participants responded by pressing the arrow keys with their right hand.

For the tactile secondary task a tactile stimulator was used, which applied a tactile rotation to the participants' left hand's palm (see Fig. 1). The tactile stimulator consisted of a rotating disk, which could be rotated via two motors. A metal wheel, connected with the rotating disk via a spring served as the actual stimulus generator. The wheel could rotate either clockwise or counterclockwise with different speeds. Rotation direction as well as the angular velocity could be changed via a low-level hardware interface (parallel port and custom C software). Participants had to insert their left hand beneath the wheel, such that the tactile rotation stimulation could be applied. The wheel had a flat surface and subtended 12 mm in diameter and 4 mm in depth. The angular velocity was kept constant in this setup and resulted in 0.33 cycles per second. To cover acoustic detection of changes in the rotation direction, an integrated loudspeaker emitted white noise, effectively masking the sound of the motors. To make the setup more convenient, the part of the device in which the participants had to insert their hand was padded. Furthermore, the stimulator was fastened with elastic straps, such that participants did not have to actively press their hands onto the stimulator. We only applied stimulation to the participants' palm (the device can also stimulate the back of the hand). To avoid confusion, we will report tactile stimulation direction from a head-centered frame of reference, as this was shown to be the critically perceived rotation direction (Butz et al., 2010). Hence, a clockwise / counterclockwise stimulation refers to a rotation stimulation on the palm from the viewpoint of the participants' head.

Material. Stimuli consisted of the letters F, P, R, and L, which were either presented in their canonical or mirrored version. Furthermore, the letters were rotated by either 70°, 100°, 130° or 160° clockwise or counterclockwise from their vertical upright. We also included a 0° condition to refresh the visual letter shapes and to keep the typical orientation salient. This might foster the identification of the letters and in turn facilitate the decision whether a clockwise, or counterclockwise rotation is required. The letters subtended an area with a radius of 2.85° of visual angle. Letters were presented in the center of the screen tinted in black in front of a white background.

Procedure. The trial schedule is shown in Fig. 2. Each trial started with the presentation of a fixation cross for 1000 ms. At the same time as the fixation cross was displayed, the tactile stimulation started. Next, a single letter was presented for 1000 ms. If a change in the tactile rotation direction took place, it occurred either 200 ms before, simultaneously with, or 200 ms after the onset of the target letter. We introduced this temporal uncertainty of the tactile rotation change onset for two reasons. First, to control for possible, unwanted cross-modal masking effects between visual onset and tactile change onset (see for instance Gallace, Auvray, Tan, & Spence, 2006). Second, we wanted to prevent temporal preparation (see e.g., Rolke & Hofmann, 2007) to the tactile change onset, which might provoke disturbance from the mental rotation task. The display remained blank after the stimulus offset for maximally 2000 ms. If participants did not respond within this interval, the trial was considered as an error trial and the participants received according feedback ("please respond faster"). The respective trials were treated as error trials and were excluded from the response time analysis, misses were extremely rare however (3‰).

Participants had to indicate whether the displayed letter was presented in its canonical or in its mirror-image version. This parity judgment was given via the arrow keys. Participants were instructed to press the left arrow key to indicate a mirrored letter, and the right arrow key to indicate a canonical letter. This mapping was chosen to avoid unnecessary incompatibility effects, because the right side is typically associated with a positive, confirming response (see for instance Natale, Gur, & Gur, 1983; Casasanto, 2009), like a correct letter presentation, whereas the left side is associated with a negative, dis-confirming response, like an incorrect, mirrored letter presentation. After providing the parity judgment, a second response screen appeared. Here the participants had to indicate whether the direction of the tactile rotation changed throughout the trial or not. Responses for the tactile change detection task were given with the number pad keys (0 when no change occurred, 1 in case the tactile rotation direction changed). To ensure that participants focused on the secondary task, we repeated trials where participants failed the secondary task at the end of the block, hence the hit rate in the tactile change detection task was 100% [1]. The number of remaining trials within the block was displayed on the lower right of the screen in the inter-trial interval. The whole experiment was self-paced. Between trials, participants had to press the up arrow button to proceed with the next trial. The experiment started with 40 training trials to familiarize the participants with the procedure. The main experiment only started when the 40 training trials were completed without error in both the primary and the secondary task. We recorded response times and error rates for the parity judgment.

After the training trials, the main experiment started. It consisted of 864 trials, presented in a single block. The trial number resulted from combining the nine disparities (0°, +/−70°, +/−100°, +/−130° or +/−160°) with the four letters (F, P, R, and L) and the two parities (canonical or mirrored), repeating each combination twelve times. The tactile stimulation conditions were balanced within the twelve repetitions. There were four tactile stimulation conditions: in the two no-change conditions, tactile stimulation was either clockwise or counterclockwise; in the two change conditions, tactile stimulation changed from clockwise to counterclockwise, or vice versa. Each condition was repeated three times. In both change conditions, the change happened in the three trials once 200 ms before, once concurrently with, and once 200 ms after

---

[1] On average, 5% (SD = 4%) of the trials had to be repeated.

visual stimulus presentation. To maintain an equal proportion of change and no-change trials, three identical no-change trials were presented for each change condition. Hence the overall chance for a change in the tactile stimulation direction was 50% throughout the experiment.

Results

Response times and error rates for the parity judgment were analyzed with repeated measures ANOVAs using R (R Core Team, 2016) and the *ez* package (Lawrence, 2015). Since the onset factor was only present in trials with tactile change, we separated the analysis of change and no-change trials. We aggregated data over the different letters and onset conditions, hence letter identity and onset were not included as separate factors [2]. As we were mostly interested in interactions between mental and tactile rotation direction, we created a mental rotation direction factor by the sign of the angular disparity. For angles below 0°, the mental rotation direction is clockwise, for angles larger than 0°, the mental rotation direction is counterclockwise. In the 0°, no mental rotation is required, hence these trials can be used to check whether tactile rotation alone had an effect on the parity judgment. We analyzed trials with non-rotated letters separately. To focus the analysis further, we aggregated over the actual mental rotation magnitudes, that is, over the four letter disparities in either direction.

For the analysis of response times, only correct trials were used. Furthermore, we used log-transformed response times to avoid possible effects of non-normality on the analysis. After the log-transform, response times above or below two times of the standard deviation were excluded. All in all, 88% of the trials yielded correct responses, due to the filtering by standard deviation, 342 trials (2.1% of all valid trials) were excluded. All post-hoc t-tests were adjusted for multiple comparisons by the method proposed by Holm (Holm, 1979).

---

[2] As noted above, the onset variation was primarily intended as a control factor. Including it in the analyses yielded only a main effect and no interactions with other factors. In general participants were faster in case of earlier onsets.

No-Change Trials. Data were analyzed with a 2 (parity) × 2 (tactile rotation direction) × 2 (mental rotation direction) factors repeated measures ANOVA. Results are shown in Tab. 1. There were no main effects, but an interaction between mental rotation direction and parity (see Fig. 4, left panel). Post-hoc t-tests revealed that participants were faster when they had to rotate mirrored letters counterclockwise compared to clockwise ($t(19) = -2.51$, $p = .042$). For canonical letters, participants responded faster when they had to mentally rotate clockwise compared to counterclockwise ($t(19) = -3.97$, $p < .001$). Tactile rotation yielded no significant effects.

Error Rates. The analysis of the error rates is shown in Tab. 2. The interaction between mental rotation direction and parity was also visible in the error data, in the same direction as for the response times. However, post-hoc t-tests showed no significant differences. Besides this interaction, there was also a main effect of parity: participants made fewer errors when judging mirrored letters.

Trials Without Mental Rotation. To further check whether tactile rotation alone had an effect on the parity judgment, we analyzed the trials with 0° disparity with a 2 (parity) × 2 (tactile rotation direction) factors repeated measures ANOVA for response times and error rates. There were no main effects or interactions.

Change Trials. Data were analyzed with a 2 (parity) × 2 (tactile rotation direction) × 2 (mental rotation direction) factors repeated measures ANOVA. Results are shown in Tab. 3. The analysis yielded three two-way interactions: for parity and tactile rotation direction, parity and mental rotation direction, and mental and tactile rotation direction.

With respect to the interaction between tactile rotation direction and parity, significant differences were only observed when the tactile stimulation switched from counterclockwise to clockwise. In this case, participants were faster to respond to canonical letters, compared to mirrored letters ($t(19) = -2.98$, $p = .03$).

The interaction between mental rotation direction and parity showed the same pattern as in no-change trials (see Fig. 4, right panel). Participants were significantly faster when rotating canonical letters clockwise, compared to when they had to rotate them counterclockwise ($t(19) = -4.77$, $p < .001$). Numerically, the opposite was true for mirrored letters, however, compared to no-change trials, the respective difference failed to reach significance ($t(19) = -1.75$, $p = .191$).

In contrast to no-change trials and most importantly, the interaction between mental rotation direction and tactile rotation direction was significant. If participants had to mentally rotate clockwise, they were faster when the tactile rotation changed to clockwise, as when it changed to counterclockwise ($t(19) = -3.19$, $p = .014$). If participants had to mentally rotate counterclockwise, they were faster when the tactile rotation changed to counterclockwise, as when it changed to clockwise ($t(19) = -3.77$, $p < .01$). The results are depicted in Fig. 3.

Error Rates. The analysis of the error rates showed a similar pattern as in the no-change trials. The results are shown in Tab. 4. Besides the main effect for parity and the two-way interaction between parity and mental rotation direction, the error rates yield an interaction between mental and tactile rotation direction, which dovetails with the results from the response times: the error rates are reduced when the tactile rotation direction after the change matched the mental rotation direction. After adjusting for multiple comparisons, however, only the respective differences for clockwise mental rotations ($t(19) = -2.63$, $p = .033$) remained significant. As for the analysis of the response times, the interaction between parity and tactile rotation direction was significant ($F(1,19) = 4.56$, $p = .046$, $\eta_p^2 = .19$). Post-hoc t-tests showed a similar tendency as for the response times: participants made fewer errors when responding to canonical letters when the tactile stimulation switched from counterclockwise to clockwise. However, after correcting for multiple comparisons, the respective difference failed to reach significance. Finally, the analysis yielded a three-way interaction between parity, mental, and tactile rotation direction. The post-hoc analysis yielded only a significant difference in case of clockwise mental rotations and canonical letters. In this case participants made fewer errors when the tactile rotation switched to clockwise instead of counterclockwise ($t(19) = -3.94$, $p = .003$).

Trials Without Mental Rotation. To further check whether tactile rotation alone had an effect on the parity judgment, we analyzed the trials with 0° disparity with a 2 (parity) × 2 (tactile rotation direction) factors repeated measures ANOVA, both for response times and error rates. There were no main effects or interactions with respect to error rates or response times.

Besides the same compatibility effect between mental rotation direction and parity that was observed for the no-change trials, the results show two additional compatibility effects. First, the hypothesized interaction between mental and tactile rotation turned out to be significant. Second, the analysis yielded a significant interaction between tactile rotation direction and parity.

Discussion

Our aim was to investigate interactions between mental and tactile rotation. Data from change trials indeed showed the hypothesized interaction. When the tactile rotation direction after the change matched the mental rotation direction, response times were significantly faster compared to trials where the directions did not match. Besides the compatibility between mental and tactile rotation direction, two additional compatibility effects reached significance. First, there was an interaction between parity and tactile rotation direction in change trials. Participants were faster to respond to canonical letters when the tactile rotation switched to clockwise, compared to counterclockwise. Second, data from change as well as no change trials yielded an interaction between mental rotation direction and parity. Participants were faster when responding to mirrored letters when they had to rotate them counterclockwise compared to when they had to rotate them clockwise. For canonical letters, the opposite was true. The reason for this interaction between mental rotation and parity cannot be easily inferred from the results of the first experiment. This is due to the fact that in our experimental setup parity and response side cannot be disentangled because participants responded with the left arrow button to mirrored letters and with the right arrow button to canonical letters. Thus, it is not clear whether mental rotation direction interacted with the parity of the letters (canonical or mirrored) or with the response mapping (left or right finger). Experiment 2 was conducted to disentangle these two potential sources of interference.

Experiment 2

To disentangle the parity effect, we varied the response mapping between blocks. In one block of trials, participants had to respond with their index finger to mirrored letters and with their middle finger to canonical letters (original response mapping applied in Experiment 1), in the other block of trials, this mapping was reversed: participants had to respond with their index finger to canonical letters and with their middle finger to mirrored letters (inverted mapping). If the interaction between mental rotation direction and parity observed in Experiment 1 is due to an effect of response side, instead of visual letter shape, there should be a three-way interaction between mental rotation direction, parity, and mapping. If the observed interaction is due to a compatibility between mental rotation direction and visual letter shape, the response mapping should not affect this interaction. To reduce the number of trials per block and to focus on the effects of primary interest, we simplified the experimental setup by discarding the onset manipulation. Furthermore, we only employed the more extreme angular disparities greater than 70°.

Participants. Twenty students from the University of Tübingen participated in this study (six males). The mean age was 25.55 years (SD = 3.5). All participants were right-handed and had normal or corrected-to-normal vision. Participants provided informed consent and received either monetary compensation or course credit for their participation. None of the participants participated in the first experiment.

Apparatus. The same apparatus as for the first experiment was used.

Material. The stimulus set was the same as in the first experiment except that we did not use the rotation of 70° in this experiment. Hence, the letters were rotated by either 0°, 100°, 130° or 160° clockwise or counterclockwise from their vertical upright.

Procedure. The course of a single trial was the same as in the first experiment, except that the tactile change onset always occurred simultaneously with the onset of the target letter.

Participants had to indicate whether the displayed letter was presented in its canonical or in its mirror-image version. The response mapping of the parity judgment varied between the two blocks of the experiment, which were conducted on two successive days. In one block they had to press the left arrow key to indicate a mirrored letter, and the right arrow key to indicate a canonical letter, in the other block the mapping was reversed. The order of the mappings was balanced over the participants. As in Experiment 1, each block started with 40 training trials to familiarize the participants with the procedure and the response mapping.

The main experiment consisted of two blocks with 896 trials each. The trial number resulted from combining the seven disparities (0°, +/−100°, +/−130° or +/−160°) with the four letters (F, P, R, and L) and the two parities (canonical or mirrored), repeating each combination 16 times. Tactile stimulation could either stay constant in a clockwise or counterclockwise fashion, or change from clockwise to counterclockwise, or vice versa. Each of the four cases occurred in 4 out of the 16 trials for each combination of disparity, letter type, and parity. As in Experiment 1, participants had to repeat trials where they failed the tactile change detection task at the end of the block[3].

Results

As in the first experiment, the miss rate was extremely low (4‰). All in all, 92% of the trials yielded correct responses, due to the filtering by standard deviation, 266 trials (2.8% of all valid trials) were excluded. Response times and error rates for the parity judgment were analyzed with 2 (mapping) × 2 (parity) × 2 (tactile rotation direction) × 2 (mental rotation direction) factors repeated measures ANOVAs. Again, we separated the analysis of change and no-change trials. To check whether tactile rotation alone had an effect on the parity judgment, we analyzed trials with non-rotated letters separately.

---

[3] On average, 7% (SD = 5%) of the trials had to be repeated.

No-Change Trials. Results are shown in Tab. 5. The analysis yielded a main effect for mapping, participants were faster when they responded with the left key to mirrored stimuli and with the right key to canonical stimuli, compared to the inverted mapping. Importantly, there was a three-way interaction between mapping, parity, and mental rotation direction. Descriptively, participants responded faster to canonical letters, when they had to mentally rotate the letter clockwise and had to respond with the right key. The opposite was true for mirrored letters. After adjusting for multiple comparisons, however, post-hoc t-tests failed to reach significance. An overview of the data is given in Fig. 6, left panel. As in the first experiment, tactile rotation had no significant effect and did not interact with other factors.

Error Rates. Results are shown in Tab. 6. The analysis of the error rates yielded a three-way interaction between mapping, parity, and mental rotation direction, which went in the same direction as for the response times, indicating that the respective response time effects were not due to speed accuracy trade-offs. The main effect for mapping was present in the error data as well: participants made fewer errors when they responded with the left key to mirrored stimuli and with the right key to canonical stimuli, compared to the inverted mapping. Additionally, the analysis of the error rates yielded a main effect for mental rotation direction. Participants made fewer errors when they had to mentally rotate clockwise compared to when they had to rotate counterclockwise. Seeing that this effect was not observed in the first experiment, we analyzed the two mapping conditions separately, and indeed the main effect was only present in case of the inverted mapping.

Trials Without Mental Rotation. To further check whether tactile rotation alone had an effect on the parity judgment, we analyzed the trials with 0° disparity with a 2 (mapping) × 2 (parity) × 2 (tactile rotation direction) factors repeated measures ANOVA, both for response times and error rates. In both cases there was a main effect for the mapping. Participants were faster and made fewer errors when they had to respond with the left key to mirrored stimuli and with the right key to canonical stimuli, compared to the inverted mapping. There were no other main effects or interactions.

Change Trials. Results are shown in Tab. 7. As in no-change trials, participants were faster when they responded with the left key to mirrored stimuli and with the right key to canonical stimuli, compared to the inverted mapping. Most importantly, the two-way interaction between parity and mental rotation direction was modified by the response mapping, yielding a three-way interaction between mapping, parity, and mental rotation direction. Breaking down the three-way interaction, post-hoc t-tests yielded significant differences in case of the original response mapping of Experiment 1, where participants responded with the left arrow key to mirrored letters. In this case, participants were faster when they had to mentally rotate mirrored letters counterclockwise compared to when they had to rotate them clockwise ($t(19) = -3.04$, $p = .020$). For canonical letters, the opposite was true ($t(19) = -4.20$, $p = .002$). Numerically, this pattern was inverted for the inverted mapping. However, the respective differences did not reach significance (see Fig. 6, right panel). Similar to the change trials in the first experiment, the analysis yielded an interaction between mental and tactile rotation direction. Participants responded faster when the direction of the tactile stimulation after the change matched the direction of the requested mental rotation (see Fig. 5, right panel). The respective difference was significant both for mental rotations in a clockwise ($t(19) = -3.05$, $p = .010$), and mental rotations in a counterclockwise fashion ($t(19) = -3.19$, $p = .010$).

Error Rates. Results are shown in Tab. 8. The analysis of the error rates yielded a main effect for mapping, as well as a two-way interaction between mapping and mental rotation direction, and a three-way interaction between mapping, parity, and mental rotation direction. Participants made fewer errors when they responded with the left key to mirrored stimuli and with the right key to canonical stimuli, compared to the inverted mapping.

Post-hoc t-tests for the interaction between mapping and mental rotation direction showed that participants made fewer errors when performing clockwise compared to counterclockwise mental rotations in case of the original mapping ($t(19) = -2.38$, $p = .056$). In case of the inverted mapping, the respective difference did not reach significance. The pattern of the three-way interaction between mapping, parity, and mental rotation direction matched the response time data. Post-hoc t-tests showed that in case of the original mapping and for canonical letters, participants made fewer errors when they had to mentally rotate clockwise compared to when they had to rotate counterclockwise ($t(19) = -3.82$, $p = .005$).

Trials Without Mental Rotation. To further check whether tactile rotation alone had an effect on the parity judgment, we analyzed the trials with 0° disparity with a 2 (mapping) × 2 (parity) × 2 (tactile rotation direction) factors repeated measures ANOVA, both for response times and error rates. While there were no significant effects with respect to error rates, the analysis of response times yielded main effects for mapping and tactile rotation direction, as well as for the respective two-way interaction. Furthermore, the two-way interaction between parity and tactile rotation direction turned out to be significant. Participants were faster when they responded with the left key to mirrored stimuli and with the right key to canonical stimuli, compared to the inverted mapping. Response times were faster when the tactile rotation direction turned to clockwise, instead of counterclockwise. Post-hoc t-tests for the interaction between mapping and tactile rotation direction showed that this was only the case for the inverted mapping ($t(19) = -3.25$, $p = .008$). Post-hoc t-tests for the two-way interaction between parity and tactile rotation direction showed faster response times for mirrored stimuli when the tactile rotation changed to clockwise ($t(19) = 4.37$, $p = .001$). Even though the three-way interaction between mapping, parity, and tactile rotation direction did not reach significance ($F(1,19) = 2.33$, $p = .143$, $\eta_p^2 = .11$), we tested whether the interaction between parity and tactile rotation direction followed the same pattern for the two mappings. Indeed, the difference between clockwise and counterclockwise stimulation was only significant for mirrored stimuli in case of the inverted mapping, where participants had to respond with the right key to mirrored letters ($t(19) = 5.49$, $p < .001$).

Discussion

The results of this second experiment replicate and extend those from the first experiment. Irrespective of the mapping condition, the results again indicate a compatibility effect between mental and tactile rotation, but only when the tactile rotation changes direction (see Fig.7, right panel). For both, change and no-change trials, the interaction between mental rotation direction and parity was affected by the mapping condition. Specifically, when participants had to respond to mirrored letters with the left key, response times were faster for counterclockwise mental rotations. For canonical letters, the opposite was true. In the case of the inverted mapping, this effect vanished. This influence of response mapping suggests that the letter identity, i.e. the fact that a letter appears in its canonical or its mirrored form, does not cause compatibility effects with mental rotation. Instead, it corroborates evidence for a compatibility effect between mental rotation direction and response side. Figure 7 provides an overview of the two compatibility effects across both experiments. To allow an estimate of the magnitude of the compatibility effects in terms of response times, Fig.7 shows non-transformed response times as well as error rates.

With respect to a third putative compatibility effect between response side and tactile rotation direction, the results remain inconclusive. In the first experiment, an interaction between parity and tactile rotation direction was observed for change trials. In the second experiment this interaction was only visible in trials that involved no mental rotation (disparity of 0°), and in case of the inverted mapping. If tactile rotation alone indeed biases the response side, this cannot be reliably determined with the current paradigm.

The second experiment again showed that mental rotation can be affected by the perception of tactile rotations and, as in the first experiment, this is only the case when the tactile rotation changes direction. Taken together, the second experiment yielded two important results. First, it confirmed the influence of tactile rotation on mental rotation. Second, it showed that mental rotation direction facilitates specific response sides.

General Discussion

According to embodied cognition, higher cognitive functions are deeply rooted in the sensorimotor system of an agent (Barsalou, 2008, 2010). In this view, active interaction with the environment allows the formation of mental models about the correlations between motor activity and environmental changes. Presumably, however, during practice the mental models become more and more independent of the original sensorimotor experiences, but they stay grounded and thus links to the sensory and motor systems remain. We employed a dual-task interference paradigm in two experiments to reveal some of those links. Mental rotation was selectively affected by a concurrently perceived tactile rotation. Furthermore, a compatibility between mental rotation direction and response side was detected.

Compatibility between Mental and Tactile Rotation

While there are a lot of data relating mental rotation performance to motor activation (see e.g. Wexler et al., 1998), little is known about the involvement of other modal codes. In two experiments, we observed a compatibility effect between mental and tactile rotation direction, when the tactile rotation changed during the trial. If the tactile rotation direction after the change matched the direction of mental rotation, response times were faster than when they mismatched. Apparently, the perceived tactile rotation change interacted with the ongoing mental rotation task. This compatibility effect is in line with the assumption of a general rotation model, which is grounded and can be applied in various modalities.

The question remains, however, how this observed interaction took place. Since the compatibility effect between the rotation directions was only visible in change trials, it seems that the tactile stimulation is not automatically integrated into the mental rotation process, but rather needed to be attended in order to be effective. We assume that the change in tactile rotation direction captured attention of the processor away from the mental rotation process to the tactile stimulation. This attentional capture most probably strengthened the tactile input, which in turn might have gained a higher processing influence and thus a higher probability to interact with the ongoing mental rotation process compared to unattended tactile rotation. In our opinion, this outlined processing idea is supported by two results of the present study. First, the compatibility effect between mental rotation and tactile stimulation happens to be based on the tactile rotation direction following the rotation change, which should be the one that is strengthened by attentional capture. Second, the tactile rotation change seems to distract processing resources away from mental rotation, which is indicated by the result that response times in change trials were elevated compared to those in no-change trials.

Taken together, our two experiments show that the direction of the tactile stimulation is integrated into the mental rotation process when the tactile simulation was sufficiently attended. Thus, by increasing the salience of the tactile stimulation, it might be possible to obtain compatibility effects even when the rotation direction does not change. This could be done, for example, by enhancing the processing demands for the tactile task, or by changing the task priorities by demanding to answer the tactile task before the parity judgment task in some of the trials.

It is worth mentioning that the rotation compatibility was obtained even though the two rotations took place in different frames of reference. Specifically, mental rotation of the letters was conducted most probably in a vertical frame of reference, while tactile rotation processing was accomplished in a horizontal frame of reference. Thus, not the specific spatial rotation axes but the general direction of the rotation processes seemed to be the most important factor, that is, whether the rotation direction was clockwise or counterclockwise. It might be that stronger compatibility effects were obtained when the spatial rotation axes match more closely. This may be achieved by using the original cube figures used by Shepard and Metzler (1971), which had to be rotated around different spatial axes. Another possibility might be that the tactile stimulation would be changed from a horizontal to a vertical axis, which would match the required mental rotation direction exactly. We expect that different alignments will modify the extent of observable stimulus interactions.

Note that we found an additional, subtle compatibility effect between tactile stimulation direction and response side during change trials in the first experiment. However, the pattern of results from the second experiment did not replicate this result. Thus, this potential compatibility effect remains inconclusive and must be investigated in further studies.

Bidirectional Connections between Mental Rotation and Motor Codes

Our results show that the direction of mental rotation facilitates specific response sides. This compatibility effect between mental rotation direction and the response side was present in both reported experiments and in tactile change trials as well as in no-change trials. The variation of the response mapping in the second experiment made clear that the mental rotation direction indeed interacted with the response side and not with parity. To the best of our knowledge, this is the first time that such a compatibility effect is reported. Moreover, in the second experiment the overall compatibility effect is stronger for the mapping in which participants responded to mirrored stimuli with the left button (same mapping as in the first experiment, see Fig. 6), compared to responses to mirrored stimuli with the right button. This modulation of the response compatibility effect might be due to the fact that the original response mapping is the more natural mapping (see e.g. Natale et al., 1983), which might be more prone to compatibility effects. Taken as a whole, the compatibility effect implies a bidirectional mapping between mental rotation and motor codes. While various studies have shown that concurrent motor activation can affect mental rotation performance (see e.g. Kosslyn, Ganis, & Thompson, 2001 for an overview), our results imply that mental rotation can prime the response side. More specifically, it seems that the end point of the rotation, that is, the rotation direction towards an upright position of the letters facilitates the response finger. For example, when a mental rotation required a rotation from left to right, the right finger responses were faster compared to when there was a mental rotation from right to left.

These results dovetail with earlier reports on the link between motion direction and perception of rotational movements (Wohlschläger, 2000), work on the spatial compatibility between rotational movements and horizontal stimulus location (Wang, Proctor, & Pick, 2003), as well as reports on motor rotation priming by means of mental rotations (Janczyk, Pfister, Crognale, & Kunde, 2012). In our experiments, mental rotations primed a response side. To generalize this compatibility effect, it would be necessary to investigate whether the compatibility is restricted to motor codes of specific effectors or different response modalities. For instance, Gardony et al. (2014) used verbal responses in their mental rotation experiments. If a similar compatibility effect can be obtained with verbal responses, this would corroborate further evidence for a general mental rotation mechanism, which interacts with different modal effector codes.

Note also that in our setup negative letter disparities required a clockwise rotation and positive disparities required a counterclockwise rotation. Consequentially, disparity prefix and mental rotation direction cannot be disentangled in our design. A separate experiment would be necessary to assure that compatibility is due to the mental rotation direction irrespective of character tilt. For an investigation like this, a design similar to the one applied by Liesefeld and Zimmer (2011) would be necessary. In their study participants had to compare a target and a reference letter with respect to parity. The letters were presented sequentially. The reference letter was either presented rotated or upright. In the first case, the target was upright, while in the second case, the target was rotated. Hence, mental rotation was necessary in both conditions, but the target letter was tilted in only one condition. The observation of a compatibility effect between mental rotation direction and response side in both conditions of such a setup would corroborate the compatibility between response side and mental rotation direction, while arguing against the alternative that the letter tilt resulted in the observed response compatibility effect.

Conclusion

Our aim was to shed light on two central questions related to mental rotation. First, we wanted to investigate the cognitive mechanisms underlying mental rotation that is *how* mental rotation is accomplished. Second, we wanted to probe the representation this process is operating on that is *what* is rotated. In two experiments we have shown that tactile rotation can interact selectively with mental rotation.

Furthermore, our results imply that mental rotation direction can prime spatially compatible motor responses. Both, the apparent multimodal grounding of mental rotation, as well as the spatial compatibility effect are hard to integrate into a model that assumes a purely visual representational basis for mental rotations, or an encapsulated, amodal one. Rather, our results suggest that a more universal, spatially-grounded rotation mechanism is at play.

Following the embodied cognition hypothesis, it can be assumed that the ability to mentally rotate objects is rooted in the sensorimotor system. Performing a physical rotation yields visual, tactile, and proprioceptive effects and is accompanied with certain motor commands. What links these effects together is the spatio-temporal change of the rotated object. Thus, learning can take place and can integrate the experiences into a predictive model. With more and more practice, the spatio-temporal patterns can be expected to develop into more abstract spatial rotation encodings, which may be implemented by suitably structured spatial predictive encodings (Butz, 2016). However, traces of the sensorimotor grounding remain and therefore mental rotation can yield (i) compatibility effects between rotational stimulations in different sensory modalities, (ii) compatibility effects with respect to motor commands, and (iii) compatibility effects between mental rotations and rotational sensory stimulations or motor activities. This mechanism provides an answer to the *how* and *what* questions. According to this view, mental rotation is realized in terms of a spatial predictions of action outcomes, the underlying representation seems to be primarily spatial, linking rotational patterns in various modalities.

The sketched-out developmental pathway should be rooted in the experiences with rotations, including self-induced rotations of objects and other stimuli as well as observed rotations. Jansen and Kellner (2015) provided direct evidence for a closer coupling of mental rotation and motor activity in children compared to adults. In their experiment, children were presented with two images of an animal. On the left side, the image was presented in its canonical orientation, while the image on the right side was rotated to some degree. The children had to judge whether the left and right images were identical, or whether either one was mirrored. While performing the mental rotation task, children were required to perform a concurrent manual rotation, which either matched mental rotation direction or not. Comparison of the age groups (7- to 8-year-old vs. 9- to 10-year-old children) revealed a compatibility effect of manual and mental rotation only for the younger group. Furthermore, the effect only occurred for boys. These results are in line with the assumed development of a general mental rotation mechanism that is grounded in sensorimotor experience and becomes progressively more abstract with practice and experience – even though this admittedly still hypothetical explanation does not account for the gender difference.

In conclusion, our results provide further evidence that our brain develops a spatially encoded rotation mechanism. According to Zacks (2008), the most likely neural areas that encode this rotation mechanism can be found in the intraparietal sulcus and neighboring regions and interact with the medial superior precentral cortex, particularly when motor simulations are involved. Seeing the available developmental psychology and embodied cognitive science literature, this mechanism most likely develops from and is thus grounded in sensory experiences of, and sensorimotor experiences with, rotating objects and other rotating stimuli. Once sufficiently well-structured, the mechanism supports the perception of rotating stimuli, the execution of mental rotations, as well as the execution of physical rotations. Further research is necessary to evaluate the universality of such a mechanism and the nature of its interaction with the three types of cognitive tasks, that is, perception, mental simulation, and action decision making and execution.

References

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22* (4), 637–660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59* , 617–645.

Barsalou, L. W. (2010). Grounded cognition: past, present, and future. *Topics in Cognitive Science*, *2* (4), 716–724.

Butz, M. V. (2016). Towards a unified sub-symbolic computational theory of cognition. *Frontiers in Psychology*, *7* (925). doi: 10.3389/fpsyg.2016.00925

Butz, M. V., Thomaschke, R., Linhardt, M. J., & Herbort, O. (2010). Remapping motion across modalities: Tactile rotations influence visual motion judgments. *Experimental Brain Research*, *207* , 1–11.

Carpenter, P. A., & Eisenberg, P. (1978). Mental rotation and the frame of reference in blind and sighted individuals. *Attention, Perception, & Psychophysics*, *23* (2), 117–124.

Casasanto, D. (2009). Embodiment of abstract concepts: good and bad in right-and left-handers. *Journal of Experimental Psychology: General*, *138* (3), 351–367.

Cohen, M. S., Kosslyn, S. M., Breiter, H. C., DiGirolamo, G. J., Thompson, W. L., Anderson, A., Belliveau, J. (1996). Changes in cortical activity during mental rotation a mapping study using functional MRI. *Brain*, *119* (1), 89–100.

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In W. G. Chase (Ed.), *Visual Information Processing* (p. 72-121). New York: Academic Press.

Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of experimental psychology: human perception and performance*, *27* (1), 229–240.

Gallace, A., Auvray, M., Tan, H. Z., & Spence, C. (2006). When visual transients impair tactile change detection: A novel case of crossmodal change blindness? *Neuroscience Letters*, *398* (3), 280–285.

Gardony, A. L., Taylor, H. A., & Brunyé, T. T. (2014). What does physical rotation reveal about mental rotation? *Psychological Science*, *25* (2), 605–612.

Gill, H. S., O'Boyle, M. W., & Hathaway, J. (1998). Cortical distribution of eeg activity for component processes during mental rotation. *Cortex*, *34*(5), 707–718.

Heil, M., & Rolke, B. (2002). Toward a chronopsychophysiology of mental rotation. *Psychophysiology*, *39*(4), 414–422.

Hoffmann, J. (1993). *Vorhersage und Erkenntnis: Die Funktion von Antizipationen in der menschlichen Verhaltenssteuerung und Wahrnehmung*. Verlag für Psychologie, Hogrefe.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 65–70.

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, *24*, 849–937.

Janczyk, M., Pfister, R., Crognale, M. A., & Kunde, W. (2012). Effective rotations: Action effects determine the interplay of mental and manual rotations. *Journal of Experimental Psychology: General*, *141*, 489-501.

Jansen, P., & Kellner, J. (2015). The role of rotational hand movements and general motor ability in children's mental rotation performance. *Frontiers in Psychology*, *6*(984). doi: 10.3389/fpsyg.2015.00984

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology*, *8*(4), 441–480.

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*(9), 635–642.

Kunde, W. (2001). Response-effect compatibility in manual choice reaction tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *27*(2), 387–394.

Lawrence, M. A. (2015). ez: Easy analysis and visualization of factorial experiments [Computer software manual]. Retrieved from `https://CRAN.R-project.org/package=ez` (R package version 4.3)

Liesefeld, H. R., & Zimmer, H. D. (2011). The advantage of mentally rotating clockwise. *Brain and cognition*, *75* (2), 101–110.

Liesefeld, H. R., & Zimmer, H. D. (2013). Think spatial: The representation in mental rotation is nonvisual. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39* (1), 167–182.

Natale, M., Gur, R. E., & Gur, R. C. (1983). Hemispheric asymmetries in processing emotional expressions. *Neuropsychologia*, *21* (5), 555–565.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, *24* (5), 939–972.

Prinz, W. (1990). *A common coding approach to perception and action*. Springer.

R Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Rolke, B., & Hofmann, P. (2007). Temporal uncertainty degrades perceptual processing. *Psychonomic Bulletin & Review*, *14* (3), 522–526.

Schwartz, D. L., & Holton, D. L. (2000). Tool use and the effect of action on the imagination. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26* (6), 1655–1665.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171* (3972), 701–703.

Wang, D.-Y. D., Proctor, R. W., & Pick, D. F. (2003). The simon effect with wheel-rotation responses. *Journal of Motor Behavior*, *35* (3), 261–273.

Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition*, *68* (1), 77–94.

Wohlschläger, A. (2000). Visual motion priming by invisible actions. *Vision Research*, *40* , 925–930.

Wohlschläger, A., & Wohlschläger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance*, *24* (2), 397–412.

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: a meta-analysis and review. *Journal of Cognitive Neuroscience, 20*(1), 1–19.

Zacks, J. M., & Michelon, P. (2005). Transformations of visuospatial images. *Behavioral and Cognitive Neuroscience Reviews*, *4*(2), 96–118.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*(1), 3–21.

Table 1

ANOVA table for the **response time** *analysis of the* **no-change trials** *in the* **first experiment**.

factor	df	F	p	η_p^2
mental rot. dir.	1	0.69	.42	.04
tactile rot. dir.	1	0.01	.91	.00
parity	1	2.66	.12	.12
mental rot. dir.× tactile rot. dir.	1	0.47	.50	.02
mental rot. dir. × parity	1	25.96	$< .001^*$	.58
tactile rot. dir. × parity	1	2.22	.15	.10
mental rot. dir. × tactile rot. dir. × parity	1	0.08	.77	.00

Table 2

ANOVA table for the **error rate** *analysis of the* **no-change trials** *in the* **first experiment***.*

factor	df	F	p	η_p^2
mental rot. dir.	1	1.73	.20	.08
tactile rot. dir.	1	0.57	.46	.03
parity	1	9.82	.01*	.34
mental rot. dir.× tactile rot. dir.	1	0.45	.51	.02
mental rot. dir. × parity	1	6.97	.02*	.27
tactile rot. dir. × parity	1	0.14	.72	.01
mental rot. dir. × tactile rot. dir. × parity	1	0.00	.95	.00

Table 3

ANOVA table for the **response time** *analysis of the* **change trials** *in the* **first experiment***.*

factor	df	F	p	η_p^2
mental rot. dir.	1	2.08	.17	.10
tactile rot. dir.	1	0.09	.77	.00
parity	1	3.09	.10	.14
mental rot. dir. × tactile rot. dir.	1	22.74	$< .001^*$	.54
mental rot. dir. × parity	1	24.61	$< .001^*$	.56
tactile rot. dir. × parity	1	7.39	$.01^*$	.28
mental rot. dir. × tactile rot. dir. × parity	1	1.46	.24	.07

Table 4

ANOVA table for the **error rate** *analysis of the* **change trials** *in the* **first experiment**.

factor	df	F	p	η_p^2
mental rot. dir.	1	2.58	.13	.12
tactile rot. dir.	1	0.27	.61	.01
parity	1	11.13	$< .01^*$	.37
mental rot. dir. × tactile rot. dir.	1	6.39	$.02^*$	.25
mental rot. dir. × parity	1	16.36	$< .001^*$	.46
tactile rot. dir. × parity	1	4.56	$.05^*$	.19
mental rot. dir. × tactile rot. dir. × parity	1	9.19	$.01^*$	.33

Table 5

ANOVA table for the **response time** *analysis of the* **no-change trials** *in the* **second experiment***.*

factor	df	F	p	η_p^2
mapping	1	8.72	.01*	.31
mental rot. dir.	1	1.27	.28	.06
tactile rot. dir.	1	2.49	.13	.12
parity	1	3.00	.10	.14
mapping × mental rot. dir.	1	0.06	.81	.00
mapping × tactile rot. dir.	1	2.87	.11	.13
mapping × parity	1	0.02	.88	.00
mental rot. dir. × tactile rot. dir.	1	2.50	.13	.12
mental rot. dir. × parity	1	1.30	.27	.06
tactile rot. dir. × parity	1	0.03	.87	.00
mapping × mental rot. dir. × tactile rot. dir.	1	0.06	.81	.00
mapping × mental rot. dir. × parity	1	6.73	.02*	.26
mapping × tactile rot. dir. × parity	1	1.18	.29	.06
mental rot. dir. × tactile rot. dir. × parity	1	1.08	.31	.05
mapping × mental rot. dir. × tactile rot. dir. × parity	1	1.06	.32	.05

Table 6

ANOVA table for the **error rate** *analysis of the* **no-change trials** *in the* **second experiment**.

factor	df	F	p	η_p^2
mapping	1	5.89	.03*	.24
mental rot. dir.	1	6.32	.02*	.25
tactile rot. dir.	1	3.71	.07	.16
parity	1	0.00	.99	.00
mapping × mental rot. dir.	1	2.41	.14	.11
mapping × tactile rot. dir.	1	0.40	.53	.02
mapping × parity	1	0.70	.41	.04
mental rot. dir. × tactile rot. dir.	1	1.21	.29	.06
mental rot. dir. × parity	1	0.10	.75	.01
tactile rot. dir. × parity	1	0.71	.41	.04
mapping × mental rot. dir. × tactile rot. dir.	1	0.39	.54	.02
mapping × mental rot. dir. × parity	1	7.48	.01*	.28
mapping × tactile rot. dir. × parity	1	3.81	.07	.17
mental rot. dir. × tactile rot. dir. × parity	1	0.44	.52	.02
mapping × mental rot. dir. × tactile rot. dir. × parity	1	0.19	.66	.01

Table 7

ANOVA table for the **response time** *analysis of the* **change trials** *in the* **second experiment***.*

factor	df	F	p	η_p^2
mapping	1	4.93	$.04^*$	.21
mental rot. dir.	1	0.58	.46	.03
tactile rot. dir.	1	0.01	.93	.00
parity	1	0.48	.50	.02
mapping × mental rot. dir.	1	0.65	.43	.03
mapping × tactile rot. dir.	1	2.22	.15	.10
mapping × parity	1	.47	.50	.02
mental rot. dir. × tactile rot. dir.	1	17.35	$< .001^*$	.48
mental rot. dir. × parity	1	6.81	$.02^*$	.26
tactile rot. dir. × parity	1	2.47	.13	.12
mapping × mental rot. dir. × tactile rot. dir.	1	1.44	.25	.07
mapping × mental rot. dir. × parity	1	12.11	$< .01^*$	.39
mapping × tactile rot. dir. × parity	1	0.06	.81	.00
mental rot. dir. × tactile rot. dir. × parity	1	0.65	.43	.03
mapping × mental rot. dir. × tactile rot. dir. × parity	1	1.39	.25	.07

Table 8

ANOVA table for the **error rate** *analysis of the* **change trials** *in the* **second experiment***.*

factor	df	F	p	η_p^2
mapping	1	4.29	$.05^*$	.18
mental rot. dir.	1	1.86	.19	.09
tactile rot. dir.	1	2.03	.17	.10
parity	1	0.01	.93	.00
mapping × mental rot. dir.	1	6.52	$.02^*$	.26
mapping × tactile rot. dir.	1	0.49	.49	.03
mapping × parity	1	0.83	.37	.04
mental rot. dir. × tactile rot. dir.	1	3.35	.08	.15
mental rot. dir. × parity	1	3.22	.09	.15
tactile rot. dir. × parity	1	0.17	.68	.01
mapping × mental rot. dir. × tactile rot. dir.	1	0.05	.83	.00
mapping × mental rot. dir. × parity	1	10.05	$< .01^*$	.35
mapping × tactile rot. dir. × parity	1	3.62	.07	.16
mental rot. dir. × tactile rot. dir. × parity	1	1.50	.24	.07
mapping × mental rot. dir. × tactile rot. dir. × parity	1	1.09	.31	.05

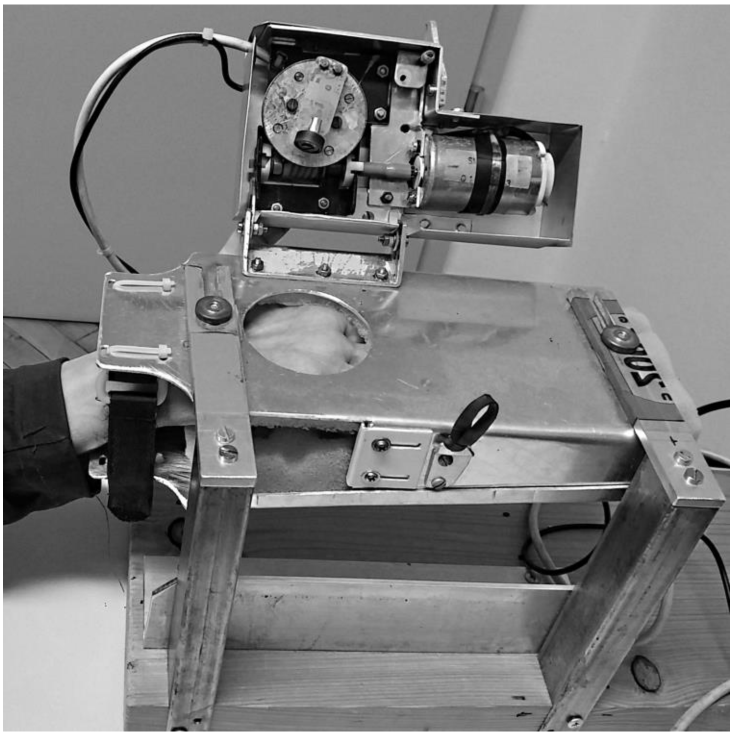


Figure 1. Tactile stimulation device. A metal wheel attached to a rotating disk propelled by two motors produced a tactile rotation on the participants left palm. Acoustic detection of changes in the rotation direction was prevented by white noise emitted from an integrated loudspeaker.

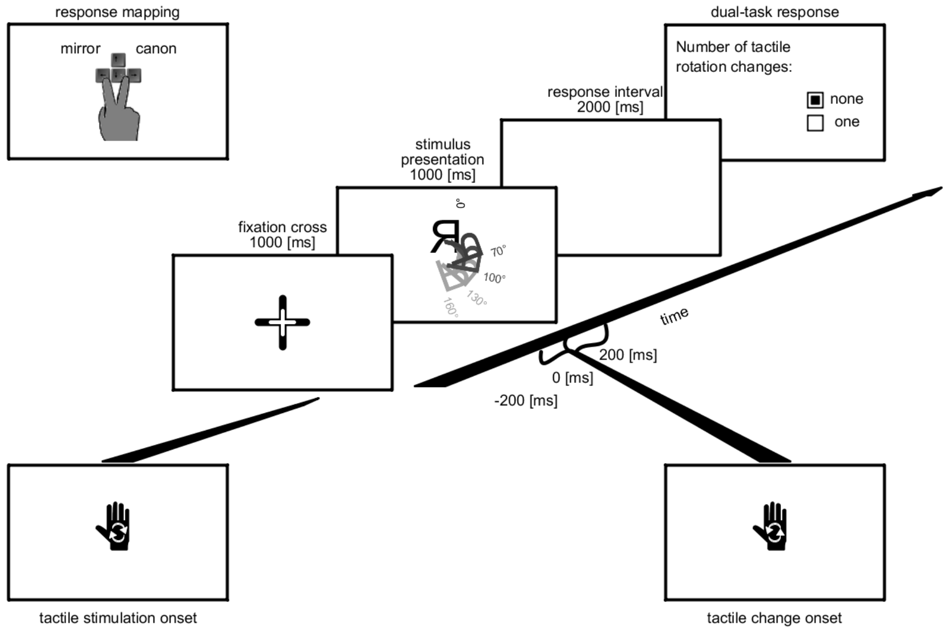


Figure 2. Schedule of a single trial: Tactile stimulation started simultaneously with the presentation of the fixation cross. After 1000 ms the target letter appeared. Participants had to judge the parity of the letter (i.e. canonical or mirrored). Please note that the letter always appeared in the center of the screen, the slight offset between the shown disparities was only included for the sake of visibility. The letter disappeared after 1000 ms, followed by a blank interval of 2000 ms or until the participant responded. After the parity judgment, the response screen for the tactile secondary task appeared. Tactile stimulation was always applied to the left hand of the participants and responses were given with the right hand. The response mapping for the parity judgment is shown in the upper left. In the second experiment, this mapping varied between blocks.

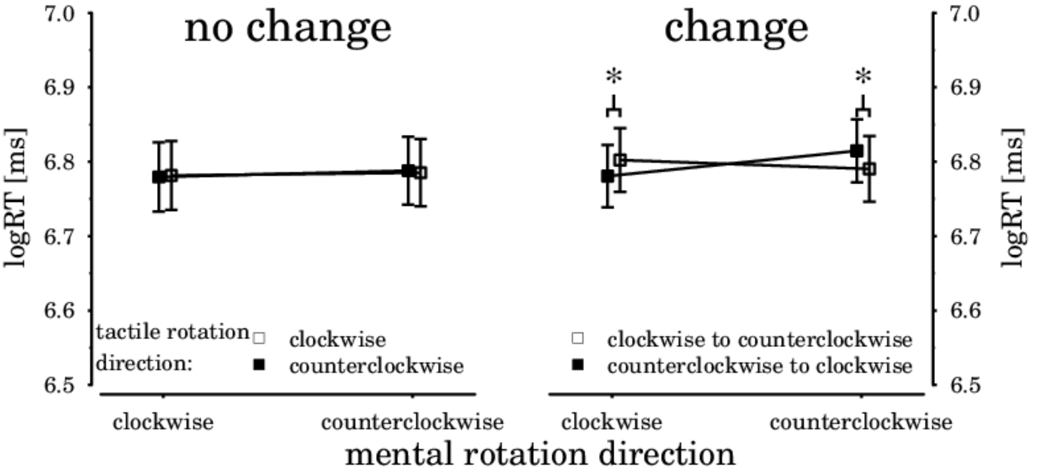


Figure 3. Interaction between **mental and tactile rotation direction** in the first experiment. The left panel shows results for no-change trials (tactile stimulation direction stayed the same), the right panel shows results for change trials (tactile stimulation direction changed). In change trials, participants responded faster when the tactile rotation direction after the change matched the mental rotation direction. In no-change trials no such compatibility was observed. Asterisks indicate significant differences with $\alpha = .05$, p-values were adjusted for multiple comparisons. Error bars indicate the standard error of the mean.

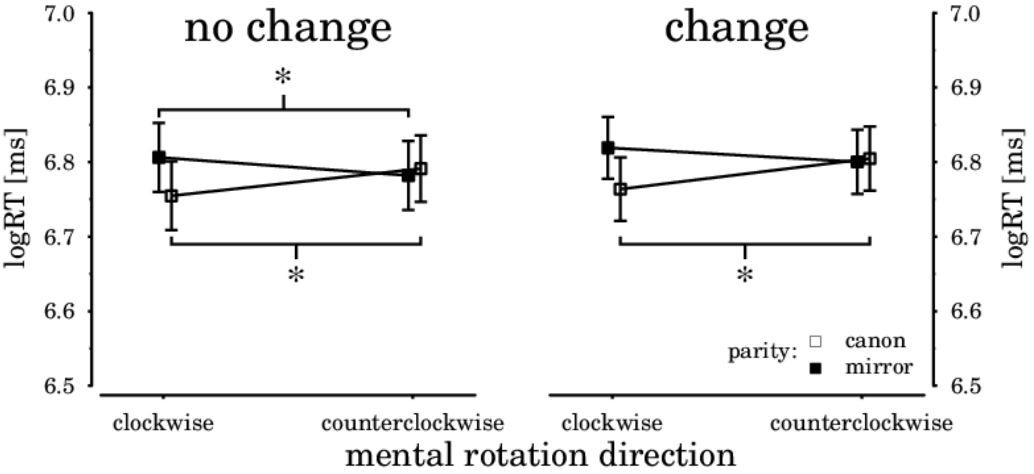


Figure 4. Interaction between **stimulus parity and mental rotation direction** in the first experiment. The left panel shows results for no-change trials (tactile stimulation direction stayed the same). The right panel shows results for change trials (tactile stimulation direction changed). For both change and no-change, participants responded faster to canonical letters when they had to mentally rotate them clockwise (to the right) compared to when they had to mentally rotate them counterclockwise (to the left). For mirrored letters the opposite was true. Asterisks indicate significant differences with $\alpha = .05$, p-values were adjusted for multiple comparisons. Error bars indicate the standard error of the mean.

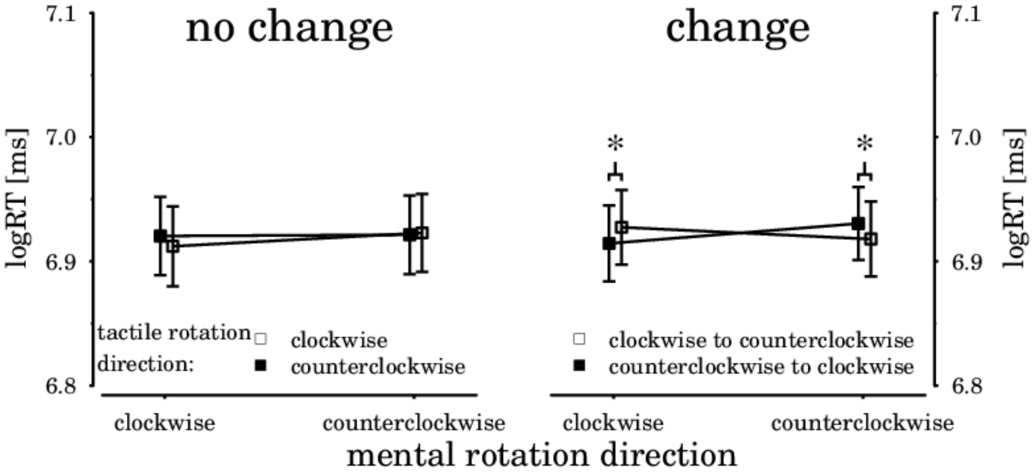


Figure 5. Interaction between **mental and tactile rotation direction** in the second experiment. The left panel shows results for no-change trials (tactile stimulation direction stayed the same), the right panel shows results for change trials (tactile stimulation direction changed). Again, a significant interaction way only observed in change trials. Participants responded faster when the tactile rotation direction after the change matched the mental rotation direction. Asterisks indicate significant differences with $\alpha = .05$, p-values were adjusted for multiple comparisons. Error bars indicate the standard error of the mean.

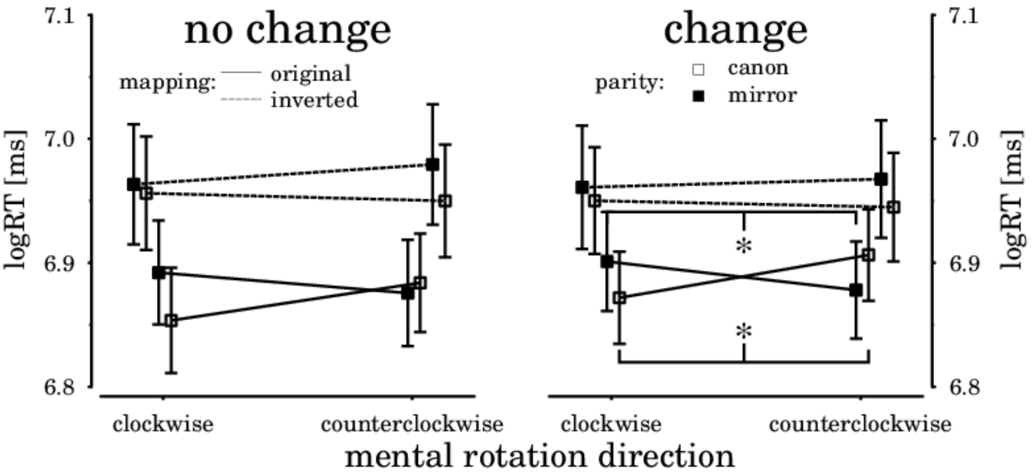


Figure 6. Interaction between **mapping, parity and mental rotation direction** in the second experiment. The left panel shows results for no-change trials (tactile stimulation direction stayed the same), the right panel shows results for change trials (tactile stimulation direction changed). Response times are elevated when mirrored stimuli have to be rotated clockwise (to the right), but only when participants had to respond with the left key. For canonical stimuli the opposite is true when participants had to respond with the right key. For the inverted mapping, the opposite is true. The respective differences only reached significance for the original mapping (left key press for mirrored letters) in change trials. Asterisks indicate significant differences with $\alpha = .05$, p-values were adjusted for multiple comparisons. Error bars indicate the standard error of the mean.

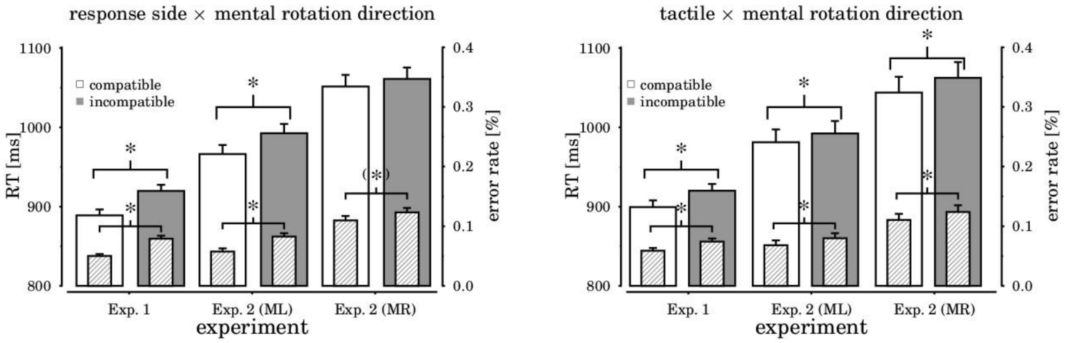


Figure 7. Compatibility effects between mental rotation direction and response side (left panel) and mental rotation direction and tactile rotation direction (right panel), across experiments. Responses with the right key were considered to be compatible with clockwise mental rotations and responses with the left key were considered to be compatible with counterclockwise mental rotations. Data of the second experiment is presented separately for the two response mappings. *ML* refers to the original mapping where participants responded to mirrored letters with the left arrow key, *MR* refers to the inverted mapping. Response times are visualized as bars, the nested, smaller bars indicate error rates. Error bars indicate the standard error of the mean. The numeric magnitude of the compatibility effects is similar in all experiments, the overall increased response time in the second experiment is due to the more extreme disparities. Please note that data in the left panel combines change as well as no-change trials, while for the data in the right panel only trials where the tactile stimulation direction changed were considered.