

Visuelle und semantische Größeninformationen in Wahrnehmung und Handlung

Dissertation zur Erlangung des Doktorgrades (Dr.rer.nat.) an der Universität Hamburg,

Fachbereich Psychologie und Bewegungswissenschaften,

Institut für Psychologie.

Vorgelegt von Karl Kunibert Kopiske,

Hamburg, Februar 2017.

Disputationsvorsitzender: Prof. Dr. Mike Wendt

1. Dissertationsgutachter und Betreuer: Prof. Dr. Volker Franz

2. Dissertationsgutachter: Prof. Dr. Martin Spieß

1. Disputationsgutachter: Prof. Dr. Ulf Liszkowski

2. Disputationsgutachter: Prof Dr. Lars Schwabe

Disputationstermin: 1.6.2017

Abstract

Visual information is vital to interact with our environment, and can give us information about our surroundings before we begin interacting with it, as well as during interaction. This visual information is often used to update what we know about an object, which may be called semantic information. Thus, it seems plausible that the two might often interact. The goal of this thesis is to give a small glimpse of how visual and semantic information of size and magnitude are processed, and when they may interact. For this thesis, several experiments were conducted to investigate whether (a) visual processing differs depending on whether the task at hand involves a direct, skilled action or conscious perception, and (b) to what degree visual features of magnitude representations matter for their processing. Study 1 examined the effect of a well-studied visual illusion of size (the Ebbinghaus illusion) on perception, as well as on grasping. Whether visual illusions, and the Ebbinghaus illusion in particular, have different effect on perception and action, respectively, has been a topic of debate for a long time, and it is a key line of evidence in the overarching debate about whether perception and action process visual size differently. To come as close as possible to settling the debate, we preregistered and conducted a large-scale experiment in four different laboratories. Controlling for potential confounds, such as context circles being perceived as obstacles, we found no difference between the Ebbinghaus illusion's effect on action and its effect on perception. The implications of this study have been discussed in two published commentaries. I discuss these commentaries – a critique of the generalizability of our experimental design, and our rebuttal – in detail, and make the case that in a confirmatory experiment like ours, what matters is how strongly predictions are made by a theory, and how well these are tested by the experiment. Under both these criteria, our study clearly presents a very strong piece of evidence. We also conducted two studies on the 'mental number line', investigating two specific questions on how numeric processing is influenced by visual stimulus properties. Study 2 investigated whether Chinese representations of number would evoke a space-response association (the so-called SNARC effect) known to occur when participants respond quickly to numeric stimuli. This effect describes a phenomenon that responses tend to be faster when the location of a response is "congruent" with its location. Research suggests that this congruency effect depends largely on reading habits, such that typically, large numbers on the right and small numbers on the left display a response time advantage in European participants. The effect is thought to be

generally independent of the modality of the stimuli, yet at the same time susceptible to spatial stimulus features. We also know of at least one study that did not find a horizontal SNARC effect in Chinese characters. Thus, we tested whether we would find a horizontal SNARC effect in participants from Mainland China, in Arabic digits, Chinese characters, and Chinese hand signs. We found a robust horizontal SNARC effect, corroborating the notion that the effect is independent of notation. This is commonly taken as one piece of evidence for the existence of an analog internal representation of magnitude on a so-called ‘mental number line’. Apart from being responsible for congruency effects between space and number, this mental number line has also been proposed to be compressed, such that nonsymbolic or nonverbal representations of magnitude tend to systematically underestimate differences, and larger magnitudes in general. Study 3 was designed to test whether the proposed compression of the mental number line can truly be ascribed to properties of magnitude processing, or whether it is caused by the way such representations are typically measured. To this end, we had participants complete a classic typed (verbal) magnitude estimation task, as well as a nonsymbolic magnitude estimation task that consisted of estimating the correct location of a stimulus on a ruler-like response bar. We found a robust nonlinearity and underestimation in both tasks that was not caused by task demands, and that was in fact even resistant to veridical feedback, showing that these properties should be considered a property of the processing and representation of nonsymbolic magnitude. Finally, I discuss what these results mean for our understanding of how visual and semantic magnitude information is processed by human observers.

Acknowledgements

First and foremost, I want to thank my supervisor, Volker Franz, for being a true mentor and for teaching me more than I had hoped to learn. Thanks also to all colleagues in Hamburg for all sorts of help and support, and for making my time at the department into what it was: Barbara, Melanie, Denise, Larissa, Owino, Martin, Alexander, Alistair, Martin, and in particular Rachel (who helped make my manuscripts readable), Christian (who helped me understand grasping), and Kao (who helped keep me sane). Thanks to my coauthors for their patience, and to Fulvio Domini for even more patience. Finally, thanks to friends and family for always being interested in my work, but also providing much-needed distraction. Especially my parents, Helga and Karsten, as well as Simon, Thomas, Steph, Pan, Sabine, and the Brodersens. I was (and am) supported by a grant of near-infinite understanding and encouragement from Yvonne Kopsike.

Index

Abstract3

Chapter 1 – Introduction7

Chapter 2 – Visual processing of object size and the two-visual-systems-hypothesis (study 1)...9

Chapter 3 – Nonsymbolic magnitude on the mental number line (studies 2 and 3).....45

Chapter 4 – Summary64

References67

List of published articles88

Appendix: Published articles89

Chapter 1 – Introduction

Size is an important property of any object in our environment, relevant for almost all interaction with the object. It is also a property that, unlike many others, can often be estimated quite accurately from visual information alone. In intentional interaction with objects, the first sensory information we have is quite often visual, so that our exploration through other senses, like touch, or perhaps smell, can be informed by the first impression we have gained through vision. With vision being such a vital source of information to plan and execute our actions, a prominent theory on the visual system proposes that the processing of visual information, for which we have two anatomically distinct cortical streams, differs functionally depending on its purpose. That is, depending on whether visual information is used to guide action, or to gather information for conscious perception, it is processed in the dorsal vision-for-action or the ventral vision-for-perception stream (Figure 1), respectively (this is the two-visual-systems-hypothesis, or TVSH; Goodale & Milner, 1992; Milner & Goodale, 1995, 2006). According to this theory, these streams have vastly different properties in many regards, one of them being how they compute the size of an object. In our first experiment (described in chapter 2), we tested one of the theory's prominently discussed predictions on this by investigating the effect of a visual illusion of size on action, thus contributing to a current debate about the veracity of a theory that, if true, would have great consequences for almost all research on vision. Discussing two commentaries on this study, we consider what evidence is available and what evidence is appropriate regarding this prediction, as well as its relevance for the TVSH overall.

In the semantic domain (which for the purpose of this thesis we take to mean: Knowledge about an object or stimulus that is not acquired through direct sensory processes), size or magnitude is typically studied with relation to numbers. These are, after all, the main part of our vocabulary to transmit information about magnitude to others, and the most common ‘symbolic’ representation of magnitude. Number and magnitude are so closely linked that indeed, multiple theories have attempted to integrate what we know about the processing of number, size, and sometimes many other dimension that can be expressed on at least an ordinal scale into one framework of ‘magnitude’ (Dehaene, 1992; Proctor & Cho, 2006; Walsh, 2003) not unlike sensory magnitude (Stevens, 1946, 1957; Teghtsoonian, 1971). In another study (described in chapter 3), we investigated whether a well-known association between numbers and space (the spatial-numerical association of response codes, or SNARC effect; Dehaene, Bossini, & Giraux, 1993) would be found in Chinese numerals that contain multiple – and in some cases conflicting – cues about numeric magnitude. Finally, we tested whether an observed nonlinearity in responses to nonsymbolically presented numeric magnitude could be better conceptualized as a phenomenon of magnitude representation, of perceptual uncertainty, or an artefact of the response mode (Figure 8).

More generally, the goal of this dissertation is to probe specific theories that posit an interaction of visual size processing with other stimulus or task features. In three studies, we examined and extended findings on how size or magnitude properties of stimuli are extracted visually, in two prominent examples where potential sources of interference – visual illusion and semantic information – have informed theories on the general processing of this size and magnitude information.

Chapter 2 – Visual processing of object size and the two-visual-systems-hypothesis

(study 1)

„Und Sie sind auch nicht wirklich so riesengroß, wenn Sie weit entfernt sind, sondern es sieht nur so aus?“
„Sehr richtig“, antwortete Herr Tur Tur. „Deshalb sagte ich, ich bin ein Scheinriese. Genauso, wie man die anderen Menschen Scheinzwerge nennen könnte, weil sie ja von weitem wie Zwerge aussehen, obwohl sie es gar nicht sind.“ – Michael Ende (1960), Jim Knopf und Lukas der Lokomotivführer.

The size of objects in our environment is a basic piece of information that our visual system has to process to facilitate interaction with our environment. It is also a property that is easy to measure and to manipulate, and as such a useful manipulation in research on perception.

According to an influential theory on visual processing, however, visual information is processed differently depending on the task it is needed for (Milner & Goodale, 1995, 2006). Processing of size information is of special importance for this theory, as it concerns a central prediction, as well as key pieces of evidence for the theory, as I will describe in the following sections.

2.1 The TVSH: Definition and evidence

In the literature on perception and action, a prominent theory proposed by Goodale and Milner (1992; Milner & Goodale, 1995, 2006) states two distinct streams of processing: A ventral ‘vision-for-perception’ stream from the primary visual cortex (V1) to higher visual areas in the inferotemporal lobe, functionally associated with conscious perception, and a dorsal ‘vision-for-action’ stream to process action-relevant visual information that leads from V1 to the posterior parietal cortex. The neuroanatomical properties of the dorsal and ventral stream were in fact described even before Goodale and Milner’s (1992) original formulation of their theory, for example by Ungerleider and Mishkin (1982), but with a different functional interpretation, as

Ungerleider and Mishkin had called the dorsal and ventral stream the ‘where’ and ‘what’ stream, supposedly responsible for object localisation and recognition, respectively. Indeed, the functional interpretation of the two streams by Milner and Goodale (1995, 2006) is rather similar to what had been called the ‘cognitive’ and ‘motor’ aspects of vision by Bridgeman and colleagues (Bridgeman, Kirch, & Sperling, 1981). Despite other theories like that of Ungerleider and Mishkin (1982) or Bridgeman et al. (1981), Milner and Goodale’s (1995, 2006) theory is often (and in this thesis) simply referred to as ‘the two-visual-systems-hypothesis’ (TVSH), due to its great impact on the field of human vision especially in recent years. It should be mentioned that the name does not imply that both streams necessarily have to be completely independent: There are interconnections between the two streams that can under some circumstances lead to information being transferred despite the separate processing (Goodale, 2008, 2014; Goodale & Milner, 2010). Indeed, the degree of interconnectivity between the two streams has been the target of many studies and discussions over the years (e.g., Goodale, 2008; Goodale & Milner, 2010; Schenk, Franz, & Bruno, 2011; Schenk & McIntosh, 2010; Westwood & Goodale, 2011), so that some authors (e.g., Bruno & Franz, 2009; Schenk et al., 2011) have argued that it may be useful to distinguish between a ‘strong’ TVSH (with little or no interaction between the dorsal and ventral stream) and a ‘weak’ TVSH (that assumes a substantial amount of interaction).

A key feature of the functional interpretation of the TVSH is that, unlike in the theory of Ungerleider and Mishkin (1982), the two streams are not thought to process different properties of a scene, but each process the entire scene (or at least all relevant information) to either form a stable percept (ventral stream) or process action-relevant information (dorsal stream). This means, among other things, that all phenomena of conscious perception processed in the ventral

stream – including perceptual biases, or interaction with other sources of information – may be completely irrelevant in action tasks. If visual processing for action is indeed separate as proposed by the TVSH, one could thus argue that a large portion of known effects in visual perception would have to be investigated twice over.

According to the TVSH, the two streams of visual processing are distinct with regards to their purpose, but also numerous other properties, as summarised in Table 1. In part, these are direct consequences of the neuroanatomy of the streams; for example, the fact that the two streams differ with regards to the speed of processing is a direct corollary of the fact that fast, magno-cellular connections exist from the thalamus to the dorsal stream, while connections to the ventral stream are mainly comprised of slow, parvo-cellular neurons (Hubel & Wiesel, 1972; Livingstone & Hubel, 1987; Milner & Goodale, 1995, 2006). Other properties are not obvious a priori and have their basis in empirical evidence, including the function of the two streams, as well as the proposed analytic (as opposed to holistic) mode of processing in the dorsal stream. Hence, such properties are often tested experimentally to test the validity of the evidence and the predictions of the TVSH.

Table 1: Properties where the dorsal and ventral stream differ according to the TVSH.

Property	Dorsal	Ventral	Reference
<i>Function</i>	Skilled actions, online correction	Conscious, stable percepts	Goodale & Milner, 1992; Milner & Goodale, 1995, 2006
<i>Processing speed</i>	Fast	Slow	Milner & Goodale, 1995, 2006
<i>Decay of representation</i>	Fast	Slow	Westwood & Goodale, 2003
<i>Mode of processing</i>	Holistic	Analytic	Aglioti, DeSouza, & Goodale, 1995
<i>Consciousness</i>	Conscious	Unconscious	Goodale, Milner, Jakobson, & Carey, 1991

Note: Compiled according to Milner and Goodale (2006, 2008), Goodale (2014). See also Schenk and McIntosh (2010) for a similar summary.

2.1.1 Evidence for the TVSH: Patient studies and neuroimaging

The first evidence for the TVSH came from patient studies. Indeed, it was also inspired by results from neuropsychological patient, as for example cortically blind patients can in rare case exhibit what is called ‘blindsight’ (Weiskrantz, 1990); that is, they may be able to adapt their actions to visual input that they are not consciously aware of (Striemer, Chapman, & Goodale, 2009). More concrete evidence for the functional dissociation of the two streams was reported by Goodale and colleagues, who tested a patient with visual form-agnosia (patient DF; Goodale et al., 1991) who was able to accurately grasp, but not perceive objects of different shapes and sizes, as well as a patient with optic ataxia, who in turn could recognise sizes and shapes, but could not scale her grip accordingly (patient RV; Goodale et al., 1994). This is commonly called a ‘double dissociation’: One manipulation, or in this case brain lesion, affects performance in task A, but not in task B; this is a single dissociation. In a double dissociation, another manipulation (or

lesion) affects performance in task B, but not task A. This is traditionally seen as strong evidence for two tasks (in this case: Perception and action) relying on different resources or brain areas (Goodale et al., 1991; Teuber, 1955, see especially p. 238). More recently, it has been noted that the difference in the required resources can in theory be infinitesimally small (Chater, 2003) – however, double dissociations remain one of the most compelling pieces of evidence in neuropsychology.

The neurological damage in patients DF and RV fit the predictions of the TVSH: Occipitotemporal lesions in the dorsal stream for RV (Goodale et al., 1994) and large occipitoparietal lesions in the ventral stream of DF (Goodale et al., 1991). In DF's case, further evidence was obtained from functional neuroimaging, as an fMRI study reported that the damaged areas (especially lateral occipital regions) corresponded well with areas that were selectively activated in healthy participants during perceptual tasks, and were not activate during these tasks in DF (James, Culham, Humphrey, Milner, & Goodale, 2003). Further fMRI studies in healthy patients investigated if classic perception and action task could be mapped to ventral and dorsal stream, respectively. Indeed, occipitoparietal areas consistent with RV's lesions (Goodale et al., 1994) where found to be selectively activated during reaching and grasping of an object, with further activation in the anterior IPS selective to grasping (Cavina-Pratesi et al., 2010). At the same time, some occipitotemporal areas (which are damaged in DF; James et al., 2003) have been demonstrated to be active specifically during object recognition (Grill-Spector, 2003).

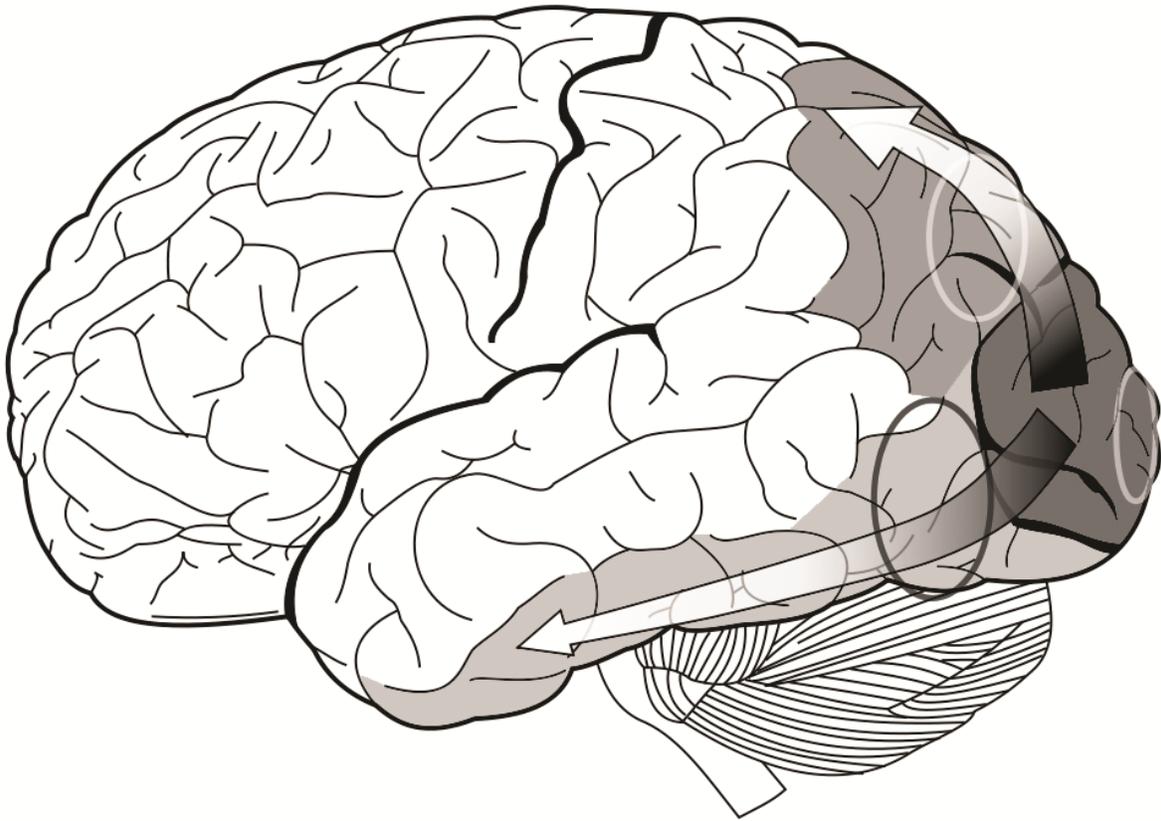


Figure 1: The two visual streams, and lesions of DF and RV. Highlighted are V1 (dark gray), dorsal stream (medium gray), and ventral stream (light gray). White ellipses indicate lesions found in patient RV (see Goodale et al., 1994), black ellipses show bilateral lesions found in patient DF (see James et al., 2003). Image of the brain and the two streams taken from wikipedia (https://en.wikipedia.org/wiki/Two-streams_hypothesis#/media/File:Ventral-dorsal_streams.svg, created by user *Selket*) and used under CC BY-SA 3.0 license.

2.1.2 Evidence for the TVSH: Behavioural studies with healthy participants

Dissociations between perception and action have also been reported in healthy participants.

Arguably the most influential studies concern grasping of visual illusions: The first piece of evidence for the TVSH from healthy participants was reported by Aglioti, deSouza, and Goodale (1995), who reported that the Ebbinghaus illusion (figure 2b), in which a central circle looks bigger or smaller depending on the size of circles surrounding it, does not affect the maximum in-flight grip aperture (MGA) measured between the thumb and the index finger during grasping.

The MGA is a frequently used indirect measure of perceived object size during motion planning. It is highly correlated with object size in so-called ‘precision grip’ reach-to-grasp movements (Jeannerod, 1984, 1986; Smeets & Brenner, 1999), that is, movements in which participants move their hand towards an object and pick it up with their thumb and index finger. Similar results to those of Aglioti, et al. (1995) have been obtained with the Ponzo illusion (Ganel, Tanzer, & Goodale, 2008; figure 2a), the Mueller-Lyer illusion (Thompson & Westwood, 2007; figure 2c) and the empty-space illusion (Stöttinger et al., 2012; figure 2d), which are all perceptual size illusions that have been shown not to influence MGA.

Besides distorted size information, other phenomena of conscious perception have been studied and not found in grasping. Two prominent examples are Weber’s law (which states that the just noticeable difference between two stimuli is proportional to the intensity of the stimuli; Fechner, 1860) and Garner’s interference (interference of task-irrelevant stimulus dimensions; Garner, 1976). A study by Ganel, Chajut, and Algom (2008) showed that Weber’s law could not be found in the MGA in a simple grasping task, although it was to be found in a perceptual task with similar task demands, manual estimation (ME) with the thumb and index finger. For Garner’s interference, Ganel and Goodale (2003) reported that irrelevant object dimensions affected response time in perceptual tasks (consistent with the original effect found by Garner, 1976, in speeded classification), but not in response time or any other dependent variable in grasping. Finally, other action tasks than grasping have been used to investigate dissociations between perception and action. For example, saccades – quick, discontinuous eye movements – have been found to affect participants’ ability to consciously perceive a spontaneous movement of an object,

while at the same time, grasping was corrected online to adapt to the movement (Goodale, Pelisson, & Prablanc, 1986).

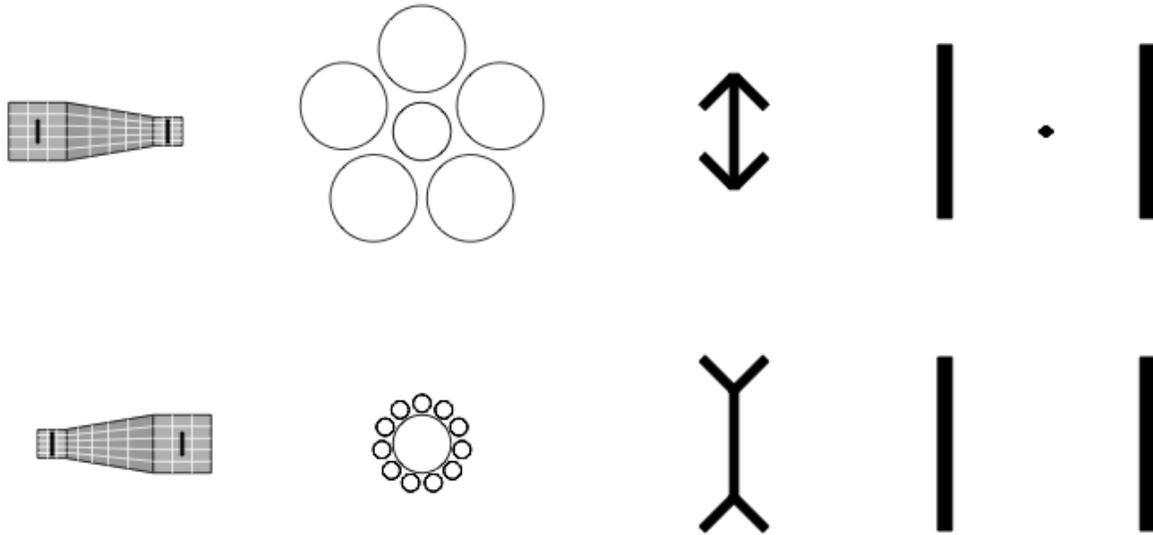


Figure 2: Examples of illusions where a perception-action dissociation has been reported. **a:** Ponzo illusion (see Ganel, Tanzer, et al., 2008). All four black rods are of equal size. **b:** Ebbinghaus illusion, classic (*LF* and *SM*) configurations (see Aglioti et al., 1995). Both central circles are of equal diameter. **c:** Mueller-Lyer Illusion (see Westwood, Heath, & Roy, 2000), both vertical lines are of equal length. **d:** Empty-space Illusion (see Stöttinger et al., 2012), in both cases the space between the two rods is equal. It tends to appear smaller when a dot is present in the middle. This is the only illusion that may get weaker when both version are presented side-by-side.

2.1.3 Criticism and doubts about the TVSH

Over the years, a substantial number of studies have reported evidence for the TVSH. However, both the interpretation and the strength of the accumulated evidence have been questioned. This concerns both patient studies and studies with healthy participants and has led to numerous modifications and additions to the original theory, and even caused some researchers to question

the TVSH's veracity altogether. I will summarise the main points of criticism below, starting with patient studies and the moving on to studies of healthy participants.

The most important results from patient studies came from studying visual form-agnosia patient DF (e.g., Goodale et al., 1991; Milner, Ganel, & Goodale, 2012; Whitwell & Buckingham, 2013; Whitwell, Milner, Cavina-Pratesi, Byrne, & Goodale, 2014; Whitwell, Milner, & Goodale, 2014), in which DF was unable to solve perceptual tasks, but performed almost as well as control participants on grasping tasks. As Schenk (2006) argued, however, it does not follow from this that her visual processing is the cause of this selective deficit. It is just as plausible to interpret DF's performance as a deficit of allocentric information, as evidenced by the facts that (a) her performance in action tasks deteriorated when a motion had to be planned relative to an external cue (Schenk, 2006, figure 2d), and (b) her perceptual judgement was relatively intact when judging distances relative to her hand (Schenk, 2006, figure 2b). Schenk further proposed that DF's preserved ability to scale her grip could be due to the fact that non-visual information, such as haptic feedback, might be used in grasping to compensate for her visual deficit (Schenk, 2010, 2012). Additionally, the claim that DF's grasping is unimpaired (Goodale et al., 1991) has also recently been scrutinised and found to be questionable (Himmelbach, Boehme, & Karnath, 2012), and her dorsal stream might not be as intact as had been assumed (Hesse, Ball, & Schenk, 2014).

Doubts have also been raised about the evidence from other neurological disorders, such as optic ataxia. Goodale et al. (1994) interpreted patient RV's deficit in controlling visually guided motion while at the same time being able to perform accurate perceptual judgements as an effect of damage to her dorsal stream. However, Himmelbach and Karnath (2005) reported that optic

ataxia patients tend to struggle with executing delayed grasping movements, which is a ventrally-controlled type of action that, according to Milner and Goodale (2006), as the dorsal stream can only store information for very short periods. In addition, Pisella and colleagues (Pisella, Binkofski, Lasek, Toni, & Rossetti, 2006), proposed that optic ataxia may be best understood as more than a pure deficit of motor control, and that the claim that the ventral and the dorsal stream are anatomically almost entirely separate may be exaggerated. More recently, it has also been proposed that the evidence of normal ‘action’ performance in blindsight patients (Striemer et al., 2009) has been overstated (A. I. Ross, Schenk, Billino, Macleod, & Hesse, 2016).

Evidence from non-clinical samples has also been scrutinised and in large part not confirmed. For example, the interpretation of Ganel and Goodale’s (2003) and Ganel et al.’s (2008) results that Weber’s law and Garner’s interference are not present in grasping as a property of the dorsal stream has been criticised due to the fact that both phenomena can also not be found in ventrally controlled action tasks (Eloka, Feuerhake, Janczyk, & Franz, 2015; Löwenkamp, Gärtner, Haus, & Franz, 2015). Thus, these effects can be more parsimoniously explained as properties of grasping (in the case of Weber’s law possibly brought on by motor constraints; Bruno, Uccelli, Viviani, & De’Sperati, 2016; Utz, Hesse, Aschenneller, & Schenk, 2015), not visual processing.

Another interesting study testing the proposed action-perception dissociation was conducted by McIntosh and Lashley (2008). They asked participants to grasp different match match-boxes of two kinds that are well known in Great Britain (and among participants in their sample) and typically have very different sizes. Thus, participants had prior knowledge about the stimuli. This influenced the MGA, as McIntosh and Lashley (2008) showed, and exactly in the predicted

direction: The boxes that would typically be larger caused a larger MGA, while the ones that would be smaller caused a smaller MGA. This is also incompatible with the TVSH, since the proposed analytic processing of object size in the dorsal stream should be impervious to external factors like prior knowledge.

Finally, the literature on visual illusions also offers much less support for the TVSH than originally proposed. The original study by Aglioti et al. (1995) was followed by a number of other studies where illusions were grasped and perceptually appraised, several of which did not find any difference between perception and action (e.g., Franz, Gegenfurtner, Bühlhoff, & Fehle, 2000; Pavani, Boscagli, Benvenuti, Rabuffetti, & Farnè, 1999). Indeed, there were illusion effects on MGA even in the original study (Aglioti et al., 1995). These were smaller than in the perceptual task, but this is not surprising for several reasons. For one, MGA is not as responsive to a change in physical size as some perceptual measures are (Franz, 2003; Smeets & Brenner, 1999), which would lead us to expect that they would also respond less to illusory distortions. Indeed, if illusion effects are scaled by the response function of the output measure, results from the most prominently studied illusion, the Ebbinghaus illusion, are consistent across tasks (Franz & Gegenfurtner, 2008). Similar results were reported in a meta-analysis of the Mueller-Lyer illusion (Bruno & Franz, 2009). It is also unclear if the display that was chosen by Aglioti et al. (1995), where two Ebbinghaus figures were presented side-by-side and participants compared the central circles, lends itself well to a comparison between perception and action, as grasping

requires participants to operate on only one of the two displays when planning their movements (Franz et al., 2000).

As an explanation for the effects of illusion displays on MGA, Haffenden and Goodale (2000) proposed, that these may indeed not be actual illusion effects in the traditional sense of changing the response through changing the perceived size, but a consequence of the unfortunate placement of the context circles. In the traditional Ebbinghaus illusion, the small context circles are placed closer to the central circle than the large context, as it is well known that a larger distance of the context elements makes the central circle look slightly smaller (Girgus, Coren, & Agdern, 1972; Roberts, Harris, & Yates, 2005), so this configuration maximises the illusion effect, but also inadvertently introduces a confound. This notion has been tested in two studies by matching the distance of the context circles (Franz, Bühlhoff, & Fahle, 2003; Haffenden, Schiff, & Goodale, 2001), but with opposite results, which means that the debate on the obstacle avoidance hypothesis is still open. Indeed, obstacle avoidance has recently been proposed as a mechanism that might also explain effects of the Ponzo illusion on MGA (Whitwell, Buckingham, Enns, Chouinard, & Goodale, 2016).

2.1.4 Alternative theories and counter-arguments

Several theories have been used to explain the mixed results in grasping visual illusions, although they were not necessarily formulated based on these findings. One is the ‘double pointing’ hypothesis by Smeets and Brenner (1999). Unlike classic theories of grasping (Jeannerod, 1984, 1986; Woodworth, 1899), this theory does not assume that the typically studied ‘precision grip’ of an object is a combination of the two components reaching (transport to the

object) and grasping (closing thumb and index finger around the object), but rather as transport combined with both the thumb and the index finger independently pointing at appropriate contact points on the object. This theory can explain the data on grasping visual illusions quite nicely, as it does not assume size processing to be relevant for grasping at all (Smeets & Brenner, 1999, 2006), so that the lack of an effect is readily explained, but it is still possible that context elements may alter the selection of contact points in such a way that it may affect the MGA. Thus, the theory can only be tested on other predictions. One such prediction concerns the timing of the MGA: According to this model, the MGA might be the same for objects that are illusorily larger and those that are physically larger, but the so-called ‘approach parameter’ should differ. This would cause the MGA to occur relatively earlier during the movement of objects that appear larger through an illusion, but later when an object is physically larger (Smeets, Glover, & Brenner, 2003). This has not been definitively refuted or confirmed, however, as the predicted effects are very small (de Grave, Biegstraaten, Smeets, & Brenner, 2005).

Another theory that could explain the different illusion effects is the so called ‘planning-control model’ (Glover, 2004; Glover & Dixon, 2002a, 2002b). This model separates grasping movements conceptually into the planning phase and the control phase, with each phase using different information. In this framework, different sizes of illusion effects can be explained as due to the fact that the illusion effect is created during the planning phase and decrease during the control phase (Glover & Dixon, 2002a), so that small technical deviations in the experimental protocol could affect the effect dramatically. However, evidence for decreasing illusion effects over the trajectory of a grasp is weak, and it is unclear whether this is anything more than an

artefact of the fact that grasping trajectories are simply generally more responsive during later phases (Franz, 2004; Franz, Scharnowski, & Gegenfurtner, 2005).

Another simple alternative to the TVSH that does not assume additional factors in grasping making the MGA less informative for the task at hand was proposed by Schenk and McIntosh (2010): Anatomically, the existence of the two streams is uncontroversial, as is the existence of connection between them. Since solid evidence for a functional dissociation is limited to very few instances, they suggest that the most sensible assumption is a single system with two relatively specialised streams (Schenk & McIntosh, 2010). Somewhat more specifically related to size processing, Franz and colleagues (Franz & Gegenfurtner, 2008; Franz et al., 2000; see also fig. 2 in Franz, Fahle, Bühlhoff, & Gegenfurtner, 2001) suggested a similar model that proposes anatomically distinct streams with a common size representation.

A frequent reply to criticism by proponents of the TVSH is to point out that the emphasis should be on cases where a dissociation can be demonstrated (see e.g. Goodale, 2008, 2014; Westwood & Goodale, 2011). As long as these exist, they argue, this speaks for the TVSH, as actions may be controlled ventrally for a number of reasons, thus no dissociation should be expected. Further, the theory states that processing differs only after V1 – consequently, illusion occurring before the split in V1 could be seen in action just like in perception (Milner & Dyde, 2003). Additionally, Goodale (2008) acknowledged that some effects may indeed be created in the ventral stream but still affect dorsal actions through interconnections; in this case, however, the effect would be weaker than in perception. Again, the emphasis is on the remaining differences between perception and action, not the similarities.

2.1.5 Open questions and suitable tests

The debate about the extent to which the evidence supports the TVSH is ongoing, especially with regards to grasping visual illusions. This leads us to the question what a severe test of TVSH predictions would be. Focussing on the prediction that contextual illusions that distort size perception should show no or only small effects on MGA, the central issues are matching task demands between perceptual tasks and grasping, and matching dependent variables with regards to their behaviour following stimulus changes. To do this, slope correction (i.e., correction for the slope of the response function of a measure) is an important instrument. Without it, illusion effects vary wildly, not just between grasping and perception, but also between perceptual tasks (see Franz & Gegenfurtner, 2008, figure 1e). At the same time, the procedure has been criticised, as the correlation between object size and MGA is not perfect and may not be completely linear, so a linear correction may be inaccurate (Westwood & Goodale, 2011, p. 807). Hence, it has been suggested that it is preferable to design experiments in such a way that no slope correction is needed (Goodale, 2014). This has been done in three prominent studies: Aglioti et al. (1995) and Haffenden and Goodale (1998) with the Ebbinghaus illusion, and Ganel et al. (2008) with the Ponzo illusion. In these studies, the illusion effect was used to create physically different, but perceptually equally-sized stimuli by embedding two different objects in different illusory contexts. In this case, any difference in MGA can be interpreted as a dissociation, as there should be no difference in perceived size. Indeed, Ganel et al. (2008) took this setup one step further:

Here, two objects were chosen such that the illusion made the physically object look smaller than the physically smaller one. Indeed, the physically smaller object led to a smaller MGA (Figure 3).

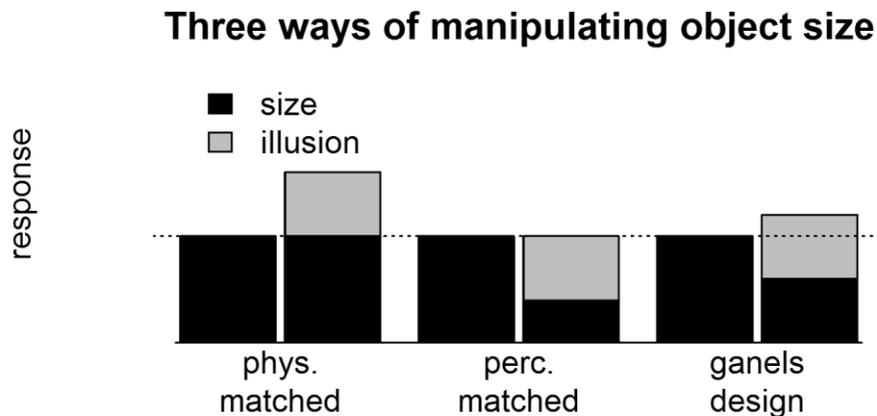


Figure 3: Approaches to manipulating physical and perceived size in the simple case of one additive illusion effect. For each way of manipulating size, we assume that there are two conditions (depicted by the two bars): Without the illusion, and with the illusion. Mean object size is shown in black and held constant for the object without illusory context. Mean illusion effect in gray. Thus, the overall height of each bar (black plus gray) indicates perceived size, as measured by the response (y-axis). **Left:** Standard physically-matched design, where mean object sizes are the same for all illusion conditions. **Middle:** Perceptually-matched design with the mean object sizes chosen such that the mean perceived size is equal. **Right:** Ganel et al.'s (2008) design, where one object is physically larger, but the other is perceived to be larger. It becomes obvious that a smaller, but still non-zero illusion effect could create a situation where the object on the right appears smaller; thus, a pattern of the perceived-larger object being grasped smaller is very much compatible with a non-zero illusion effect in grasping.

Such a method is elegant, but it also has a number of problems. Firstly, the illusion effect cannot be measured without uncertainty. This is especially important, since physical objects have to be grasped, which necessarily vary discretely in size. Indeed, objects as much as 1 mm apart have been used in this method (Aglioti et al., 1995), in an illusion with a typical illusion effect size of less than 2 mm (Franz & Gegenfurtner, 2008). Thus, the best achievable perceptual match may not be a very good match overall, and uncertainty introduced through slope correction (when using physically matched stimuli) is replaced by uncertainty introduced by an imperfect perceptual matching procedure. Which of these designs is better suited to reliably detect small

difference in illusion effects depends on the precision of the match and the measured slope. It has to be noted that the slope-correction method has the advantage that it allows for quantifying the illusion effect in grasping in a way that is comparable to perceptual measures. Although it is of course possible to use multiple matched pairs and calculate a response slope within each illusion configuration, this would still require applying the slope correction to a noisily measured difference between residual perceptual differences (also slope-corrected) between matched configurations, and MGA differences. Since one would have to apply two different slope corrections just to estimate one illusion effect, the main advantage of the perceptual matching method would be negated and this would not be a sensible method.

Ganel et al.'s (2008) method does not require an exact match or exact quantification of the illusion effect in grasping (see Figure 3). Object sizes are deliberately chosen in a way that the physical difference exists, but is smaller than the perceptual illusion effect. This way, the physically smaller of two objects is perceived to be the larger one. If now we find a smaller MGA for this (perceived to be smaller) object than for the larger (perceived to be smaller) object, this clearly demonstrates a dissociation between perception and action. However, the interpretation of this dissociation is tricky. Ganel et al. (2008) speak of a 'double dissociation' between perception and action in the title of their article. This 'double dissociation' is suggested to be such that (a) perceived size affects perception, but not grasping, and (b) physical size affects grasping, but not perception.

The argument thus relies on conceiving perceived and actual size as two independent concepts. However, in Ganel et al.'s (2008) study, perceived size is nothing else than the physical size, plus

(or minus) the illusion effect, so that functionally there is no difference between saying (a) ‘physical size affects grasping’ and ‘perceived size does not affect grasping’ on the one hand, and on the other hand saying (b) ‘the illusion effect does not affect grasping’. As a consequence, this interpretation of a double dissociation between physical size and perceived size is functionally indistinguishable from the claim of a single dissociation that assumes grasping to be immune to illusions. This matters, as single dissociations are typically interpreted not as strongly and taken to represent evidence of little more than that at least one task component is more critical in one task than another, or a ‘hierarchy of function’ (McCloskey, 2003; Teuber, 1955, p. 283). It is also much more parsimonious to talk about the illusion effect, as this is essentially the only difference between perceived and physical size, and perceived size is highly correlated with physical size: Even within a visual illusion the same principle holds that all other things being equal, a larger object is perceived to be larger. Thus, in this context it makes more sense to conceptualise the illusion effect as the manipulation.

Thus, this study presents a single dissociation. This could still present quite strong support for the TVSH. However, there are several other problems with the methodology. One such problem is that participants had full vision of their hand; we know that this makes illusion effects in grasping look smaller, as the visual information becomes more informative and less distorted as the hand approaches the object (Post & Welch, 1996). The tasks were also such that participants were instructed to grasp or estimate the ‘short’ or ‘long’ object – an instruction that could quite plausibly affect conscious perception more than grasping. Another point concerns the analysis, as the illusion effect was indeed there, if a proper control was chosen. As can be seen in Figure 3, the pattern of results (physically larger object, perceived smaller, larger MGA) can be achieved

even with a non-zero illusion effect in MGA, as long as the physical difference is smaller than the perceptual illusion effect, but larger than the illusion effect in grasping. For the reasons outlined above, a smaller illusion effect on MGA is hardly surprising. This also emphasises that the Ganel et al. (2008) paradigm is really logically equivalent to the perceptual matching by Aglioti et al (1995): Whenever the illusion effect on MGA is smaller than the illusion effect on perception, there will be a difference between MGAs for matched objects. Under exactly these circumstances (and only under these!), it is also possible to create a situation as described by Ganel et al. (2008). Hence, the difference is a matter of statistical sensitivity, not a conceptual one. This point is still important, as illusion effects tend to be rather small, and MGA quite variable.

So how can we test the predictions of the TVSH? A comparison between two vastly different tasks is inherently difficult, and there are many possible confounds. Thus, it is arguably best to use a paradigm that has been tested before, which as the additional benefit that the TVSH's predictions are clearly defined. This brings us back to the Ebbinghaus illusion: Here, the TVSH makes the clear prediction that the MGA (if confounds are avoided) should not be affected by the illusion. Methodological concerns have been discussed in various articles (Franz, 2003; Franz et al., 2003, 2000; Franz, Hesse, & Kollath, 2009; Franz et al., 2005; Gonzalez, Ganel, Whitwell, Morrissey, & Goodale, 2008; Haffenden & Goodale, 1998; Haffenden et al., 2001; Post & Welch, 1996), but the data is still considered unclear ('There is still no consensus on whether this dissociation reflects a fundamental difference in ventral- and dorsal-stream visual processing, as is outlined in the TVSH' – Whitwell et al., 2016, S. 2).

2.2 Is there an effect of the Ebbinghaus illusion on grasping? Study 1

Our experiment (Kopiske, Bruno, Hesse, Schenk, & Franz, 2016b) was designed to answer the question if the Ebbinghaus illusion affects the MGA in grasping in a similar way to the perceptual illusion effect. Since studies examining this question are among the first and most cited works in the TVSH literature on healthy participants (Aglioti et al., 1995; Haffenden & Goodale, 1998; Haffenden et al., 2001; 529, 238 and 110 citations via isiknowledge.com; Clarivate Analytics, 2016, retrieved 04.10.2016), this is a central question for the TVSH. The long discussion on this topic also means that (a) the TVSH makes explicit predictions here, which (b) we can test with confidence in our methods. As there are still two opposing interpretations of the data – obstacle avoidance on the one hand, a common size representation for perception and action on the other hand – the question is still open.

2.2.1 Experimental design and procedure

The basic design of our study was closely modelled after the studies by Haffenden et al. (2001) and Franz et al. (2003). More precisely, we conducted a direct replication of these studies – the same stimuli, tasks, and number of repetitions per condition – which was augmented by some additional conditions.

Discs of different sizes were presented embedded within Ebbinghaus-context circles and either a grasp or a perceptual judgement was performed (Figure 4a). Grasping was performed in the typical way, ‘open-loop’ (i.e., without visual feedback after the movement was initiated, such that there was no closed feedback loop; visibility was controlled with PLATO LCD-goggles, Milgram, 1987). Hand motions were recorded with an Optotrak Certus (Northern Digital,

Waterloo, Canada) with three infrared cameras, and diodes attached to the thumb, index finger and wrist of the participants. The main dependent variable was the MGA, from which the illusion effect was derived as the difference between the mean responses to size-matched discs embedded in different illusion contexts.

The illusion effect in grasping was compared to three different perceptual illusion effects: One was obtained in a classic perception task of matching the central disk to a circle from a graded series of circles without Ebbinghaus surrounds (Figure 4b), two other perceptual measures were manual estimates obtained from ME ‘open-loop’, as well as ‘closed-loop’ (i.e., with full vision of the hand at all times). ME has been used in both variants in several previous studies (Dewar & Carey, 2006; Haffenden & Goodale, 1998; Haffenden et al., 2001) and according to its proponents can be seen as a ‘manual read-out’ of perceived size (Haffenden & Goodale, 1998), comparable with cross-modal perceptual matching (Stevens, 1959).

The context circles corresponded to the ones used by Franz et al. (2003): The traditional Ebbinghaus conditions, with two additional configurations so that size and distance of the context circles were varied on two levels each, independently, creating the four configurations ‘large-far’ (LF), ‘large-near’ (LN), ‘small-far’ (SF), and ‘small-near’ (SN). Our targets were PVC discs of 3 mm thickness and 28, 30, and 32 mm diameter.

An additional condition that was tested neither by Haffenden et al. (2001) nor Franz et al. (2003) but in some previous studies (Aglioti et al., 1995; Haffenden & Goodale, 1998), was our so called ‘perceptually-matched’ condition (the standard condition where the physical object size was manipulated directly was called ‘physically-matched’). In this condition, two discs were

embedded in the LF and SF configurations, respectively. These discs were made to look equally large by separately matching them to a neutral 30.5 mm circle before each participant started the main experiment, by employing two 1-up, 1-down staircase procedures for each disc (one starting at 28 mm, one starting at 32 mm). Our step-size for the objects was 0.25 mm. These discs were then presented within all tasks at randomised positions. For these discs, the TVSH predicts a difference in MGA, while it predicts no difference between physically-matched discs embedded in different Ebbinghaus configurations. By including the perceptually-matched discs in all perceptual tasks, we also assessed the quality of the match we achieved. As laid out in section 2.1.5, achieving a good match is far from trivial and can produce spurious effects in the data.

2.2.2 Procedure: Preregistration, replication, and a ‘confirmatory experiment’

This study was preregistered at the journal Cortex, meaning that the introduction, methods and proposed analyses went through peer-review before data collection with the promise that the finished manuscript would be published if this first stage of review was successful. For details, see Box 1.

Box 1 | A ‘Registered Report’ at Cortex

Registered Reports are format of the journal *Cortex* that is somewhat unusual in psychology. A manuscript is submitted and reviewed with just the introduction, methods and proposed analyses included, and before any data have been collected. After a regular review process, *in-principle-acceptance* is given, and only then may data be collected. The finished article is then re-submitted after data collection and will be published barring gross negligence on the authors’ part.

Our project was initially submitted in March 2014 with the title „The functional subdivision of the visual brain: Is there a real illusion effect on action? A multi-lab replication study“ and given in-principle-acceptance in September 2014. The full manuscript was resubmitted in September 29th, 2015, and accepted for publication March 14th, 2016.

The experiment was conducted in four laboratories: In the Department of General Psychology, University of Hamburg, as well as the groups of Constanze Hesse (University of Aberdeen), Nicola Bruno (University of Parma), and Thomas Schenk (University of Erlangen). I visited each laboratory to verify that the protocol was indeed executed in always the same way.

This procedure is meant to ensure that studies that are theoretically motivated in a coherent way and have been conducted thoroughly get published regardless of the outcome, while also preventing undisclosed flexibility in the data analysis by the authors (Chambers, 2013c). Selectively not publishing data (‘file drawer’ problem – Rosenthal, 1979) or analyses (Simmons, Nelson, & Simonsohn, 2011) is a known problem that can lead to systematic distortions in the literature (Dwan, Gamble, Williamson, & Kirkham, 2013; Sterling, 1959; Sterling, Rosenbaum, & Weinkam, 1995). Preregistration is meant to prevent these practices and emphasise accuracy and reliability of research over novelty (Chambers, Dienes, McIntosh, Rotshtein, & Willmes, 2015; Nosek & Lakens, 2014). In our case, preregistration was particularly attractive, as our experiment was specifically designed to be a confirmatory experiment (Wagenmakers, Wetzels,

Borsboom, van der Maas, & Kievit, 2012) in a field where a large portion of the literature has come from a small number of laboratories with clear patterns in their results.

It is important to note that replication is not necessarily a useful arbiter of the veracity of an effect. If we assume that the original experiment was done in a thorough and conscientious manner, that is if all questions were theoretically motivated and all corresponding tests were published, then the original data would not lose any of their value once a replication has been published. In this case, the replication can be considered simply as additional data and a meta-analysis might be the best tool to examine the effect (Francis, 2013a, 2013b). If, however, we assume that the field suffers from a file-drawer problem (Rosenthal, 1979; Sterling et al., 1995), then indeed it is prudent to assign more weight to data from replications, as it is unlikely that the original data were truly selected randomly. This of course requires the replication to be free of the same file-drawer phenomenon, which is not necessarily the case, in which case meta-analyses become near-useless, as aggregating biased data will give you a biased aggregate (for an intuitive and vivid explanation, see Gelman & Loken, 2014). In our case, both approaches (more data, and confirmatory preregistered experiments) are satisfied: We chose to conduct a preregistered replication study, in which our sample was more than twice as large as the largest previously studied sample on the same question.

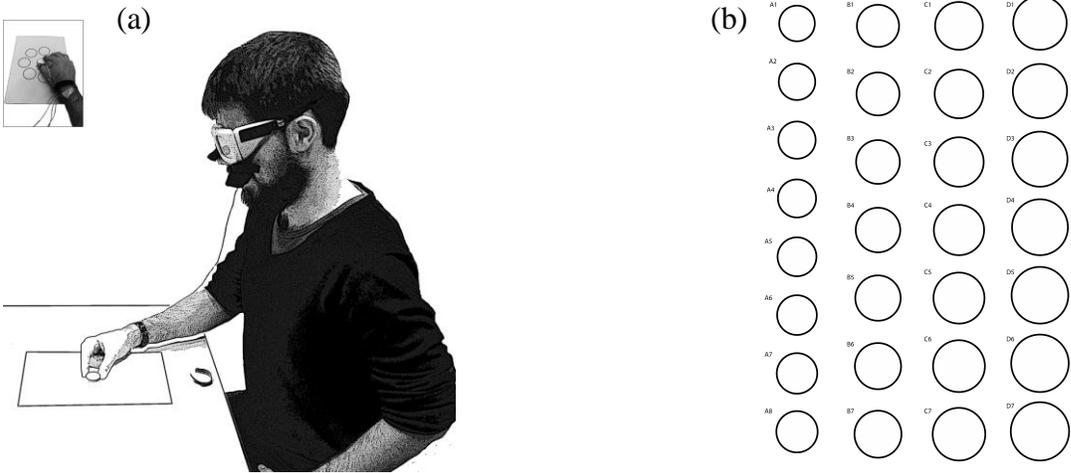


Figure 4: Setup used in study 1. **a:** Grasping task with the participant wearing LCD goggles (Milgram, 1987) and infrared diodes on the wrist, index finger and thumb while grasping a white PVC disc. **b:** Graded series of comparators as used in the classic perceptual task in study 1.

2.2.3 Power and Bayes factors

Our large sample had another advantage in that is guaranteed sufficient power to reliably detect any effects of interest. This is especially relevant if we consider that the point of a confirmatory experiment like ours is to decide between two theories, one of which predicts a certain effect, while the other predicts the absence of the same effect. Thus, it is important to be able to draw strong conclusions in both directions, which in turn means doing at least one of two things: Controlling for both type I and type II errors, or using a measure that would allow us to measure evidence in favour of the H0.

We controlled both error rates by ensuring a high statistical power, while also employing Bayes factors, calculated following the logic by Dienes (2008, 2011). These compare the likelihoods of two statistically modelled hypothesis (in our case: Normal or uniform distributions around certain values, see published article 1, section 2.4), which are divided by one another to compute a Bayes

factor, that is, a measure of how strongly prior beliefs about the effect should be updated because of the data. Bayes factors have been much discussed over the past few years (see e.g. Dienes, 2011; Rouder, Speckman, Sun, Morey, & Iverson, 2009; Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011; Wetzels et al., 2011) and promise multiple advantages over frequentist statistics, such as optional stopping, no need to correct for multiple comparisons, measuring strength of evidence directly, and the possibility of gathering evidence for the H₀. This is not to say that Bayes factors are to be preferred in general: They rely on some additional assumptions such as assigning prior beliefs, do not control error rates, which is problematic and negates their advantages with regards to optional stopping and multiple comparisons if they are interpreted in a binary way (see e.g. Mayo & Spanos, 2011). Indeed, in practice Bayes factors also very often lead to the same inferences as frequentist tests (Wetzels et al., 2011, figure 3). The advantages of either method shall not be discussed here. However, our reasoning for including Bayes factors was that we wished to strengthen our methods by examining the data from multiple perspectives. Should these converge, this would give a clearer picture and strengthen our conclusions; if not, this would draw attention to the fact that the data may not be as clear as either analysis might indicate.

2.2.4 Results and conclusions

In our study, we found illusion effects in all tasks and in all comparisons of small and large illusory contexts. Importantly, this includes contrasts in grasping with a matched context circle distance (SF-LF and SN-LN) – results here did not differ significantly from those in the same contrasts in perceptual tasks, indicating that obstacle avoidance cannot explain these results.

Bayesian and frequentist analysis converged on this conclusion. Data, compared to the predictions, can be seen in Figure 5b. The second important test for a difference in illusion effects was the comparison of MGA for perceptually-matched discs. Here, we found no evidence of a difference in MGA, despite a physical difference of on average 1.15 mm. This is despite our large sample, and constitutes very strong evidence for the H0 under the Bayesian framework. Data can be seen in Figure 5a. Indeed, small, non-significant differences do exist; however, these are in the same direction, and of the same magnitude, as residual differences that also exist in perceptual tasks, indicating that our staircase procedure may have measured a slightly too large illusion effect. It follows from the fact that the *physically* larger object was also *perceived* to be slightly larger that object size differences in our perceptually-matched conditions were slightly too large. That is, the conditions were actually over-sensitive to detecting an effect of physical size over perceived size on MGA, emphasising that our non-finding indicates that such an effect did not exist.

Thus, our study very strongly speaks for the notion that the Ebbinghaus illusion affects grasping just like perception, and that obstacle avoidance does not matter for this. However, grasping the Ebbinghaus illusion is only one of many possible paradigms to test grasping in visual illusions, which raises the question: What do our results mean for the debate on visual illusions in action?

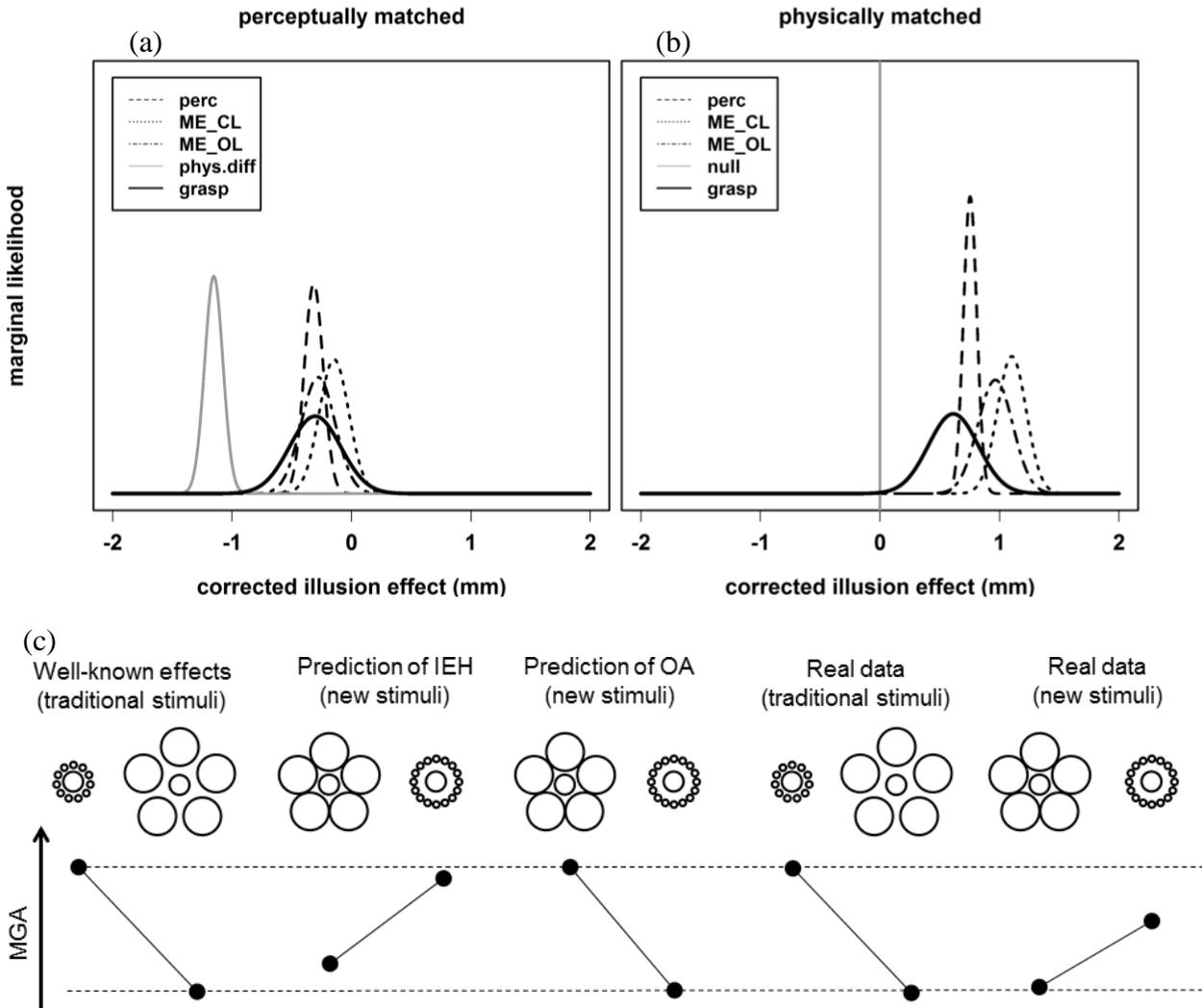


Figure 5: Predictions and data from study 1. Relevant contrasts from our study, predictions and actual data. Curves in panels a and b indicate means and standard errors as measured in perceptual tasks and grasping; height chosen such that the area under each curve is constant. Thus, curves are a graphical representation of the hypotheses used for Bayes factor calculation. **a:** Illusion effect SF-LF, perceptually-matched. **b:** SF-LF, physically-matched. **c:** Obstacle avoidance predictions tested by comparing new configurations LN and SF.

2.3 Weighing the evidence and advancing the debate: What does grasping the Ebbinghaus illusion really tell us? Public debate about study 1

What we are doing in investigating whether grasping is affected by the Ebbinghaus illusion is testing a prediction of the TVSH – namely, that this is not the case. Thus, it becomes clear that

action-perception dissociations are not per se directly relevant to the TVSH, but only when two conditions are met: (a) The TVSH makes predictions as to whether to expect a dissociation or not, and (b) we can test for such a dissociation in a confound-free way.

In the Ebbinghaus illusion, our arguments has been that both conditions are met and that this is why the paradigm lends itself well to testing TVSH predictions: It is very clear that the TVSH predicts a dissociation (Aglioti et al., 1995; Milner & Goodale, 2006), with a very explicit reasoning why (Milner & Dyde, 2003). There has also been a long discussion of methodological issues in grasping the Ebbinghaus illusion (Franz et al., 2003, 2000; Haffenden et al., 2001; Pavani et al., 1999), unearthing general pitfalls such as comparing dependent variables with different responsiveness, but also pitfalls specific to this design such as obstacle avoidance or superadditivity. Knowing these issues should make us more confident in our methods when using the Ebbinghaus illusion than we would be with less-studied distortions. Our study (in line with many before it, see Figure 6) now claims that even when avoiding these pitfalls, the illusion effect on grasping remains.

However, in a commentary on our article Whitwell and Goodale (2016) argued, in essence, that condition (b) was not met in our study. In addition, they make the case that stronger tests of the TVSH's predictions do exist in older studies, with results that are opposite to and cannot be explained by our study. We responded to this in another commentary (Kopiske, Bruno, Hesse, Schenk, & Franz, 2016a) by reanalysing data from these older studies and considering the methodological criticisms, which we do not consider problematic for our study. In the following sections, I will spell out the criticism as well as our response in more detail.

2.3.1 Did our study adequately test TVSH predictions? The Whitwell & Goodale arguments

Whitwell and Goodale (2016) accept the conclusion that obstacle avoidance cannot explain our results in grasping (“The focus of the Kopiske et al. study was ostensibly on attempting to replicate two later studies that examined the obstacle avoidance explanation directly [...] We actually have no issue with this aspect of their study.”), which moves the discussion solely to the perceptual processes underlying grasping and perceptual tasks. Here, however, they take issue with several points. Briefly, this concerns the use of a single-illusion display (see fig 2b), as opposed to the dual-illusion display used for example by Aglioti et al. (1995). Whitwell and Goodale (2016) claim that this creates problems with regard to (i) the phenomenology of the illusion, and (ii) the strength of the illusion effect. Further, they claim (iii) that some of the methodological advances have made it necessary to revisit some old data from a study by Haffenden and Goodale (1998), which show a very different pattern of results compared to our study (and other studies, see Figure 6). I will go through these points one-by-one.

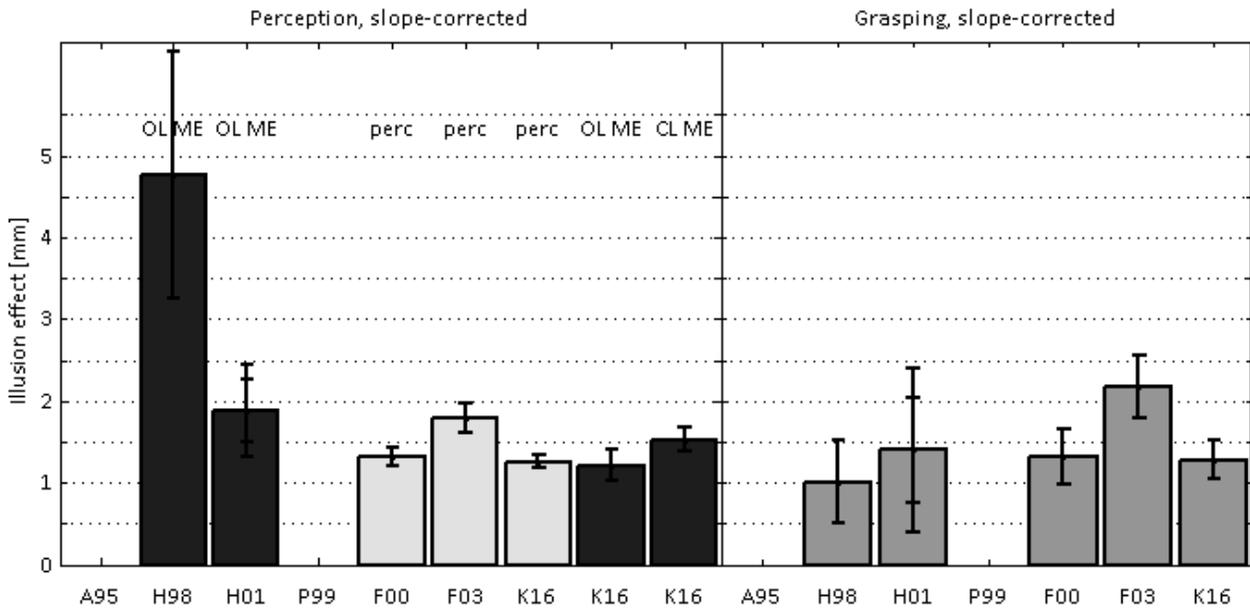
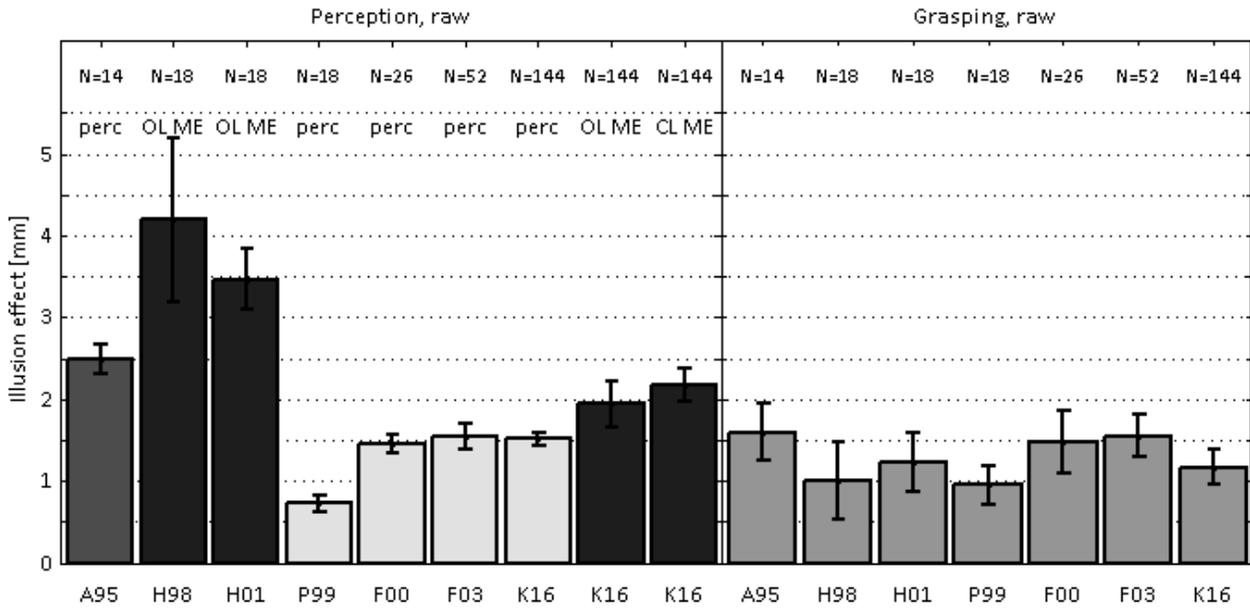


Figure 6: Illusion effects in grasping and different perceptual measures. Bars in light grey show results from classic perceptual size-matching tasks, dark grey shows manual estimates. Error bars indicate within-participant S.E.M. **Top row:** Uncorrected illusion effects. **Bottom row:** Slope-corrected illusion effects (Franz, 2007; Franz et al., 2009). Slopes required to compute corrected illusion effects were not available for Aglioti et al. (1995) and Pavani et al. (1999). Included are studies using traditional configurations (SN, LF): Aglioti, et al. (1995), A95; Haffenden & Goodale (1998), H98; Haffenden, et al. (2001), H01; Pavani, et al. (1999), P99; Franz, et al. (2003, 2000), F00 and F03; Kopsike, et al. (2016b), K16. For H01, the S.E.M. could only be estimated as the formula for the Taylor estimation uses the covariance between the slope and the illusion effect, which is not available. Two estimates are given: For a low correlation ($r = -.5$; larger estimate of the S.E.M.), and for a relatively high correlation ($r = .5$).

In more detail, their claim that (i) the single-illusion display used in our study is phenomenologically different from the dual-illusion display is rather straightforward, but it is not clear why this would be problematic. Whitwell and Goodale (2016) point out that participants never got to experience the “striking, real-time phenomenology of the standard two-configuration Ebbinghaus illusion”. The use of ‘standard’ could be disputed, as a vast number of studies concerning grasping (Franz et al., 2000; Haffenden et al., 2001; Pavani et al., 1999) as well as perception only (Coren & Enns, 1993; Coren & Girgus, 1972; Girgus et al., 1972; Knol, Huys, Sarrazin, & Jirsa, 2015; Roberts et al., 2005) was conducted using single-illusion Ebbinghaus displays. Despite this, they argue that using the single-illusion display made the perceptually-matched condition obsolete, as no direct match was created, thus marring our attempts to create a good test for a dissociation with the TVSH as the H0.

To this, there are several responses. It is doubtlessly true that the phenomenology as described by Whitwell and Goodale was not achieved in our study. However, the question remains why this should matter. Our goal was to replicate the studies by Haffenden, et al. (2001) and Franz, et al. (2003), as we considered these to be the methodologically superior, confound-free studies. We did not aim to include a replication of Aglioti, et al. (1995) or Haffenden and Goodale (1998). We did include a similar perceptual matching condition where two objects should be perceived as

equally large, with the purpose to include a condition with the TVSH as the H0; thus, the relevant question would be whether we succeeded in presenting participants with two discs they perceived to be equal in size. Data from our perceptual tasks indicate that this was the case. Indeed, our test was arguably more stringent in this than the methods used by Haffenden and Goodale (1998), as we employed a standard perceptual task which is much more sensitive than ME (used in Haffenden & Goodale, 1998).

Now, one could argue that the difference in phenomenology matters beyond the sheer magnitude of the illusion effect by making the illusion qualitatively different. However, this would not only be a new, untested assumption, for it to undermine our study one would also have to assume that this other illusion would not be a ventrally processed illusion, which is inconsistent with Milner and Dyde's (2003) claim that what matters is whether the illusion is contextual (and the single-illusion display still is contextual), but also with the way the single-illusion Ebbinghaus display has been used to support the TVSH (e.g., Haffenden et al., 2001).

As for their other claim that (ii) the strength of the illusion is critically diminished by the single-illusion display, this is simply a matter of statistical power. We conducted an *a priori* power analysis as part of registering our design, in which we based our calculations on the effect size typically achieved in grasping a single-illusion Ebbinghaus display. Our calculations were based on wanting to be able to detect an effect only 70% as large as the typical effect, or around the smallest effect size found. For this, we still had over 99% power. Bayes factors provide convergent evidence that sensitivity was not an issue in our study. If we want to go into more detail, we know that due to superadditivity the effect tends to be approximately 1.5 times larger

in dual-illusion Ebbinghaus displays as opposed to single-illusion displays. Since the S.E.M. decreases with the square root of the sample size, to ‘compensate’ for an effect size reduced by a factor of 1.5, our sample would have to be $1.5^2 = 2.25$ times as large to achieve the same power. Our sample was larger than the largest dual-illusion display study by a factor of 8, which reemphasises that power was not the problem and brings us to the next point.

Finally, Whitwell and Goodale (2016) also note that (iii) other results need to be considered before drawing conclusions about grasping illusions in general. Specifically, they point to a study by Haffenden and Goodale (1998). This study is interesting, as it uses the classic, not distance-matched illusion displays, as well as a dual-illusion display, and was therefore criticised as being potentially problematic from a methodological standpoint (Franz et al., 2000; Pavani et al., 1999). Indeed, it is only recent investigations that have shown that this criticism may not be so problematic after all: Foster and Franz (2014) investigated the conditions under which superadditivity applies, coming to the conclusion that superadditivity does not occur when a comparator (that is, the stimulus to be adjusted to the size of the target) is positioned outside of the illusion. Whitwell and Goodale (2016) argue that this can be taken to mean that manual estimation as used by Haffenden and Goodale (1998) – where the participant’s hand is the comparator – is not affected by superadditivity. Likewise, with our study (Kopiske, Bruno, et al., 2016b) providing evidence that obstacle avoidance cannot explain Ebbinghaus display effect on MGA (see also figure 5c), the problem of interpreting effects on MGA also disappears. Haffenden and Goodale’s (1998) study, then, also provides a confound-free comparison of perception and grasping, and comes to the conclusion that illusion effects are larger in perception.

In response to this, two things should be pointed out. One is the fact that in their study, Haffenden and Goodale (1998) found very inconsistent illusion effect within their perceptual measures (this becomes very clear in article A.2, Figure 1). This (a) casts doubt on the validity of these measurements that are supposed to measure the same (i.e., perception), and (b) raises the question of which perceptual measure should be compared to grasping. As it stands, a dissociation is demonstrated between grasping and one measure (manual estimation) but not another (size matching). In our view, this does not support the conclusion that grasping and the general concept of ‘perception’ are dissociable. The second issue, as I have alluded to earlier, is the sample size. Even if one were to accept that the comparison between grasping and manual estimation should be the most decisive one (and for some reason ignore the classic size matching task), one would find a situation with one study, N=18 (Haffenden & Goodale, 1998) finding a dissociation that may or may not be a false positive, and on the other hand a preregistered study (Kopiske, Bruno, et al., 2016b), N=144, finding no dissociation.

2.3.2 Why our study adequately tested TVSH predictions: Considering other data and designs

As recurring theme has been the emphasis on our study being the best available test of the TVSH prediction that under the right conditions, grasping is unaffected by visual illusions. However, as Whitwell and Goodale (2016) correctly point out, there are other designs which may be considered good tests, and where the data may speak more strongly for a dissociation.

To appraise the evidence, the criteria established above may be applied: Knowing how to test for a dissociation, and knowing that testing this is meaningful. This can be applied to many different illusions, but an especially interesting case is the Ponzo illusion (Fig 2c), where

dissociations have been demonstrated in several high-profile studies (Ganel, Tanzer, et al., 2008; Whitwell et al., 2016). Despite these studies, it is not clear if the Ponzo illusion is actually processed before or after the dorsal-ventral split in V1 (Murray, Boyaci, & Kersten, 2006; Schenk & McIntosh, 2010; but see also Goodale & Milner, 2010). The same argument can be used to question the validity of results from the empty-space illusion (Milner & Dyde, 2003). In fact, the Ebbinghaus illusion is one of the few where the prediction of the TVSH is entirely clear. In combination with our knowledge of specific methodological pitfalls, we argue that our study stands as a strong test of TVSH predictions in healthy participants. Criticism by Whitwell and Goodale (2016b) does not make a consistent case against this, and thus cannot explain our findings and should not weaken our conclusions.

2.4 Object size in visual perception and action: Current state of the debate

The TVSH (Goodale & Milner, 1992; Milner & Goodale, 1995, 2006) continues to be an influential theory, despite criticism. The question if object size is processed separately for perception and action has not been conclusively answered; at the same time, many predictions of the TVSH on this matter have not been confirmed (Schenk et al., 2011). Our study contributes to the debate by emphasising that illusory distortions of visual size information are found in grasping just like in perception. The pessimistic view that one would have to investigate separately whether each phenomenon of visual processing that could potentially be action-relevant is actually processed by the dorsal stream is still a possibility; however, this may also be seen in a more optimistic light, in that the TVSH continues to provide a framework that can be useful to guide research on the visual control of motor actions (see Schenk et al., 2011, p. 802).

Chapter 3 – Nonsymbolic magnitude on the mental number line (studies 2 and 3)

When talking about magnitude in the semantic domain, this will most commonly refer to numbers. These can be represented in several ways, be it symbolically through numerals (e.g., 5, or V, or Ⅴ, or a spoken word ‘five’) or nonsymbolically through the numerosity of items in a set (e.g., III, or Ⅲ, or three short ‘beep’ sounds, or three fingers held up on one hand). In spite of the differences between these types of number representation, the principles by which humans manipulate numbers remain the same. However, there are some robust interactions of numerical cognition and visuospatial processing of numerical stimuli, which raises the question of whether the semantic content and the visual appearance of numbers are truly independent.

Two concepts have had a prominent position in the field of numeric cognition for the past twenty years: The mental number line (MNL; Dehaene, 1992; Moyer & Landauer, 1967), a metaphor of the representation of magnitude, and the SNARC effect (Dehaene et al., 1993), whereby numerical magnitude and spatial properties of responses are associated. While they have been much-studied, questions remain about the scope of these concepts: Do properties of the SNARC effect and the MNL that hold for symbolic representations of magnitude hold for nonsymbolic representations? And are these properties based on an automatic, general, amodal magnitude representation that is independent on sensory stimulus features? In two studies, we tried to approach these questions and tested properties of these two phenomena with regards to numbers, as well as non-symbolic magnitude.

3.1 The MNL and the SNARC effect

The MNL as a metaphor has been used at least since the 1960s (Moyer & Landauer, 1967), but the version currently most referred to comes from Dehaene's (1992) so called 'triple code model', which postulates three main representations of number: A visual digit form, a verbal number-word form, and an analog magnitude representation that may be conceptualised as a number line. The MNL's function in this model is to enable estimation of numeric magnitudes to enable nonverbal manipulation or comparison. This may be relevant in approximate calculation with numbers (e.g., 17×21 is 'about 350'), or in appraising and manipulating numeric magnitude that is not precisely known, such as the number of dots in a cloud. This feature is often called the stimulus' *numerosity* (corresponding to the cardinality of elements contained in it - Butterworth, 1999). Since the numerosity of a stimulus is indeed a way to represent a certain numeric magnitude that is not reliant on an arbitrary symbol, I will in this thesis use the more general term 'nonsymbolic magnitude' to refer to numerosity, as well as other forms of numeric magnitude not represented by numbers or number words.

The SNARC effect refers to a stimulus-response association of numeric magnitude and space (Dehaene et al., 1993). This association is most typically observed in classification tasks, such as pressing a button depending on the parity (Dehaene et al., 1993) or magnitude (Herrera, Macizo, & Semenza, 2008) of a number, which will typically show a speed advantage for small-left and large-right responses (but this depends on reading habit: Dehaene et al., 1993; Ito & Hatta, 2004; Shaki, Fischer, & Petrusic, 2009). Importantly, the association persists in tasks where numeric magnitude is not strictly relevant to the task, such as parity judgement (Dehaene et al., 1993). The association seems to extend to SNARC-like effects that have been observed in other tasks,

including grasping (Chiou, Wu, Tzeng, Hung, & Chang, 2012), although other associations that might be predicted from the MNL metaphor have not been found (like, e.g., number and movement direction – Santens & Gevers, 2008). The SNARC is most frequently taken as evidence for an MNL that organises magnitudes from different notations internally (and, importantly, spatially) and is accessed in tasks related to magnitudes (Dehaene, 1992; Feigenson, Dehaene, & Spelke, 2004). This is a core assumption of the triple-code model (Dehaene, 1992) or direct-mapping model (Santens & Gevers, 2008). However, other interpretation of the SNARC effect exist, some assuming a mental number line (e.g., the dual-route model postulating several routes of activation, a semantic route and an automatic route that has different properties than we numeric processing under most circumstances; Fias, 2001), while others explain it in terms of more general stimulus-response polarity effects (Cohen Kadosh & Walsh, 2009; Proctor & Cho, 2006; Walsh, 2003; but see Santiago & Lakens, 2015). Recently, it has even been proposed that the effect can be explained in terms of language statistics (with more frequent words being associated with left responses; Hutchinson & Louwerse, 2014).

3.1.1 The MNL as a link between input and output

As reviewed above, the MNL's is primarily seen as an analog representation of magnitudes (Dehaene, 1992). This representation is thought to be automatic when participants are presented with numeric stimuli, as can be seen by for example by the automaticity of the SNARC effect: There is no need for analog magnitude representation to judge a number's parity, yet the effect occurs (Dehaene et al., 1993) in this task.

Thus, there has been a lot of research on what the properties of the MNL may be. For example, it can be shown that responses to quantities that cannot be assessed precisely (and thus should rely heavily on the MNL, according to Dehaene, 1992) follow a nonlinear function, that is a logarithmic or a power function. This is often taken as evidence that the representation of magnitudes is compressed logarithmically (e.g., Dehaene, 2003), with different shapes being explained by output demands (e.g., Izard & Dehaene, 2008, proposed an ‘output grid’ that could be adapted to the task that was being presented).

However, it is quite difficult to uncouple the effects of input, internal representation, and output (in essence, the three components of the triple code model, Dehaene, 1992), which in turn makes it difficult to deduce MNL properties from number line tasks, as an example may illustrate: The interpretation of a logarithmic MNL was based mainly on results from one task – estimating the approximate number of dots in a cloud of dots (Dehaene, Izard, Spelke, & Pica, 2008; Gallistel & Gelman, 1992; Whalen, Gallistel, & Gelman, 1999). The problem is that it is quite possible to explain the shape as a consequence of a linear representation with scalar variability (Cantlon, Cordes, Libertus, & Brannon, 2009). That is, if the variability of responses increases proportionally with stimulus magnitude, then we should see more relatively small responses (because a higher proportion of relatively small magnitudes will lead to relatively small responses than the proportion of relatively large magnitudes leading to relatively large responses), making the response function look logarithmic without assuming a logarithmic representation. Importantly, this would be a property of the response, not of the MNL. Indeed, scalar variability of number line responses has been found (Dehaene, 1992; Gallistel & Gelman, 1992), leading to at least three possible conceptualisations of the MNL: A logarithmic function, a linear function

with scalar variability, or a third possibility of an internal representation that conforms to two psychophysical principles often found in sensory perception: *Weber's law* (Fechner, 1860), in that it represents magnitudes with more uncertainty the greater they are, and Stevens' *power law* used by Stevens (1957) to describe how participants match intensities in different senses, according to which the mapping of one sensory modality to another – in this case, visual input to magnitude – tends to follow a power function (Dehaene, 2003).

Both Weber's law and the response power function are typical of sensory perception and have been part of an argument that proposes that numeric magnitude may in fact be sensed directly from visual input, without it being transformed into a semantic representation (Arrighi, Togoli, & Burr, 2014; Burr & Ross, 2008; J. Ross & Burr, 2010), a hypothesis that has been further corroborated by results that nonsymbolic magnitude displays adaptation effects across sensory modalities (Arrighi et al., 2014; see Anobile, Cicchini, & Burr, 2015 for a recent overview of the argument).

It should be noted at this point that methodology is quite central to these questions. For example, scalar variability of a response can only be taken as evidence for the existence of Weber's law (Fechner, 1860) if the response is free of any systematic effects of response magnitude on response. Likewise, sensory matching in the manner of Stevens would not be immune to alternative explanations like the one by Cantlon, et al. (2009) that the characteristic shape may be a result of increasing uncertainty rather than compressive representation; indeed, it has been proposed for a long time that absolute judgements of noisy input may exhibit nonlinear properties as somewhat of a general rule (Haubensak, 1992; Parducci & Perret, 1971). Hence, it

is important to look at results from different methods to be able to draw conclusions about properties of the MNL. The most typical task is simple number estimation: Participants are presented with a nonsymbolic magnitude (like a cloud of dots) and asked to report the magnitude contained in the stimulus. However, it is also common to use an actual number line to measure the metaphorical mental number line. In such a task, participants move a cursor over a line and indicate where they think a given magnitude belongs on the line; this has been done with healthy Western adults (Siegler & Opfer, 2003), children (Barth & Paladino, 2011; Siegler & Opfer, 2003), indigenous people who have no concept of verbal number (Dehaene et al., 2008), with results consistently showing a systematic nonlinearity that, however, has been interpreted in several different ways. It stands to reason that the properties of encoding and representation would be similar between these methods; however, output demands differ greatly. Hence, this warrants further investigation.

3.1.2 The SNARC effect, Chinese characters, and visual features of notation

The SNARC effect has been taken as a tool to deduce general properties about numerical cognition (Cohen Kadosh & Walsh, 2009; Dehaene, Piazza, Pinel, & Cohen, 2003; Fias, 2001; Santens & Gevers, 2008), primarily due to its considerable ubiquity and robustness, having been demonstrated in Arabic digits and written number words (Dehaene et al., 1993), non-Arabic digits (Hung, Hung, Tzeng, & Wu, 2008; Ito & Hatta, 2004), spoken words (Nuerk, Wood, & Willmes, 2005), hand signs for numbers (Bull, Blatto-Valle, & Fabich, 2006; Iversen, Nuerk, Jäger, & Willmes, 2006), dice patterns (Nuerk et al., 2005), letters (Gevers, Reynvoet, & Fias, 2003), months (Gevers et al., 2003), and more (for a review, see Wood, Willmes, Nuerk, &

Fischer, 2008). Following this wealth of evidence for a SNARC effect in many conditions, it has been proposed that the SNARC effect is completely amodal and independent of notation (Dehaene, Molko, Cohen, & Wilson, 2004; Nuerk et al., 2005).

However, there are certain exceptions where a SNARC effect could be reduced or reversed by visual properties, such as visual complexity (Chinello, de Hevia, Geraci, & Girelli, 2012), or the typical context of a notation (like Chinese characters that may be read vertically; Hung et al., 2008). The latter is especially interesting, as there are several possible explanations that would warrant experimental testing: First, Chinese characters differ in visual complexity, too, and have been reported to be processed in different (and maybe multiple) stages (Cao, Li, & Li, 2010; Liu, Tang, Luo, & Mai, 2011), which is consistent with similar findings from Korean hand-sign number representations (Domahs et al., 2012). Second, reading habit may be different than for Arabic digits; however, this is only the case in Taiwan (where horizontal writing was not officially mandated until 2004), but not for Mainland China (where horizontal writing was mandated in 1955). Thus, reading experience – one of the key factors determining the SNARC effect (Dehaene et al., 1993; Shaki et al., 2009 - although there is evidence that this effect may be more unstable than previously thought, see Fischer, Shaki, & Cruise, 2009; Pfister, Schroeder, & Kunde, 2013) – is a confounding variable in Hung et al. (2008), but can potentially be controlled for by comparing Taiwanese participants with Mainland Chinese participants. This brings us to the questions of study 2: Does a horizontal SNARC effect persist in a visually complex notation where a SNARC effect has previously been questioned, that is in Chinese characters? And does it persist in a mixed symbolic-nonsymbolic notation with higher visual complexity, Chinese hand

signs? Both questions relating to the larger issue of whether the SNARC effect is indeed independent of notation, and what visual features may influence it.

3.2 The SNARC effect in Chinese numerals: Study 2

Our main question in study 2 (Kopiske, Löwenkamp, et al., 2016) was whether we would find a horizontal, left-to-right SNARC effect in Chinese characters and Chinese hand signs, as well as Arabic digits, testing participants from Mainland China. This was supposed to give us insight into two things: (a) Whether processing of numerical magnitude would show different patterns in Chinese participants, as has been suggested (Cao et al., 2010; Liu et al., 2011), and (b) whether the notation would matter with regards to the SNARC effect (Hung et al., 2008; Liu et al., 2011). This was especially interesting, as Chinese characters presented a notation that was similar to Arabic digits with regards to context, but differed with in visual complexity (Chinese characters; this has been proposed to influence numerical processing: Chinello et al., 2012), while Chinese hand signs not only differed in visual complexity and context, but also presented a mixed notation of nonsymbolic (for numbers up to 5; see fig 1, article A.3) and symbolic magnitude representation (for higher numbers).

3.2.1 Sample and design

We ran two experiments: A standard parity judgement task, and a magnitude judgement task. Both were conducted at Tsinghua University in Beijing, with two samples of N=26 and N=25 Mainland Chinese native speakers living in Beijing. In both experiments, participants were presented with one stimulus (a number between 1 and 9, with 5 being excluded) at a time, and asked to make a binary decision: Either whether this number was odd or even (experiment 1:

Parity judgement) or whether it was smaller or larger than 5 (experiment 2: Magnitude judgement). Each experiment consisted of six blocks: Three types of stimuli (Arabic digits, simple-form Chinese characters, and Chinese hand signs; see article A.3, Fig 1), each presented in two separate blocks that differed in response mapping (left-odd/right-even and left-even/right-odd in experiment 1, left-small/right-large, left-large/right-small in experiment 2). We ran the second experiment since it was not clear whether the mapping of stimuli to responses was confound-free with regards to visual properties in the first experiment. Indeed, we do not consider the mapping to be confound-free in magnitude judgement, but would consider it stronger evidence if a SNARC-effect were to persist in two (albeit imperfect) conditions. It has also been shown that somewhat different resources are needed to complete parity judgement and magnitude judgement tasks, respectively (Herrera et al., 2008; van Dijck, Gevers, & Fias, 2009). Responses times (RTs) of button presses of the ‘s’ and ‘l’ buttons on a standard USB keyboard were recorded as the main dependent variable.

3.2.2 Results and discussion

From RTs as a dependent measure, a SNARC-effect can be investigated through the very simple process of subtracting for each number left-handed RTs from right-handed RTs and testing if the resulting differences of response times (dRTs) decrease with higher number. That is, it is typically tested if a linear regression of dRT over number has a negative slope (Fias, Brysbaert, Geypens, & d’Ydewalle, 1996). Indeed, this is equivalent to a repeated-measures ANOVA over RTs with factors ‘number’, and ‘hand’, another popular approach to test for a SNARC effect (Pinhas, Tzelgov, & Ganor-Stern, 2012; Tzelgov, Zohar-Shai, & Nuerk, 2013).

The two measures differ when it comes to quantifying the SNARC effect: Whereas the slope indicates the magnitude of the effect, with explained variance as a measure of how well dRTs are predicted by numeric magnitude, the ANOVA only provides a partial eta-squared as a measure of the effect size in terms of variance explained. Similarly, R^2 is typically given along with dRT-slopes. At the same time, the ANOVA approach allows a rather straightforward way of comparing SNARC-effects in different notations by simply adding this as a factor. Ours was somewhat of a hybrid approach: For each experiment, we ran an omnibus-ANOVA over RTs with factors *numeric magnitude*, *side*, and *notation*, and conducted separate dRT analyses for each notation to confirm if a SNARC effect occurred there.

Our results can be seen in Figure 7. Broadly speaking, we found a SNARC effect in both experiments and in all notations. In both ANOVAs, we found interactions between the factors *hand* and *number*, which indicated a possible SNARC-effect, and no interaction with the factor *notation*. Note that this only indicated a possible SNARC-effect. To investigate the association between number and space more closely, we looked at the results of the dRT ~ number regression which we ran for each notation, in each of the experiments. Here, results mirrored almost perfectly the typical slopes from a recent meta-analysis (Schiller, Eloka, & Franz, 2016), but were less clear with regards to differences between notations. While slopes in all three notations differed significantly from 0 in experiment 1 (see appendix, article A.3, Table 1 – note that there is a typesetting error in the published version, as the rightmost column and the note should read ‘ R^2 ’ where it currently reads ‘R’) and not significantly from each other in parity judgement, they were still markedly different. Slopes were similar in experiment 2 (magnitude judgement), but in fact not significantly different from 0 for Chinese hand signs and Chinese characters. ANOVA

results mirrored those in experiment 1, in that the ‘SNARC-interaction’ *hand * number*, but no interaction with *notation* was statistically significant.

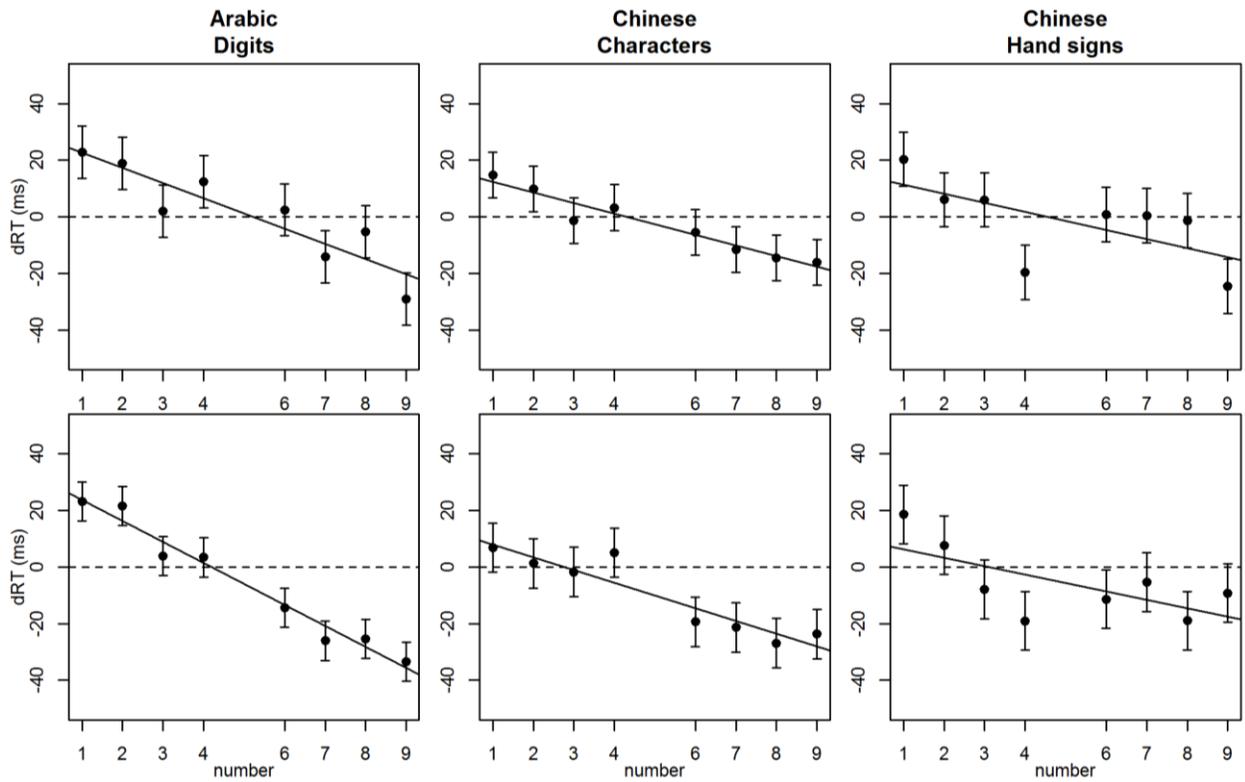


Figure 7: The SNARC effect in our study. Mean RTs in ms of right-handed responses – mean RTs of left-handed responses (i.e., dRTs, Fias et al., 1996) are shown. Lines indicate best linear fit of $dRT \sim number$, also given in Table 1 of appended article A.3. Results from experiment 1 (parity judgement) in the top row, experiment 2 (magnitude judgement) in the bottom row. Negative slopes are what we expect due to the SNARC effect. Error bars show pooled within-subject SEMs for the differences between numbers (see Franz & Loftus, 2012; Loftus & Masson, 1994).

This may just be a matter of a lack of statistical sensitivity to detect relatively small differences; indeed, a Bayesian analysis (Dienes, 2011) reported just that, with Bayes factors close to 1 for comparisons between notations. It is certainly true that a USB-standard keyboard like we used is not ideal to measure RTs. The temporal precision of such devices is not very good, so that measurements may be off by typically around 30 ms, but up to 70 ms (Shimizu, 2002). Indeed,

our data were rather noisy, as indicated by the fact that although visual inspection shows a good fit (see Figure 7), the R^2 were not impressively high, (see article A.3, Table 1).

So what do these results leave us with? Going back to the initial question of whether spatial properties known from Arabic digits hold for other notations, the evidence indicates that they do. At the same time, our data very much leave open the question of whether the effect was any different in nonsymbolic as opposed to symbolic stimuli.

3.3 Modelling the MNL: Study 3

In study 2, we found corroborating evidence for the ubiquity of spatial-numerical associations, emphasising the usefulness of the MNL metaphor. In a third study, we sought to investigate the mechanisms behind the number line's nonlinearity more closely.

To do this, we tested different modalities of input and output to get a better sense of which part of the input-output transformation gives rise to the characteristic number line shape, as well as testing for interaction effects between different modalities. Participants completed both a standard number estimation task, as well as a number line task, presented in different versions. Our stimuli consisted of Arabic digits, as well as clouds of dots where participants whose numerosity participants were asked to indicate. Feedback and stimuli were manipulated to highlight between-modality effects.

3.3.1 Experimental design

We conducted a total of five experiments. The first four served to investigate properties of a digital number line task, similar to the ones used in several previous studies (Arrighi et al., 2014;

Dehaene et al., 2008; Opfer, Siegler, & Young, 2011). In line with Dehaene's (1992) idea of a non-veridical response grid, we tested how Arabic digits would be mapped onto such a number line (experiment 1; N=6), expecting a near-perfect linear relationship in this rather simple task (see Anobile, Cicchini, & Burr, 2012). In three further experiments, we investigated how nonsymbolic stimuli – clouds of dots of varying visual features (Gebuis & Reynvoet, 2011) – would be mapped to the same number line. This was tested in a task in which participants were presented with a number line bounded by traditional Arabic digits (experiment 2; N=8), as well as two further ones with a number line bounded by nonsymbolic magnitudes (experiment 3; N=8), and nonsymbolic magnitudes in a random order and with a random starting position, to prevent participants from learning motions instead of magnitude mapping (experiment 4; N=8). In each of experiments 2...4, mapping Arabic digits to the number line was included as a separate control condition, and in intermixed blocks to shed light on another important question: The question of whether, as proposed by Burr and Ross (2008; J. Ross & Burr, 2010) as well as Gebuis and colleagues (Gebuis, Gevers, & Cohen Kadosh, 2014), visual number is indeed better understood in terms of a sensory process that has little to do with semantic encoding – in which case we would predict no interaction between trials with symbolic stimuli and nonsymbolic stimuli.

We also conducted a fifth experiment (N=36) for a more severe test of this relationship. In this experiment, we tested participants on both a number line task like in experiments 1...4, and on a classic magnitude estimation task in which they typed the number of dots they estimated to be in a cloud of dots. The presented magnitudes were all nonsymbolic. The key manipulation was that in this experiment, feedback was presented that was either veridical or systematically distorted.

This was done independently for both tasks, such that any effect that feedback for previous trials would have on a specific *response* would not transfer between tasks, but effects on the *representation* and mapping of magnitudes would.

3.3.2 Modeling the number line

In section 3.1, I have discussed several versions of the MNL as a representation of approximate magnitude that have been proposed to explain the shape of observed response function to nonsymbolic magnitudes. A schematic overview of how these internal representations could be conceptualised, along with simulated data that would follow from each model can be seen in Figure 8. However, looking at the bottom row (simulated responses) makes obvious the biggest problem in telling apart these models, which is that the predicted response functions are quite similar. This makes sense, of course, seeing that all models were derived from similar sets of observed data. At the same time, it emphasises the need for experimental designs that may enable us to tell apart the influence of different subcomponents of tasks that typically produce these response functions.

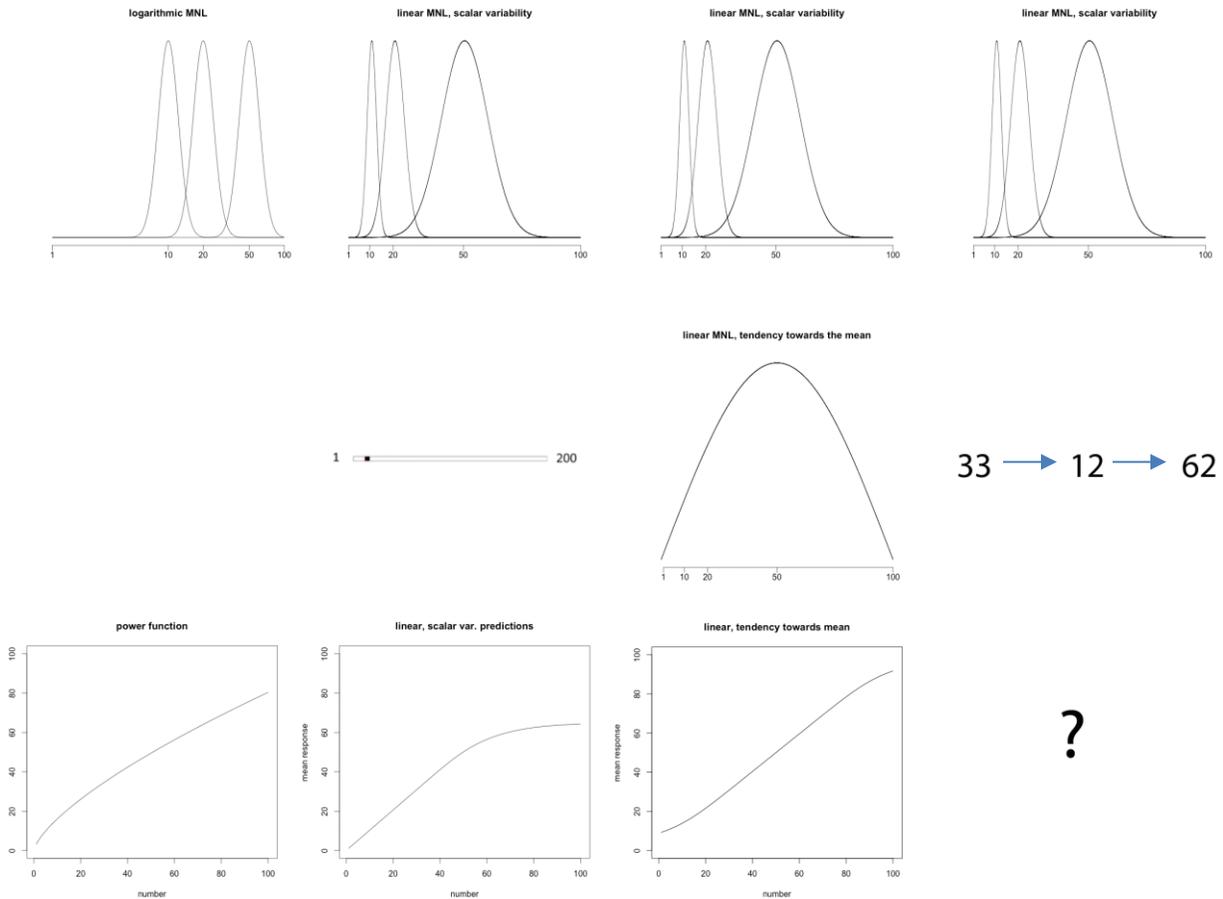


Figure 8: Models of the MNL, and the corresponding response functions. Models of internal magnitude representation (top row), response biases (middle row) and predicted response functions (bottom row). **a:** A logarithmic MNL with a nonlinear response grid (as posited by Dehaene, 1992), resulting in a power function with an exponent < 1 . **b:** Linear MNL with scalar variability and a bounded response range. **c:** Linear MNL with scalar variability and a tendency towards the mean. **d:** Linear MNL, scalar variability, and a response that is dependent on previous trials. This makes the response function dependent on previous trials and not clear what to expect.

In our analyses of experiments 1...4, we fit the responses from each task to three basic models (roughly followed guidelines by Knoblauch & Maloney, 2012): A linear model of the form $y = a + b \cdot x$, a logarithmic model of the form $y = a + b \cdot \log(x)$, and a power function model of the form $y = a \cdot x^b$ (fit through a linear model $\log(y) \sim \log(x)$ with a fixed intercept of 0 to conform with standard models of sensory matching, Stevens, 1957, 1959; Teghtsoonian, 1971). Thus, each model had two free parameters. These models were fit on a by-number basis, that is with responses

aggregated over trials and participants by numbers, for the sake of comparability with other studies. To investigate the effect of previous trials, we also compared these simpler models to a more complex model with the *previous magnitude* as an added linear predictor, as well as an interaction term of *previous magnitude*previous type* (coded as ‘same’ or ‘different’ relative to the current trialtype) to investigate if the effect would be moderated by whether the previous trial was of the same type. Here, we used Akaike’s Information Criterion (AIC; Akaike, 1974) as a means of comparing the explanatory power, since more complex models may explain more variance even when they do not add explanatory power, and indeed nested models will always explain at least the same amount of variance.

This was all done in addition to a simpler approach of running a simple repeated-measures ANOVA on relative error (defined as $(response(x)-x)/x$) with the factors oddball and block type and, in the case of experiment 5, two ‘feedback’ factors for verbal and nonverbal feedback. These ANOVAs were then followed up with t-tests to examine the effects more closely.

3.3.3 Results and discussion

Fitting the models revealed three main findings. First, and unsurprisingly, the responses to symbolic stimuli were fitted best (and indeed almost perfectly) by a linear function. In all experiments, this was a function with a slope slightly larger than 1 and a negative y-intercept, indicating that while there was not a general bias in estimating the location of the stimuli, differences tended to be slightly exaggerated. Importantly, however, this finding also underlines that there is nothing inherent to this task that would produce a nonlinear response function.

Second, nonsymbolic stimuli were usually fitted best by a power function with an exponent smaller than 1, indicating the typical nonlinear shape with an overestimation in lower ranges, and underestimation for relatively large stimuli. It should be noted that in experiment 2, it was in fact the linear function that fit the data best, and differences in fit were quite small, so that this analysis can hardly serve as strong evidence of one model over another.

Finally, in all but one experiment we found an improved fit (indicated by a negative AIC) when previous magnitude was included as a predictor, and in all experiments where mixed blocks were presented this fit was then improved further by including a previous magnitude*previous type-interaction term. Previous magnitude generally was assigned a positive weight when trial type was the same, but not when a nonsymbolic trial was preceded by a symbolic trial or vice-versa. The last point is illustrated further by the ANOVA we ran on relative error with the three-level factor oddball. Results from this analysis show that there was indeed an effect of oddball on relative error in all experiments, but any effect of oddball was qualified by an oddball*block type interaction that was indicative of an oddball effect in homogenous, but not in mixed blocks, as running post-hoc t-tests showed. We also see no evidence for any transfer of calibration between tasks in experiment 5. This would have been expected if feedback had indeed functioned as a calibration of the input-to-representation mapping. Instead, it looks like it was the response that was calibrated.

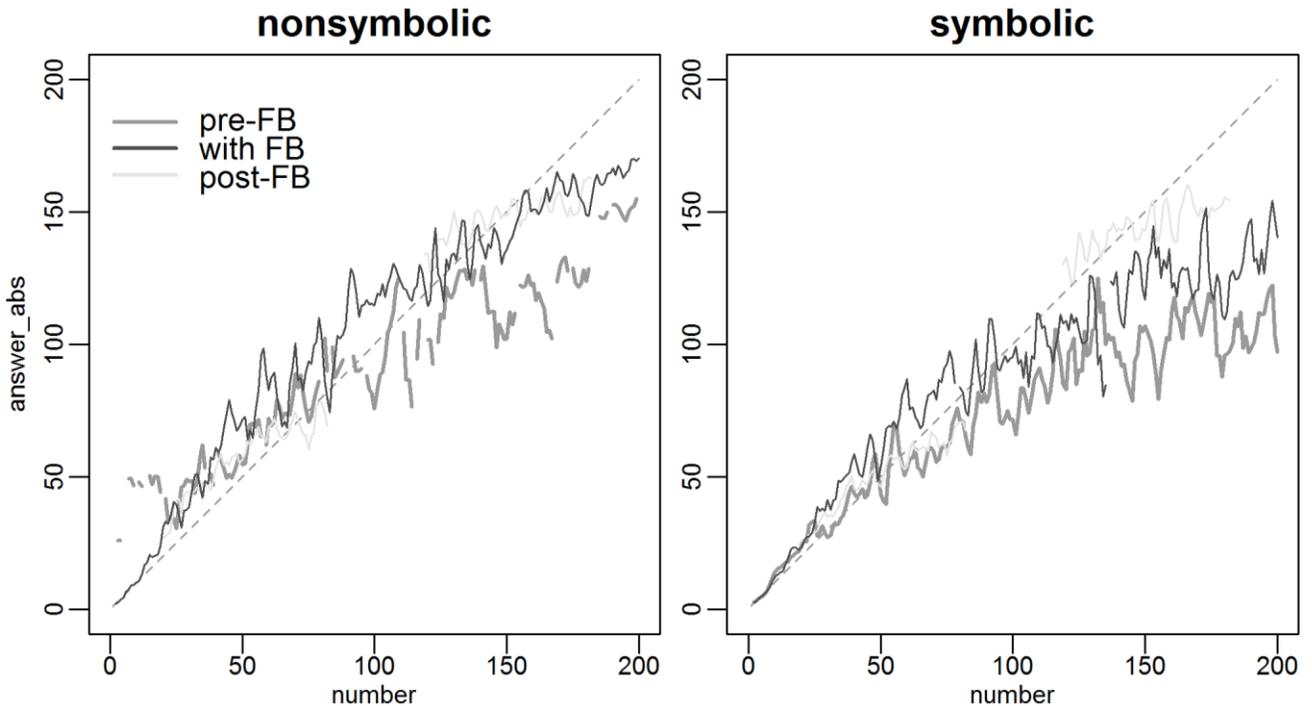


Figure 9: Nonsymbolic (number line) and symbolic (verbal) number-line responses, and the effect of veridical feedback. Data from experiment 5 of study 3, veridical-feedback condition. Moving average (± 1 cell) applied. **Left:** Number line responses, **right:** Verbal responses. In both cases, responses during and after feedback phase are closer to the veridical dashed line, but still exhibit the typical nonlinear shape.

In summary, these results do not support the notion of a dynamic response to a general magnitude system, affected by feedback or through adapting responses to previous trials. Instead, we see effects of nonlinearity for nonsymbolic input that are independent of the task, as well as feedback given in experiment 5.

3.4 The MNL – what do we know?

Our results clearly indicate that the spatial associations and nonlinear response functions observed in magnitude tasks are robust phenomena that are not artifacts of tasks, nor of stimulus properties. A spatial-numerical association of response codes is as ubiquitous as previously thought and does not, as had been proposed (Hung et al., 2008) have exceptions to the rule based

on visual properties, although differences remain between notations, and it is not obvious what they tell us. Similarly, we found that response functions in judgements of nonsymbolic magnitudes followed the typical power function, which was influenced (but not brought about) by previous trial magnitude and, importantly, was relatively robust to feedback and calibration.

A number of theories have tried to explain these phenomena. Some are incompatible with our data. For example, our results show that a nonlinear shape could be observed in the absence of Weber's law (see figure 4 of article A.4), as well as in an unbounded task that included veridical feedback. Thus, they cannot be caused by a linear representation by simply assuming an increase in variability. Similarly, while there is no question that previous trials influence the next trial, nonlinear models achieved a good fit even on a per-trial level without including previous trials as predictors. Furthermore, previous trials of a different type seemed to have no impact on succeeding trials, yet the nonlinear shape did not differ between mixed and homogenous blocks. This would indicate that while previous trials do influence the response function of magnitude estimates, a theory relying on this influence cannot explain its shape.

The model by Dehaene (1992) of a logarithmic mental number line with scalar variability can partially explain our results. This model successfully predicts the shape of the MNL-response function, although it generally assumes that the power-shape (instead of a logarithmic shape) is due to a variable response-grid. However, feedback did nothing to linearize the response function in either task in experiment 5 of study 3. A logarithmic representation would still predict data very similar to ours, and the fit of a power model is not that much better compared to a logarithmic model to conclusively rule out the latter; however, the hypothesis of a variable

response grid as a cause of the power-shape, a central point of Dehaene's theory (Dehaene et al., 2008; Izard & Dehaene, 2008) makes predictions that could not be confirmed. It should also be noted that the response function in the number line task with symbolic stimuli was almost perfectly linear, indicating that any nonlinearity arguably was wholly due to the internal representation. The response-grid, or more generally representation-to-output mapping, may be variable, but arguably not in the way predicted by this model.

Indeed, the model that best predicted our results is that of a general tendency towards the mean. This model predicts the typical a logarithmic or power function (although it could be argued that its predictions may be modeled even better by a logistic model, see Figure 9), which is what we found. Fitting a linear function, a tendency towards the mean would predict a substantially positive intercept with a slope smaller than 1, which is exactly what we find for all responses to nonsymbolic stimuli. As the results showed this pattern not only in number line tasks, but also in the unbounded task in experiment 5, this theory relies strongly on the assumption that participants lean towards responding close to the mean of their previous responses. This is indeed one of the main propositions put forward by Haubensak (1992), and certainly warrants further examination.

Chapter 4 – Summary

We conducted three studies to investigate (1) visual perception of object size, and if this process differs depending on the task, leading to a debate about what results from study 1 can tell us about the streams of visual processing, (2) the ubiquity of an association between space and numbers, and thus the concept of the MNL, (3) what causes the nonlinear shape of the MNL.

These investigations were designed to investigate different aspects of visual and semantic magnitude representation, as well as possible interactions between the two.

Our experiments (study 1; Koppke, Bruno, et al., 2016b) showed that indeed the cortical processing of visual information in the context of a visual illusion appeared to be similar across tasks in our first study. This is not to say that the processing of visual information does not depend on the task. In fact, one of our main arguments is that the comparison between perceptual measures and action measures is so difficult because participants attend to - and thus process - different bits of information in each task (see e.g. Franz et al., 2000), respectively. However, the evidence in favour of a theory of different cortical processing when the same information is presented and attended to is much weaker than sometimes presented (e.g., Goodale, 2014; Westwood & Goodale, 2011). Questions remain as to what this means for the TVSH (Milner & Goodale, 1995, 2006), since we tested and refuted one specific hypothesis of the larger theory, while there is still a large body of corroborating evidence on other predictions. We argue in another article (Koppke, Bruno, et al., 2016a) that our experiment constitutes a strong test of a critical prediction - however, for the discussion to reach a conclusion, both the importance of the illusion literature for the theory, and likely the evidence in other domains would have to be examined, which was beyond the scope of our studies.

We can also say that the MNL is as ubiquitous (see study 2) and its properties possibly even more robust (see study 3) than is sometimes assumed. We found the characteristic association of space and numbers even in notations where this association had previously been questioned (Hung et al., 2008; but see Koppke, Löwenkamp, et al., 2016). This is especially interesting

considering that the effect occurred even when visuospatial properties of these notations made it possible to solve our number-related tasks without accessing numeric magnitude.

With regards to the shape of the MNL, we found the classic power-function shape in both a number line task and a verbal estimation task. These are also classic examples of a bounded task and an unbounded task – we found the same nonlinearity in both, although it would be interesting to compare the shapes more precisely, as bounded responses are known to affect the response function exponent (but not the qualitative shape) in sensory matching (Stevens, 1959; Teghtsoonian, 1971). The nonlinearity also was not caused by sequential dependency effects, and did not seem to be a property of the response in general, as evidenced by the fact that responses to symbolic stimuli were linear, and that the shape of the response function was robust to calibration. In general, there seemed to be little interaction between trials with symbolic stimuli and trials with nonsymbolic stimuli. Responses appeared to be malleable to some degree, but this was specific to a stimulus type. Thus, we did not find much interplay between visual and semantic magnitude information – if anything, our data support theories that assume a separate processing for visual and semantic magnitude information, even when they concern the same type of magnitude.

References

- Aglioti, S., DeSouza, J. F. X., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology*, *5*(6), 679–685.
- Akaike, H. (1974). A new look at the statistical model identification. In *IEEE Transactions on Automatic Control AC-19* (pp. 716–723).
- Anobile, G., Cicchini, G. M., & Burr, D. C. (2012). Linear mapping of numbers onto space requires attention. *Cognition*, *122*(3), 454–459.
<http://doi.org/10.1016/j.cognition.2011.11.006>
- Anobile, G., Cicchini, G. M., & Burr, D. C. (2015). Number As a Primary Perceptual Attribute: A Review. *Perception*. <http://doi.org/10.1177/0301006615602599>
- Arrighi, R., Togoli, I., & Burr, D. C. (2014). A generalized sense of number. *Proceedings of the Royal Society B*, *281*(1797), 20141791. <http://doi.org/10.1098/rspb.2014.1791>
- Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: evidence against a representational shift. *Developmental Science*, *14*(1), 125–135.
<http://doi.org/10.1111/j.1467-7687.2010.00962.x>
- Bridgeman, B., Kirch, M., & Sperling, A. (1981). Segregation of cognitive and motor aspects of visual function using induced motion. *Perception & Psychophysics*, *29*(4), 336–342.
<http://doi.org/10.3758/BF03207342>
- Bruno, N., & Franz, V. H. (2009). When is grasping affected by the Müller-Lyer illusion? A quantitative review. *Neuropsychologia*, *47*, 1421–1433.

- Bruno, N., Uccelli, S., Viviani, E., & De'Sperati, C. (2016). Both vision-for-perception and vision-for-action follow Weber's law at small object sizes, but violate it at larger sizes. *Neuropsychologia*, *91*, 327–334. <http://doi.org/10.1016/j.neuropsychologia.2016.08.022>
- Bull, R., Blatto-Valle, G., & Fabich, M. (2006). Subitizing, magnitude representation, and magnitude retrieval in deaf and hearing adults. *Journal of Deaf Studies and Deaf Education*, *11*(3), 289–302.
- Burr, D., & Ross, J. (2008). A visual sense of number. *Current Biology*, *18*, 425–428.
- Butterworth, B. (1999). A head for figures. *Science*, *284*, 928–929.
- Cantlon, J. F., Cordes, S., Libertus, M. E., & Brannon, E. M. (2009). Comment on “Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures.” *Science*, *323*, 38.
- Cao, B., Li, F., & Li, H. (2010). Notation-dependent processing of numerical magnitude: Electrophysiological evidence from Chinese numerals. *Biological Psychology*, *83*, 47–55. <http://doi.org/10.1016/j.biopsycho.2009.10.003>
- Cavina-Pratesi, C., Monaco, S., Fattori, P., Galletti, C., McAdam, T. D., Quinlan, D. J., ... Culham, J. C. (2010). Functional magnetic resonance imaging reveals the neural substrates of arm transport and grip formation in reach-to-grasp actions in humans. *Journal of Neuroscience*, *30*(31), 10306–10323. <http://doi.org/10.1523/JNEUROSCI.2023-10.2010>
- Chambers, C. D. (2013). Registered Reports: A new publishing initiative at Cortex. *Cortex*, *49*, 609–610.

- Chambers, C. D., Dienes, Z., McIntosh, R. D., Rotshtein, P., & Willmes, K. (2015). Registered reports: Realigning incentives in scientific publishing. *Cortex*, *66*, A1–A2.
- Chater, N. (2003). How much can we learn from double dissociations? *Cortex*, *39*, 167–169.
- Chinello, A., de Hevia, M. D., Geraci, C., & Girelli, L. (2012). Finding the spatial-numerical association of response codes (SNARC) in signed numbers: notational effects in accessing number representation. *Functional Neurology*, *27*(3), 177–185.
- Chiou, R. Y. C., Wu, D. H., Tzeng, O. J. L., Hung, D. L., & Chang, E. C. (2012). Relative size of numerical magnitude induces a size-contrast effect on the grip scaling of reach-to-grasp movements. *Cortex*, *48*(8), 1043–1051. <http://doi.org/10.1016/j.cortex.2011.08.001>
- Clarivate Analytics. (2016). Web of Science. Retrieved October 4, 2016, from <http://isiknowledge.com>
- Cohen Kadosh, R., & Walsh, V. (2009). Numerical representation in the parietal lobes: Abstract or not abstract? *Behavioral and Brain Sciences*, *32*, 313–373.
- Coren, S., & Enns, J. T. (1993). Size contrast as a function of conceptual similarity between test and inducers. *Perception & Psychophysics*, *54*(5), 579–588.
- Coren, S., & Girgus, J. S. (1972). A comparison of five methods of illusion measurement. *Behavior Research Methods & Instrumentation*, *4*(5), 240–244. <http://doi.org/10.3758/BF03210006>
- de Grave, D. D. J., Biegstraaten, M., Smeets, J. B. J., & Brenner, E. (2005). Effects of the Ebbinghaus figure on grasping are not only due to misjudged size. *Experimental Brain*

Research, 163, 58–64.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, 44, 1–42.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences*, 7(4), 145–147. [http://doi.org/10.1016/S1364-6613\(03\)00055-X](http://doi.org/10.1016/S1364-6613(03)00055-X)

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

Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or Linear? Distinct Intuitions of the Number Scale in Western Amazonian Indigene Cultures. *Science*, 320, 1217–1220. <http://doi.org/10.1126/science.1156540>

Dehaene, S., Molko, N., Cohen, L., & Wilson, A. J. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology*, 14, 218–224.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, 20, 487–506.

Dewar, M. T., & Carey, D. P. (2006). Visuomotor “immunity” to perceptual illusion: A mismatch of attentional demands cannot explain the perception-action dissociation. *Neuropsychologia*, 44(8), 1501–1508. <http://doi.org/10.1016/j.neuropsychologia.2005.11.010>

Dienes, Z. (2008). *Understanding psychology as a science. An introduction to scientific and statistical inference*. New York: Palgrave Macmillan.

Dienes, Z. (2011). Bayesian versus Orthodox statistics: Which side are you on? *Perspectives on*

Psychological Science, 6(3), 274–290.

Domahs, F., Klein, E., Moeller, K., Nuerk, H. C., Yoon, B. C., & Willmes, K. (2012).

Multimodal semantic quantity representations: Further evidence from Korean sign language.

Frontiers in Psychology, 3, 1–10. <http://doi.org/10.3389/fpsyg.2011.00389>

Dwan, K., Gamble, C., Williamson, P. R., & Kirkham, J. J. (2013). Systematic review of the

empirical evidence of study publication bias and outcome reporting bias - an updated

review. *PLoS ONE*, 8(7), e66844. <http://doi.org/10.1371/journal.pone.0066844>

Eloka, O., Feuerhake, F., Janczyk, M., & Franz, V. H. (2015). Garner-Interference in left-handed

awkward grasping. *Psychological Research*, 79(4), 579–589. [http://doi.org/10.1007/s00426-](http://doi.org/10.1007/s00426-014-0585-1)

014-0585-1

Ende, M. (1960). *Jim Knopf und Lukas, der Lokomotivfuehrer*. Stuttgart, Germany: Thienemann.

Fechner, G. T. (1860). *Elemente der Psychophysik*. Leipzig: Breitkopf und Haertel.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive*

Sciences, 8(7), 307–314. <http://doi.org/10.1016/j.tics.2004.05.002>

Fias, W. (2001). Two routes for the processing of verbal numbers: Evidence from the SNARC

effect. *Psychological Research*, 65, 250–259. <http://doi.org/10.1007/s004260100065>

Fias, W., Brysbaert, M., Geypens, F., & d’Ydewalle, G. (1996). The Importance of magnitude

information in numerical processing: Evidence from the SNARC effect. *Mathematical*

Cognition, 2(1), 95–110.

Fischer, M. H., Shaki, S., & Cruise, A. (2009). It takes just one word to quash a SNARC.

Experimental Psychology, 56(5), 361–366.

Foster, R. M., & Franz, V. H. (2014). Superadditivity of the Ebbinghaus and Müller-Lyer illusions depends on the method of comparison used. *Perception*, 43(8), 783–795.
<http://doi.org/10.1068/p7802>

Francis, G. (2013a). Replication, statistical consistency, and publication bias. *Journal of Mathematical Psychology*, 57(5), 153–169. <http://doi.org/10.1016/j.jmp.2013.02.003>

Francis, G. (2013b). We don't need replication, but we do need more data. *European Journal of Personality*, 27, 125–126.

Franz, V. H. (2003). Manual size estimation: a neuropsychological measure of perception? *Experimental Brain Research*, 151, 471–477.

Franz, V. H. (2004). Is there a dynamic illusion effect in the motor system? *Behavioral and Brain Sciences*, 27(1), 34–35.

Franz, V. H. (2007). *Ratios: A short guide to confidence limits and proper use*. Retrieved from <http://arxiv.org/pdf/0710.2024v1.pdf>

Franz, V. H., Bühlhoff, H. H., & Fahle, M. (2003). Grasp effects of the Ebbinghaus illusion: obstacle avoidance is not the explanation. *Experimental Brain Research*, 149, 470–477.

Franz, V. H., Fahle, M., Bühlhoff, H. H., & Gegenfurtner, K. R. (2001). Effects of visual illusions on grasping. *Journal of Experimental Psychology: Human Perception and Performance*, 27(5), 1124–1144.

Franz, V. H., & Gegenfurtner, K. R. (2008). Grasping visual illusions: Consistent data and no

- dissociation. *Cognitive Neuropsychology*, *25*, 920–950.
- Franz, V. H., Gegenfurtner, K. R., Bühlhoff, H. H., & Fahle, M. (2000). Grasping visual illusions: No evidence for a dissociation between perception and action. *Psychological Science*, *11*(1), 20–25.
- Franz, V. H., Hesse, C., & Kollath, S. (2009). Visual illusions, delayed grasping, and memory: No shift from dorsal to ventral control. *Neuropsychologia*, *47*, 1518–1531.
- Franz, V. H., & Loftus, G. R. (2012). Standard errors and confidence intervals in within-subjects designs: Generalizing Loftus and Masson (1994) and avoiding the biases of alternative accounts. *Psychonomic Bulletin & Review*, *19*, 395–404.
- Franz, V. H., Scharnowski, F., & Gegenfurtner, K. R. (2005). Illusion effects on grasping are temporally constant not dynamic. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(6), 1359–1378.
- Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*, 43–74.
- Ganel, T., Chajut, E., & Algom, D. (2008). Visual coding for action violates fundamental psychophysical principles. *Current Biology*, *18*(14), R599–R601.
- Ganel, T., & Goodale, M. A. (2003). Visual control of action but not perception requires analytical processing of object shape. *Nature*, *426*, 664–667.
- Ganel, T., Tanzer, M., & Goodale, M. A. (2008). A double dissociation between action and perception in the context of visual illusions: Opposite effects of real and illusory size.

Psychological Science, 19(3), 221–225.

Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes.

Cognitive Psychology, 8(1), 98–123. [http://doi.org/10.1016/0010-0285\(76\)90006-2](http://doi.org/10.1016/0010-0285(76)90006-2)

Gebuis, T., Gevers, W., & Cohen Kadosh, R. (2014). Topographic representation of high-level cognition: Numerosity or sensory processing? *Trends in Cognitive Sciences*, 18(1), 1–3.

Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior Research Methods*, 43(4), 981–986. <http://doi.org/10.3758/s13428-011-0097-5>

Gelman, A., & Loken, E. (2014). The AAA tranche of subprime science. *Chance*, 27(1), 51–56. <http://doi.org/10.1080/09332480.2014.890872>

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequence is spatially organized. *Cognition*, 87, B87–B95.

Girgus, J. S., Coren, S., & Agdern, M. (1972). The interrelationship between the Ebbinghaus and Delboeuf illusions. *Journal of Experimental Psychology*, 95(2), 453–455.

Glover, S. (2004). Separate visual representations in the planning and control of action. *Behavioral and Brain Sciences*, 27(1), 3–24.

Glover, S., & Dixon, P. (2002a). Dynamic effects of the Ebbinghaus illusion in grasping: Support for a planning/control model of action. *Perception & Psychophysics*, 64(2), 266–278.

Glover, S., & Dixon, P. (2002b). Semantics affect the planning but not the control of grasping. *Experimental Brain Research*, 146, 383–387.

Gonzalez, C. L. R., Ganel, T., Whitwell, R. L., Morrissey, B., & Goodale, M. A. (2008). Practice

makes perfect, but only with the right hand: Sensitivity to perceptual illusions with awkward grasps decreases with practice in the right but not the left hand. *Neuropsychologia*, 46, 624–631.

Goodale, M. A. (2008). Action without perception in human vision. *Cognitive Neuropsychology*, 25(7–8), 891–919.

Goodale, M. A. (2014). How (and why) the visual control of action differs from visual perception. *Proceedings of the Royal Society B*, 281, 20140337.
<http://doi.org/http://dx.doi.org/10.1098/rspb.2014.0337>

Goodale, M. A., Meenan, J. P., Bühlhoff, H. H., Nicolle, D. A., Murphy, K. J., & Racicot, C. I. (1994). Separate neural pathways for the visual analysis of object shape in perception and prehension. *Current Biology*, 4(7), 604–610.

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

Goodale, M. A., & Milner, A. D. (2010). Two visual streams: Interconnections do not imply duplication of functions. *Cognitive Neuroscience*, 1(1), 65–68.

Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, 349, 154–156.

Goodale, M. A., Pelisson, D., & Prablanc, C. (1986). Large adjustments in visually guided reaching do not depend on vision of the hand or perception of target displacement. *Nature*, 320, 748–750. <http://doi.org/10.1038/320748a0>

- Grill-Spector, K. (2003). The neural basis of object perception. *Current Opinion in Neurobiology*, *13*(2), 159–166. [http://doi.org/10.1016/S0959-4388\(03\)00040-0](http://doi.org/10.1016/S0959-4388(03)00040-0)
- Haffenden, A. M., & Goodale, M. A. (1998). The effect of pictorial illusion on prehension and perception. *Journal of Cognitive Neuroscience*, *10*(1), 122–136.
- Haffenden, A. M., & Goodale, M. A. (2000). Independent effects of pictorial displays on perception and action. *Vision Research*, *40*, 1597–1607.
- Haffenden, A. M., Schiff, K. C., & Goodale, M. A. (2001). The dissociation between perception and action in the Ebbinghaus illusion: Nonillusory effects of pictorial cues on grasp. *Current Biology*, *11*, 177–181.
- Haubensak, G. (1992). The consistency model: A process model for absolute judgements. *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 303–309.
- Herrera, A., Macizo, P., & Semenza, C. (2008). The role of working memory in the association between number magnitude and space. *Acta Psychologica*, *128*, 225–237.
<http://doi.org/10.1016/j.actpsy.2008.01.002>
- Hesse, C., Ball, K., & Schenk, T. (2014). Pointing in visual periphery: Is DF's dorsal stream intact? *PLoS ONE*, *9*(3), e91420. <http://doi.org/10.1371/journal.pone.0091420>
- Himmelbach, M., Boehme, R., & Karnath, H. O. (2012). 20 years later: A second look on DF's motor behaviour. *Neuropsychologia*, *50*, 139–144.
- Himmelbach, M., & Karnath, H. O. (2005). Dorsal and ventral stream interaction: Contributions from optic ataxia. *Journal of Cognitive Neuroscience*, *17*(4), 632–640.

- Hubel, D. H., & Wiesel, T. N. (1972). Laminar and columnar distribution of geniculo-cortical fibers in the macaque monkey. *The Journal of Comparative Neurology*, *146*(4), 421–450.
<http://doi.org/10.1002/cne.901460402>
- Hung, Y., Hung, D. L., Tzeng, O. J. L., & Wu, D. H. (2008). Flexible spatial mapping of different notations of numbers in Chinese readers. *Cognition*, *106*, 1441–1450.
- Hutchinson, S., & Louwerse, M. M. (2014). Language statistics explain the spatial-numerical association of response codes. *Psychonomic Bulletin & Review*, *21*(2), 470–478.
<http://doi.org/10.3758/s13423-013-0492-2>
- Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition*, *32*(4), 662–673.
- Iversen, W., Nuerk, H.-C., Jäger, L., & Willmes, K. (2006). The influence of an external symbol system on number parity representation, or What's odd about 6? *Psychonomic Bulletin & Review*, *13*(4), 730–736.
- Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*(3), 1221–1247.
- James, T. W., Culham, J., Humphrey, G. K., Milner, A. D., & Goodale, M. A. (2003). Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study. *Brain*, *126*, 2463–2475.
- Jeannerod, M. (1984). The timing of natural prehension movements. *Journal of Motor Behavior*, *16*(3), 235–254.

- Jeannerod, M. (1986). The formation of finger grip during prehension. A cortically mediated visuomotor pattern. *Behavioural Brain Research*, *19*, 99–116.
- Knoblauch, K., & Maloney, L. T. (2012). *Modeling psychophysical data in R*. New York, NY: Springer.
- Knol, H., Huys, R., Sarrazin, J. C., & Jirsa, V. K. (2015). Quantifying the Ebbinghaus figure effect: Target size, context size, and target-context distance determine the presence and direction of the illusion. *Frontiers in Psychology*, *6*, 1–11.
<http://doi.org/10.3389/fpsyg.2015.01679>
- Kopiske, K. K., Bruno, N., Hesse, C., Schenk, T., & Franz, V. H. (2016a). Do visual illusions affect grasping? Considerable progress in a scientific debate. *Cortex*.
- Kopiske, K. K., Bruno, N., Hesse, C., Schenk, T., & Franz, V. H. (2016b). The functional subdivision of the visual brain: Is there a real illusion effect on action? A multi-lab replication study. *Cortex*, *79*, 130–152. <http://doi.org/10.1016/j.cortex.2016.03.020>
- Kopiske, K. K., Löwenkamp, C., Eloka, O., Schiller, F., Kao, C. S., Wu, C., ... Franz, V. H. (2016). The SNARC effect in Chinese numerals: Do visual properties of characters and hand signs influence number processing? *PLoS One*, *11*(9), e0163897.
<http://doi.org/10.1371/journal.pone.0163897>
- Liu, C., Tang, H., Luo, Y., & Mai, X. (2011). Multi-representation of symbolic and nonsymbolic numerical magnitude in Chinese number processing. *PLoS One*, *6*(4), e19373.
- Livingstone, M. S., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the

- perception of form, color, movement, and depth. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 7(11), 3416–3468.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1, 476–490.
- Löwenkamp, C., Gärtner, W., Haus, I. D., & Franz, V. H. (2015). Semantic grasping escapes Weber's law. *Neuropsychologia*, 70, 235–245.
<http://doi.org/10.1016/j.neuropsychologia.2015.02.037>
- Mayo, D. G., & Spanos, A. (2011). Error statistics. In P. S. Bandyopadhyay & M. R. Forster (Eds.), *Philosophy of statistics* (Vol. 7, pp. 153–198). Amsterdam, NL: Elsevier.
<http://doi.org/10.1016/B978-0-444-51862-0.50005-8>
- McCloskey, M. (2003). Beyond Task Dissociation Logic: A Richer Conception of Cognitive Neuropsychology. *Cortex*, 39(1), 196–202. [http://doi.org/10.1016/S0010-9452\(08\)70101-1](http://doi.org/10.1016/S0010-9452(08)70101-1)
- McIntosh, R. D., & Lashley, G. (2008). Matching boxes: Familiar size influences action programming. *Neuropsychologia*, 46, 2441–2444.
- Milgram, P. (1987). A spectacle-mounted liquid-crystal tachistoscope. *Behavior Research Methods, Instruments & Computers*, 19(5), 449–456.
- Milner, A. D., & Dyde, R. (2003). Why do some perceptual illusions affect visually guided action, when others don't? *Trends in Cognitive Sciences*, 7(1), 10–11.
- Milner, A. D., Ganel, T., & Goodale, M. A. (2012). Does grasping in patient D.F. depend on vision? *Trends in Cognitive Sciences*, 16(5), 256–257.

- Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action* (1st ed.). Oxford: Oxford University Press Inc.
- Milner, A. D., & Goodale, M. A. (2006). *The visual brain in action* (2nd ed.). Oxford: Oxford University Press Inc.
- Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, *46*, 774–785.
- Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*, 1519–1520.
- Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience*, *9*(3), 429–434.
- Nosek, B. A., & Lakens, D. (2014). Registered reports: A method to increase the credibility of published results. *Social Psychology*, *45*(3), 137–141. <http://doi.org/10.1027/1864-9335/a000192>
- Nuerk, H.-C., Wood, G., & Willmes, K. (2005). The universal SNARC effect. The association between number magnitude and space is amodal. *Experimental Psychology*, *52*(3), 187–194.
- Opfer, J. E., Siegler, R. S., & Young, C. J. (2011). The powers of noise-fitting: Reply to Barth and Paladino. *Developmental Science*, *14*(5), 1194–1204. <http://doi.org/10.1111/j.1467-7687.2011.01070.x>
- Parducci, A., & Perret, L. F. (1971). Category rating scales: Effects of relative spacing and frequency of stimulus values. *Journal of Experimental Psychology*, *89*(2), 427–452.

- Pavani, F., Boscagli, I., Benvenuti, F., Rabuffetti, M., & Farnè, A. (1999). Are perception and illusion affected differently by the Titchener circles illusion? *Experimental Brain Research*, *127*, 95–101.
- Pfister, R., Schroeder, P. A., & Kunde, W. (2013). SNARC struggles: Instant control over spatial-numerical associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1953–1958.
- Pinhas, M., Tzelgov, J., & Ganor-Stern, D. (2012). Estimating linear effects in ANOVA designs: The easy way. *Behavior Research Methods*, *44*, 788–794.
- Pisella, L., Binkofski, F., Lasek, K., Toni, I., & Rossetti, Y. (2006). No double-dissociation between optic ataxia and visual agnosia: Multiple sub-streams for multiple visuo-manual integrations. *Neuropsychologia*, *44*, 2734–2748.
- Post, R. B., & Welch, R. B. (1996). Is there dissociation of perceptual and motor responses to figural illusions? *Perception*, *25*(5), 569–581.
- Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychological Bulletin*, *132*(3), 416–442.
- Roberts, B., Harris, M. G., & Yates, T. A. (2005). The roles of inducer size and distance in the Ebbinghaus illusion (Titchener circles). *Perception*, *34*, 847–856.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641.
- Ross, A. I., Schenk, T., Billino, J., Macleod, M. J., & Hesse, C. (2016). Avoiding unseen

- obstacles: Subcortical vision is not sufficient to maintain normal obstacle avoidance behaviour during reaching. *Cortex*. <http://doi.org/10.1016/j.cortex.2016.09.010>
- Ross, J., & Burr, D. C. (2010). Vision senses number directly. *Journal of Vision*, *10*(2), 1–8. <http://doi.org/10.1167/10.2.10>
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*, 225–237.
- Santens, S., & Gevers, W. (2008). The SNARC effect does not imply a mental number line. *Cognition*, *108*, 263–270.
- Santiago, J., & Lakens, D. (2015). Can conceptual congruency effects between number, time, and space be accounted for by polarity correspondence? *Acta Psychologica*, *156*, 179–191. <http://doi.org/10.1016/j.actpsy.2014.09.016>
- Schenk, T. (2006). An allocentric rather than perceptual deficit in patient D.F. *Nature Neuroscience*, *9*(11), 1369–1370.
- Schenk, T. (2010). Visuomotor robustness is based on integration not segregation. *Vision Research*, *50*, 2627–2632.
- Schenk, T. (2012). No dissociation between perception and action in patient DF when haptic feedback is withdrawn. *The Journal of Neuroscience*, *32*(6), 2013–2017.
- Schenk, T., Franz, V. H., & Bruno, N. (2011). Vision-for-perception and vision-for-action: Which model is compatible with the available psychophysical and neuropsychological data? *Vision Research*, *51*, 812–818.

- Schenk, T., & McIntosh, R. D. (2010). Do we have independent visual streams for perception and action? *Cognitive Neuroscience*, *1*(1), 52–62.
- Schiller, F., Eloka, O., & Franz, V. H. (2016). Using key distance to evaluate theories on the SNARC. *Perception*, *45*, 196–221.
- Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers contribute to the SNARC effect. *Psychonomic Bulletin & Review*, *16*(2), 328–331.
- Shimizu, H. (2002). Measuring keyboard response delays by comparing keyboard and joystick inputs. *Behavior Research Methods, Instruments, & Computers*, *34*(2), 250–256.
<http://doi.org/10.3758/BF03195452>
- Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychological Science*, *14*(3), 237–243.
<http://doi.org/10.1111/1467-9280.02438>
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366. <http://doi.org/10.1177/0956797611417632>
- Smeets, J. B. J., & Brenner, E. (1999). A new view on grasping. *Motor Control*, *3*(3), 237–271.
- Smeets, J. B. J., & Brenner, E. (2006). 10 Years of Illusions. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1501–1504.
- Smeets, J. B. J., Glover, S., & Brenner, E. (2003). Modeling the time-dependent effect of the Ebbinghaus illusion on grasping. *Spatial Vision*, *16*(3–4), 311–324.

- Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—Or vice versa. *Journal of the American Statistical Association*, *54*, 30–34.
- Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician*, *49*(1), 108–112. <http://doi.org/10.2307/2684823>
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680.
- Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153–181.
- Stevens, S. S. (1959). Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *Journal of Experimental Psychology*, *57*(4), 201–209.
- Stöttinger, E., Pfusterschmied, J., Wagner, H., Danckert, J., Anderson, B., & Perner, J. (2012). Getting a grip on illusions: replicating Stöttinger et al [Exp Brain Res (2010) 202:79–88] results with 3-D objects. *Experimental Brain Research*, *216*(1), 155–157.
<http://doi.org/10.1007/s00221-011-2912-8>
- Striemer, C. L., Chapman, C. S., & Goodale, M. A. (2009). “Real-time” obstacle avoidance in the absence of primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(37), 15996–16001. <http://doi.org/10.1073/pnas.0905549106>
- Teghtsoonian, R. (1971). On the exponents in Stevens’ law and the constant in Ekman’s law. *Psychological Review*, *78*(1), 71–80. <http://doi.org/10.1037/h0030300>
- Teuber, H. L. (1955). Physiological Psychology. *Annual Review of Psychology*, *6*, 267–296.

- Thompson, A. A., & Westwood, D. A. (2007). The hand knows something that the eye does not: Reaching movements resist the Müller–Lyer illusion whether or not the target is foveated. *Neuroscience Letters*, *426*(2), 111–116.
- Tzelgov, J., Zohar-Shai, B., & Nuerk, H.-C. (2013). On defining quantifying and measuring the SNARC effect. *Frontiers in Psychology*, *4*, 302.
- Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In J. Ingle, M. A. Goodale, & J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.
- Utz, K. S., Hesse, C., Aschenneller, N., & Schenk, T. (2015). Biomechanical factors may explain why grasping violates weber’s law. *Vision Research*, *111*(Part A), 22–30.
<http://doi.org/10.1016/j.visres.2015.03.021>
- van Dijck, J. P., Gevers, W., & Fias, W. (2009). Numbers are associated with different types of spatial information depending on the task. *Cognition*, *113*(2), 248–253.
<http://doi.org/10.1016/j.cognition.2009.08.005>
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Why psychologists must change the way they analyze their data: The case of psi: Comment on Bem (2011). *Journal of Personality and Social Psychology*, *100*, 426–432.
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, *7*, 627–633.

- Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends in Cognitive Sciences*, 7(11), 483–488.
- Weiskrantz, L. (1990). Outlooks for blindsight: Explicit methodologies for implicit processes. *Proceedings of the Royal Society B*, 239, 247–278.
- Westwood, D. A., & Goodale, M. A. (2003). Perceptual illusion and the real-time control of action. *Spatial Vision*, 16(3–4), 243–254. <http://doi.org/10.1163/156856803322467518>
- Westwood, D. A., & Goodale, M. A. (2011). Converging evidence for diverging pathways: Neuropsychology and psychophysics tell the same story. *Vision Research*, 51, 804–811.
- Westwood, D. A., Heath, M., & Roy, E. A. (2000). The effect of a pictorial illusion on closed-loop and open-loop prehension. *Experimental Brain Research*, 134, 456–463.
- Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E. J. (2011). Statistical evidence in experimental psychology: An empirical comparison using 855 t Tests. *Perspectives on Psychological Science*, 6(3), 291–298.
- Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, 10(2), 130–137.
- Whitwell, R. L., & Buckingham, G. (2013). Reframing the action and perception dissociation in DF: Haptics matters, but how? *Journal of Neurophysiology*, 109, 621–624.
- Whitwell, R. L., Buckingham, G., Enns, J. T., Chouinard, P. A., & Goodale, M. A. (2016). Rapid decrement in the effects of the Ponzo display dissociates action and perception. *Psychonomic Bulletin & Review*, 23(4), 1157–1163. <http://doi.org/10.3758/s13423-015->

Whitwell, R. L., & Goodale, M. A. (2016). Real and illusory issues in the illusion debate (Why two things are sometimes better than one): Commentary on Kopiske et al. (2016). *Cortex*.
<http://doi.org/10.1016/j.cortex.2016.06.019>

Whitwell, R. L., Milner, A. D., Cavina-Pratesi, C., Byrne, C. M., & Goodale, M. A. (2014). DF's visual brain in action: The role of tactile cues. *Neuropsychologia*, *55*, 41–50.
<http://doi.org/http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.019>

Whitwell, R. L., Milner, A. D., & Goodale, M. A. (2014). The two visual systems hypothesis: New challenges and insights from visual form agnostic patient DF. *Frontiers in Neurology*, *5*(DEC), 1–8. <http://doi.org/10.3389/fneur.2014.00255>

Wood, G., Willmes, K., Nuerk, H.-C., & Fischer, M. H. (2008). On the cognitive link between space and number: a meta-analysis of the SNARC effect. *Psychology Science Quarterly*, *50*(4), 489–525.

Woodworth, R. S. (1899). The accuracy of voluntary movement. *Psychological Review-Monograph Supplements*, *3*(3), 1–114.

List of published articles

The following four articles have been written on the basis of the experiments described above and submitted to peer-reviewed journals. They have been attached in order as follows.

Article 1: Grasping the Ebbinghaus illusion (published in Cortex)

Kopiske, K. K., Bruno, N., Hesse, C., Schenk, T., & Franz, V. H. (2016b). The functional subdivision of the visual brain: Is there a real illusion effect on action? A multi-lab replication study. *Cortex*, 79, 130-152.

Article 2: Appraising the evidence on grasping illusions (published in Cortex)

Kopiske, K. K., Bruno, N., Hesse, C., Schenk, T., & Franz, V. H. (2016a). Do visual illusions affect grasping? Considerable progress in a scientific debate. A reply to Whitwell & Goodale, 2016. *Cortex*, published online 28 October 2016.

Article 3: The SNARC-effect in Chinese numerals (published in PLoS One)

Kopiske, K. K., Löwenkamp, C., Eloka, O., Schiller, F., Kao, C. S., Wu, C., Gao, X., & Franz, V. H. (2016). The SNARC effect in Chinese numerals: Do visual properties of characters and hand signs influence number processing? *PLoS One*, 11(9), e0163897.

Article 4: Calibrating the mental number line (submitted to Advances in Cognitive Psychology)

Kopiske, K. K., & Franz, V. H. (submitted). Comparing symbolic and nonsymbolic number lines: Consistent effects of notation across output measures.

Appendix: Published articles

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Registered report

The functional subdivision of the visual brain: Is there a real illusion effect on action? A multi-lab replication study



Karl K. Kopiske ^{a,f,*}, Nicola Bruno ^b, Constanze Hesse ^c,
Thomas Schenk ^d and Volker H. Franz ^{a,e}

^a University of Hamburg, Department of Psychology, Hamburg, Germany

^b Dipartimento di Neuroscienze, Università di Parma, Unità di Psicologia, Parma, Italy

^c University of Aberdeen, School of Psychology, Kings College, Old Aberdeen, United Kingdom

^d Ludwig-Maximilians Universität München, Department of Psychology, Munich, Germany

^e University of Tübingen, Department of Computer Science, Experimental Cognitive Science, Tübingen, Germany

^f Center for Neuroscience and Cognitive Systems@UniTn, Istituto Italiano di Tecnologia (IIT), Rovereto, TN, Italy

ARTICLE INFO

Article history:

Protocol Received 11 March 2014

Protocol Accepted 02 September 2014

Received 29 September 2015

Reviewed 12 December 2015

Revised 10 February 2016

Accepted 14 March 2016

Action editor Rob McIntosh

Published online 6 April 2016

Keywords:

Action perception

Visual processing

Illusions

Grasping

Manual size estimation

ABSTRACT

It has often been suggested that visual illusions affect perception but not actions such as grasping, as predicted by the “two-visual-systems” hypothesis of Milner and Goodale (1995, *The Visual Brain in Action*, Oxford University press). However, at least for the Ebbinghaus illusion, relevant studies seem to reveal a consistent illusion effect on grasping (Franz & Gegenfurtner, 2008. Grasping visual illusions: consistent data and no dissociation. *Cognitive Neuropsychology*). Two interpretations are possible: either grasping is not immune to illusions (arguing against dissociable processing mechanisms for vision-for-perception and vision-for-action), or some other factors modulate grasping in ways that mimic a vision-for-perception effect in actions. It has been suggested that one such factor may be obstacle avoidance (Haffenden Schiff & Goodale, 2001. The dissociation between perception and action in the Ebbinghaus illusion: nonillusory effects of pictorial cues on grasp. *Current Biology*, 11, 177–181). In four different labs (total $N = 144$), we conducted an exact replication of previous studies suggesting obstacle avoidance mechanisms, implementing conditions that tested grasping as well as multiple perceptual tasks. This replication was supplemented by additional conditions to obtain more conclusive results. Our results confirm that grasping is affected by the Ebbinghaus illusion and demonstrate that this effect cannot be explained by obstacle avoidance.

© 2016 Elsevier Ltd. All rights reserved.

* Corresponding author. Center for Neuroscience and Cognitive Systems@UniTn, Istituto Italiano di Tecnologia, Corso Bettini, 31, 38068 Rovereto, TN, Italy.

E-mail addresses: karl.kopiske@iit.it (K.K. Kopiske), nicola.bruno@unipr.it (N. Bruno), c.hesse@abdn.ac.uk (C. Hesse), thomas.schenk@psy.lmu.de (T. Schenk), volker.franz@uni-tuebingen.de (V.H. Franz).

<http://dx.doi.org/10.1016/j.cortex.2016.03.020>

0010-9452/© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

1.1. Visual illusions and the two-visual-streams hypothesis (TVSH)

Current theories on the fundamental architecture of the primate brain suggest that there are two functionally and anatomically distinct cortical processing routes for visual information: the dorsal vision-for-action route and the ventral vision-for-perception route. This two-visual-streams hypothesis (TVSH, Goodale & Milner, 1992; Milner & Goodale, 1995, 2006, 2008) is supported by multiple lines of evidence, including evidence from neuropsychology (e.g., action perception-double dissociations after brain damage) and from psychophysics (e.g., action-perception double dissociations in healthy participants responding to visual illusions). Neuropsychological evidence has come from patients with blindsight (Weiskrantz, 1990), optic ataxia (Milner et al., 2001), as well as visual form agnosia (Goodale & Milner, 1992; Goodale, Milner, Jakobson, & Carey, 1991). However, there is an ongoing debate on the question to which degree the neuropsychological data support the TVSH (Milner, Ganel, & Goodale, 2012; Milner & Goodale, 2008; Whitwell, Milner, Cavina-Pratesi, Byrne, & Goodale, 2014), or allow for alternative interpretations (Himmelbach, Boehme, & Karnath, 2012; Schenk, 2006, 2010, 2012). For recent reviews, see Schenk, Franz, and Bruno (2011), Schenk and McIntosh (2010), and Westwood and Goodale (2011). This debate suggests that patient studies may not provide conclusive evidence for the TVSH, so that evidence from healthy participants becomes especially important.

Aglioti, DeSouza, and Goodale (1995) conducted a seminal study that is often cited as key evidence that the TVSH also holds for healthy human observers. In this study they investigated how perception and action are affected by size contrast illusions (i.e., the Ebbinghaus or Titchner illusion). In this illusion, a central disc is surrounded by larger (or smaller) context circles, which creates a size-contrast illusion, meaning that the central disc is perceived as being smaller (or larger) than without context circles. Aglioti et al. (1995) found that this illusion only affected the perceptual judgements of the central disc, but not the maximum grip aperture (MGA) when grasping the central disc. They argued that this dissociation between perceptual and visuomotor tasks is best explained by assuming that the Ebbinghaus illusion is generated in the vision-for-perception stream, whereas the vision-for-action stream processes size independent of the context. They further suggested that, when performing an action such as grasping, our vision-for-action stream calculates a veridical and metrically accurate representation of the target object that is not accessible to our perceptual awareness. This notion has been dubbed a “motoric zombie” (Ramachandran & Blakeslee, 1999). In consequence, the perception-action dissociation as observed in the Ebbinghaus illusion was considered a strong argument in support of the TVSH (Carey, 2001).

However, since then other researchers have reported different results based on which they have argued that the effect of Ebbinghaus illusion displays on grasping may be comparable to the effects observed in perceptual tasks (Franz,

Gegenfurtner, Bülthoff, & Fahle, 2000; Pavani, Boscagli, Benvenuti, Rabuffetti, & Farnè, 1999). This seems contradictory at first sight, but a closer look at the data across different illusion studies suggests that the findings are relatively consistent. In summary, the two key findings are that (a) perceptual measures show large differences between illusion effects (see Fig. 1a), and (b) grasping shows a consistent illusion display effect across all studies (see Fig. 1b). We will first discuss (a) and then (b). Furthermore, we will argue that after careful analysis, the dissociation between perceptual measures and grasping disappears (Franz & Gegenfurtner, 2008).

1.2. Illusion effects on perception

The question of why perceptual measures yield such inconsistent effects was investigated in several studies by Franz and colleagues (for a review, see Franz & Gegenfurtner, 2008). In a nutshell, their main argument was that perceptual measures have varying response functions. Most importantly, manual size estimation (ME), which has been used in many studies, has been shown to differ from most other measures (see Franz, 2003). When performing ME, participants indicate the size of an object using their index finger and thumb. Proponents of the TVSH have interpreted this as a ‘manual “read-out” of what participants perceive’ (Haffenden & Goodale, 1998, p. 125), i.e., a form of cross-modal matching (Stevens, 1959). In consequence, ME has been widely used in studies on perception-action dissociations.

However, ME will typically exaggerate a physical change of object size. For example, in the study by Haffenden and Goodale (1998), a physical increase in object size of 1 mm led to an increase of app. 1.6 mm in ME. We can therefore expect that an illusory increase in object size of 1 mm would also result in a 1.6 mm (and not 1 mm) increase in ME. This is different from more classic perceptual measures such as a size adjustment task in which a physical increase in object size of 1 mm typically also leads to app. 1 mm increase in a size adjustment task (Franz, 2003). In consequence, we cannot interpret raw illusion effects found in a ME¹-task. We first have to correct ME for the steeper response function. Because ME depends linearly on object size, the correction can be done by simply dividing the measured illusion effect by the slope of the response function (this corresponds to a calibration in metrology, see also Bruno & Franz, 2009; Franz, Fahle, Bülthoff, & Gegenfurtner, 2001; Franz, Scharnowski, & Gegenfurtner, 2005; Glover & Dixon, 2002; Schenk et al., 2011 for details). Although correction may not be as necessary for other measures, as the slopes of their response functions are typically closer to one, we nevertheless performed such a correction for all measures (for a detailed discussion of when calibration is necessary and when it is optional, see Franz et al., 2001). Once the correction is performed, the

¹ It should be noted that ME does not always seem to exaggerate a physical change of size. If ME is performed closed-loop such that the hand is seen all the time the exaggeration seems to vanish. For an example, see de Grave et al. (2005). Because this has not been investigated systematically, we include two ME conditions in our experiment: One open-loop and one closed-loop.

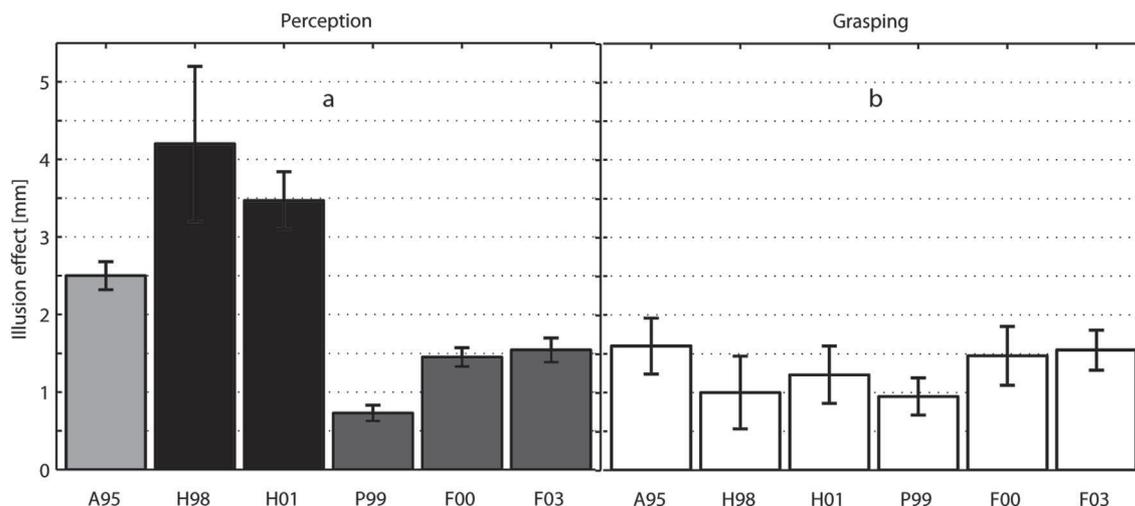


Fig. 1 – Results of previous studies on the Ebbinghaus/Titchener illusion with configurations identical to the original study by Aglioti et al. (1995): (a) Illusion effects on perceptual measures. These are colour-coded as black: Manual size estimation (ME) (open-loop). Dark grey: Perceptual comparison of a central target disc surrounded by illusion inducing circles to a neutral comparison element; light grey: Perceptual comparison of a central target disc surrounded by small illusion inducing circles to another central target disc surrounded by large illusion inducing circles, as used by Aglioti et al. (1995). This method was criticised by Franz et al. (2000) because it overestimates the relevant part of the illusion for grasping by app. 50%. It was therefore not used by subsequent studies and hence we will not discuss this measure in further detail here. (b) Illusion display effects on MGA in grasping. A95: Aglioti et al. (1995), H98: Haffenden and Goodale (1998), H01: Haffenden, Schiff, and Goodale (2001), P99: Pavani et al. (1999), F00 and F03: Franz et al. (2000; 2003). Error bars depict the SEM of the illusion effect.

perceptual effects become very consistent and can now be compared to the (equally consistent) illusion display effects on grasping (Franz & Gegenfurtner, 2008).

1.3. Illusion effects on grasping

For the Ebbinghaus illusion, reported effects on grasping range from not significantly different from 0 (e.g., Haffenden & Goodale, 1998) to significantly different from 0, but still smaller than the perceptual effect (e.g., Aglioti et al., 1995; Glover & Dixon, 2002) to significantly different from 0 and comparable to the perceptual effect (e.g., Franz et al., 2000; Pavani et al., 1999). However, unlike the perceptual effects discussed above (see Fig. 1a), the absolute size of the motor effect has not varied much between studies (Fig. 1b). This gives a very consistent picture of the effect of illusion displays on grasping. Since grasping shows a response function slope that is similar to the slopes found for classic perceptual measures, we can compare the raw illusion display effects between these measures.² Visual inspection shows that while statistical significance varies, these illusion display effects are actually quite

² More specifically: Grasping has been found to have a response function slope of app. .82 (Smeets & Brenner, 1999). If we perform the correction discussed above, the raw illusion display effects will be multiplied by roughly $1/.82 = 1.22$, to result in the corrected illusion effects. For classic perceptual measures the correction has hardly any effect (slope is close to 1), such that overall the match between perceptual illusion and grasping illusion is even better if we perform the correction. This better comparison was done in the present study but is omitted for the sake of brevity here.

similar in size between studies (Fig. 1b; see also Franz, 2003 and Franz & Gegenfurtner, 2008). In conclusion, it seems that the effects of the Ebbinghaus illusion displays on grasping might be very similar to the observed perceptual effects. However, the cause of the effect on grasping has been much debated.

1.4. Why do Ebbinghaus displays influence grasping?

If it was true that the Ebbinghaus illusion affects perception and action similarly, then this would directly contradict the notion of Aglioti et al. (1995) that grasping is immune to the Ebbinghaus illusion as predicted by the TVSH. However, this conclusion may be premature for two reasons.

First, Goodale (2008) suggested that some studies have measured grasping in ways that are so intrusive that the movement becomes awkward (Gonzalez, Ganel, & Goodale, 2006; Gonzalez, Ganel, Whitwell, Morrissey, & Goodale, 2008). According to the TVSH, awkward movements are controlled by the vision-for-perception system and therefore it would be no surprise that those studies found illusion display effects on grasping. Although this argument has been tested and refuted (Franz, Hesse, & Kollath, 2009), we took great care in our study to measure grasping in exactly the same way as done in the original study of Aglioti et al. (1995) such that this concern cannot apply.

Second, Haffenden and Goodale (2000) argued that in the Ebbinghaus display used by all studies in this field (starting with the first study by Aglioti et al., 1995 and as used in all the studies in Fig. 1), the context circles caused unexpected motor effects on grasping. Specifically, they argued that, even

though these motor effects look like illusion effects, they are in fact unrelated. Data supporting this notion was provided in a subsequent study (Haffenden et al., 2001). The main idea of Haffenden et al. (2001; Haffenden & Goodale, 2000) is that in some conditions the context circles of the Ebbinghaus display are treated as obstacles by the vision-for-action system. Such an obstacle avoidance effect may look like a perceptual effect, but would in fact be a motor effect. If obstacle avoidance can indeed explain the effects of Ebbinghaus displays on grasping, then the finding that grasping is affected by the illusion could be reconciled with the TVSH. There is, however, some contradictory data on this topic (Franz, Bühlhoff, & Fahle, 2003; Franz et al., 2001; Haffenden & Goodale, 1998; Pavani et al., 1999). In the following section, we will discuss the suggestion that obstacle avoidance may be the cause of illusion effects on grasping in more detail.

1.5. Can obstacle avoidance explain the effects of the Ebbinghaus display on grasping?

According to the obstacle avoidance hypothesis by Haffenden and Goodale (2000), the traditional distance between the target and the large context circles (approx. 9.5...14 mm in Aglioti et al., 1995, and Haffenden & Goodale, 1998) is just big enough for participants to fit their fingers between the annulus and the target, which reduces the in-flight aperture size. Conversely, the traditional distance between the targets and the small context circles (approx. 2...5 mm) is assumed to be not big enough to fit the fingers in between. As a consequence, participants tend to adjust their aperture size to fit around the whole stimulus, including annulus (see Fig. 2 for details). Thus, according to the obstacle avoidance hypothesis, the size of the MGA in grasping for Ebbinghaus illusion displays depends on annulus distance, rather than context circle size.

Haffenden et al. (2001) tested the obstacle avoidance hypothesis by comparing three grasp responses: targets surrounded by small context circles that were far away (Fig. 3: *small-far*), targets surrounded by the traditional small-context configuration (Fig. 3: *small-near*), and the traditional large-context configuration (Fig. 3: *large-far*). They found that the *small-far* responses were markedly different from *small-near* responses, but almost identical to *large-far* responses, which

is precisely what the obstacle avoidance hypothesis would predict.

Franz et al. (2003) repeated the study by Haffenden et al. (2001) and added another condition (*large-near*: large context circles, small distance; see Fig. 3). In this study, they found that participants grasped smaller in conditions with large context circles than in conditions with small context circles, regardless of context circle distance, contradicting the predictions of the obstacle avoidance account. Instead, the effects on grasping followed the same pattern as the perceptual effects. In conclusion, two studies (Franz et al., 2003; Haffenden et al., 2001) obtained opposite results using the same conditions. Importantly, this is the only case of obvious data inconsistency in the visual illusions and grasping literature on the Ebbinghaus illusion. Proponents of the obstacle avoidance account argue that the results of Haffenden et al. (2001) are more in line with results from other studies (Goodale, 2008; Westwood & Goodale, 2011), while sceptics argue that the study by Franz et al. (2003) had more statistical power due to a larger sample size, as well as clearer predictions due to an additional illusion condition (Franz & Gegenfurtner, 2008; Schenk et al., 2011).

Another study that tested the notion of whether obstacle avoidance may influence grasping in Ebbinghaus displays in a slightly different way was conducted by de Grave, Biegstraaten, Smeets, and Brenner (2005). They rotated the 2D context elements of Ebbinghaus figures to manipulate the extent to which the context elements might be perceived as blocking the path between the fingers and the object during a grasping movement and thereby manipulating the extent to which the context elements might act as obstacles. The authors found an effect of context element size on MGA, consistent with an illusion effect on grasping. They also found effects of context element rotation on several grasping parameters (grip orientation, final grip aperture), suggesting that context elements affect grasping movements in several other ways than just by altering perceived size. However, they did not find an effect of context circle rotation on MGA (Fig. 5c of de Grave et al., 2005). This is important, as MGA is the critical dependent variable which has been used in all studies to test for illusion effects on grasping. Moreover, the obstacle avoidance hypothesis has been specifically suggested by

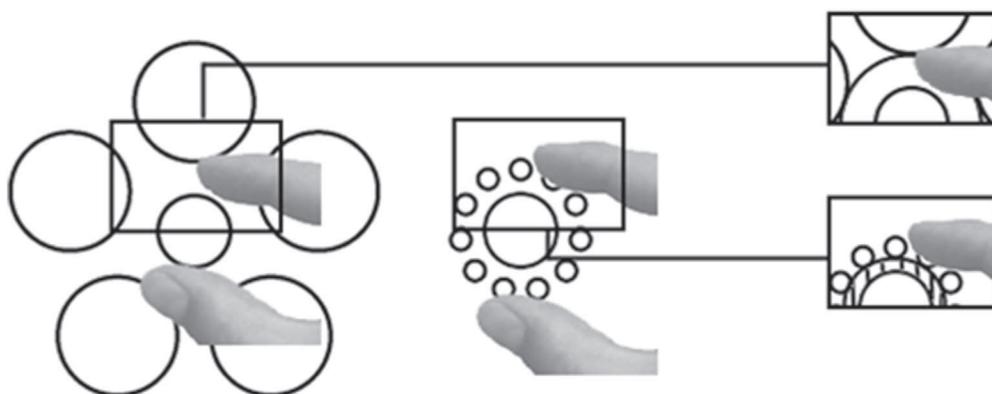


Fig. 2 – Obstacle avoidance as proposed by Haffenden and Goodale (2000) and Haffenden et al. (2001): The gap between the target and the large context circles (left) is just large enough to fit fingers in. Conversely, the gap between the target and the small context circles (right) is assumed to be too small to fit the fingers in and thereby causes a larger grip aperture.

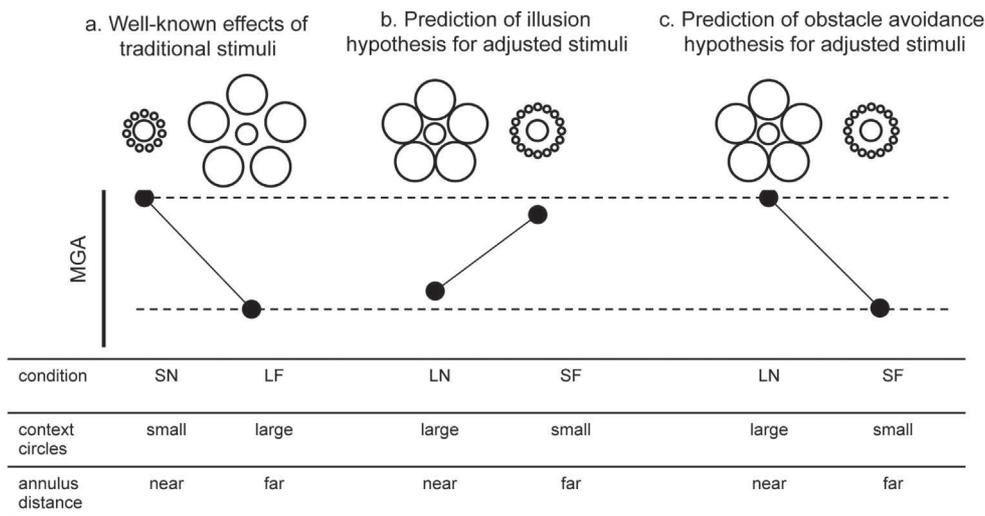


Fig. 3 – Stimuli and predictions of our study by context circle size and distance: small-near (SN), small-far (SF), large-near (LN), large-far (LF). (a) Traditional Ebbinghaus configurations create a known effect: larger grasping (MGA) in the SN condition than in the LF condition. This can be explained either by a single mechanism creating the illusion in grasping and in perception (illusion effect hypothesis: IEH) or by an obstacle avoidance mechanism operating independently of the perceptual illusion in grasping (obstacle avoidance hypothesis). We used a new set of conditions with oppositional predictions for the competing hypothesis: (b) According to the IEH, varying the distance and size of the context circles should have similar effects on grasping and on perception: large context circles lead to a smaller perceived size of the central circle. Larger context circle distance also leads to a slightly smaller perceived size (cf. the quantitative model of Roberts, Harris, & Yates, 2005,³). (c) According to the obstacle avoidance hypothesis, the gap between context circles and central circle is the critical parameter for grasping, not the size of the context circles. Consequently, a small gap should always lead to a large MGA, independent of the size of the context circles (compare SN and LN conditions), because the gap is too small to fit the fingers in. A larger gap, about finger-width, should lead to relatively small MGA due to finger-fitting, again independent of the size of the context circles (compare SF and LF conditions). Note the opposite pattern of predictions in (b) and (c).

Haffenden and Goodale (2000) to account for the illusion effect on MGA that was found in some studies. Given the fact that de Grave et al. (2005) did not find an effect of context-rotation on MGA, their results cannot be used to support the Haffenden and Goodale (2000) claim that obstacle avoidance processes account for illusion effects in grasping.

To summarise: according to de Grave et al. (2005) there are obstacle avoidance effects of context elements on certain grasping parameters, but not on MGA. Therefore, the question of whether obstacle avoidance can reconcile the TVSH with the effects of Ebbinghaus displays on grasping remains unresolved. At the core of this issue is an inconsistency in the empirical data (Franz et al., 2003; Haffenden et al., 2001). Resolving this issue is the main goal of our study.

³ For our figure, we fit the data from Franz et al. (2003) to the decay function proposed by Roberts et al. (2005): Illusion magnitude in mm = $-.07883 + .37616 \exp(-x/2.3076)$, where x = distance (in mm) from center of target to center of inducers. Plotted are the best linear fits. For large context circles: Illusion magnitude = $.02527 + 2.0983 \cdot f(x)$, small context circles: illusion magnitude = $.2004 + 2.1475 \cdot f(x)$. Note that there is a typing error in the published version of the original model, as it says $-.7883$ for parameter a , instead of $-.07883$. This being a typing error has been confirmed by Brian Roberts and Mike Harris (personal communication, January 10, 2014).

1.6. How to test for a dissociation: the issue of perceptually matched stimuli

Testing for a dissociation between perception and grasping requires comparing the effects of the illusion on different dependent variables. This is not trivial and is somewhat unusual. Illusion studies have employed three approaches to solve this problem (of which we employed the first two in our study).

The most common approach is to use different illusion displays and keep the physical size of the target object constant (“physically-matched condition”). The illusion effect is calculated by subtracting responses to the two different illusion configurations with the same target size (cf. Fig. 1). Although simple and straightforward, this approach has one major drawback: since the TVSH predicts grasping to be unaffected by the illusion, it predicts a null-effect (H0), which raises the problem of how to argue in favour of a statistical null-hypothesis (Westwood & Goodale, 2011; but see Schenk et al., 2011; Schenk & McIntosh, 2010). Some remedies can be used to tackle this issue, such as the use of Bayes factors or methods that test the predictions of the TVSH as the alternative hypothesis (H1). A prominent method to achieve the latter is described in the following paragraph and was employed together with Bayes factors and the physically-matched condition in our study.

To create a situation in which the TVSH predicts the H1 and not the H0, some studies (Aglioti et al., 1995; Haffenden &

Goodale, 1998) used a perceptual nulling method. Perceptual nulling is done by selecting two targets that look perceptually equal when embedded in different illusion configurations, although they differ in physical size (“perceptually-matched condition”). Because the TVSH assumes grasping to be veridical, it should follow the physical size of the targets such that the TVSH now predicts an effect (H1) between conditions with different context circles, while the IEH predicts a null-effect. Therefore, Westwood and Goodale (2011) argued that this nulling procedure provides a better test of the TVSH. Nevertheless, it also has its drawbacks: (a) because physical size and illusion are confounded it is difficult to quantify the illusion display effect if there is some effect on grasping that needs to be compared quantitatively to the perceptual effect, and (b) matching two targets in figure-surround configurations to be perceptually equal is in principle very difficult to do, especially when the surrounds have opposite effects (incremental vs decremental). For an example from lightness perception, see Jandó, Agostini, Galmonte, and Bruno (2003). In practice, this presents even more of a problem since the physical size of stimuli will always increase in steps, rather than continuously (e.g., Aglioti et al. 1995 used step-sizes of 1 mm, which may be too coarse for a good perceptual match). To partially account for this issue, we used smaller step sizes of .25 mm and also tested for the consistency of the perceptual matching by running the same perceptual tasks on the selected pair of matched stimuli, thereby providing additional information about the quality of the perceptual match.

Ganel, Tanzer, and Goodale (2008) took the nulling paradigm one step further in a study on the Ponzo illusion by creating opposing predictions for TVSH and IEH. Starting with perceptually-matched stimuli, reducing the size of the physically larger stimulus will make it appear perceptually smaller, such that TVSH and IEH predict opposite effects on grasping: the TVSH predicts the physically larger object to still be grasped with larger apertures (because TVSH assumes no effect of the illusion on grasping), while the IEH predicts smaller grip apertures for this object (because IEH assumes grasping to follow perception).⁴ However, a problem arises when neither hypothesis’ “strong” version is true, i.e., when there is a partial dissociation between perception and action: then, if the physical change in size is larger than the difference between the illusion effect in perception and in grasping, the results will seem to support the IEH; if it is

⁴ Ganel et al. (2008) found opposite effects on grasping and perception for the Ponzo illusion. They made use of the fact that placing objects in contrasting illusory contexts can create a situation where the physically smaller object is perceived as being bigger than the physically larger one. Nevertheless, the obtained MGAs were larger for the physically larger (but perceptually smaller) object and smaller for the physically smaller (but perceptually larger) one. Discussion whether this result for the Ponzo illusion constitutes evidence for the TVSH independent of data from the Ebbinghaus illusion, would go beyond the scope of this article. Such a discussion would need to resolve questions similar to those discussed for the Ebbinghaus illusion. This includes issues such as whether the task-demands were well matched, whether tasks should be performed in an open or closed-loop fashion, and whether the Ponzo illusion arises before or after the dorsal-ventral split (Murray, Boyaci, & Kersten, 2006).

smaller, the TVSH seems to be supported. While this allows for an upper (or lower) bound of the effect on grasping, this method suffers from the same problems mentioned above: the illusion effects are difficult to quantify due to the confounding of illusion size and physical size and the accuracy of the method is limited by the step size of the targets. In fact, the opposite effects procedure is equivalent to the nulling procedure used by Aglioti et al. (1995): whenever nulling works as proposed by the TVSH, it is possible to create an opposite effect situation, and whenever an opposite effect situation works as proposed by the TVSH, it is possible to create a nulling situation. In our study, we therefore decided to employ physically-matched conditions as well as perceptually-matched conditions to cover and compare the validity of the most widely used methods.

1.7. The present study

In the present study, we studied grasping movements using Ebbinghaus illusion displays to investigate whether or not actions are immune to visual illusions. We know of no study that has tried to account for all points of criticism and to identify the factors that may have led to the conflicting results regarding the obstacle avoidance account. This makes it difficult to interpret the studies in question (as shown in Fig. 1) which constitute key evidence in the debate about the TVSH. To solve this issue, we replicated the study by Haffenden et al. (2001) and investigated existing data inconsistencies to test the obstacle avoidance hypothesis. We also introduced some additional conditions to better generalise to grasping overall. Specifically, we aimed to assess to what extent Ebbinghaus illusion displays affect grasping and whether the possible effects can be attributed to a size contrast illusion or an obstacle avoidance strategy. Our main dependent variable was MGA. Additionally, we report the relative time to MGA, as it has been proposed that the presence of obstacles would result in a relatively earlier MGA (Smeets & Brenner, 1999; Smeets, Glover, & Brenner, 2003) and that MGA alone may not be sufficient to investigate the influence of visual illusions on grasping (Smeets & Brenner, 2006). An effect of visual illusions on MGA would then have to be explained in terms of not just size perception, but other grasping parameters as well.

This study includes a direct replication of the studies by Haffenden et al. (2001) and Franz et al. (2003), the only studies for which we identified contradictory results on effects of the Ebbinghaus illusion and obstacle avoidance on grasping. Hence, our stimuli were identical to those used by Haffenden et al. (2001) and Franz et al. (2003). We used four different conditions (see Fig. 3): the traditional Ebbinghaus conditions SN (*small context circles, annulus near target*) and LF (*large context circle, annulus far from the target*) to test the size of the illusion effect, as well as two non-traditional conditions (SF, “*small-far*” and LN, “*large-near*”), to test the proposed obstacle avoidance account. For the latter two conditions, the obstacle avoidance hypothesis and the IEH predict opposite patterns of results in grasping: the obstacle avoidance hypothesis predicts a distance effect (small distance → large MGA), while the IEH predicts a context circle size effect (small context circles → large MGA). Thus, the obstacle avoidance hypothesis predicts a larger MGA in the *large-near* and a

smaller MGA in the *small-far* condition (see Fig. 3c), while the IEH predicts the opposite pattern (Fig. 3b).

Two different procedures were used to vary the size of the central target: in one condition the physical size of the target was controlled (physically-matched), and in the other condition the perceptual size of the target was controlled (perceptually-matched). These conditions complement each other since the TVSH predicts differences in grasping in the perceptually-matched conditions but not in the physically-matched conditions, while the IEH predicts the opposite pattern.

We used three different perceptual measures: matching size perception to a graded series of stimuli (a classic perception task), ME without visual online feedback (open-loop), and ME with visual online feedback (closed-loop). We expected ME open-loop, but not the other perceptual measures to have a slope larger than one (cf. Footnote 1). Testing the variations in response functions of different perceptual tasks also provides novel information on the appropriateness of slope correction procedures as proposed by Franz (2003). Finally, by measuring the responses to “perceptually-matched” configurations in multiple perceptual measures, we also assessed the validity of the perceptual nulling procedure.

In addition to the overall size of the illusion effect, the correlation between perceptual measures and the MGA can provide information about the underlying visual representation. If grasping is guided by the same visual representation as perception, then one would predict a positive correlation between grasping and perceptual measures across participants. We tested this prediction, accounting for the fact that noisy measures predict a reduced correlation size (cf. Section 2.4).

We conducted the experiment in four different labs using exactly the same procedures and stimuli. By doing so, we obtained a precise estimate of the size of the illusion effect that combines advantages of a meta-analytical approach (large sample, multiple labs) with those of a single study (carefully controlled and comparable conditions).

We tested the following key hypotheses: (1) In the physically-matched conditions, participants grasp larger in the *large-near* condition than in the *small-far* condition (a test of the obstacle avoidance hypothesis; cf. Fig. 3). (2) In the perceptually-matched condition, participants grasp larger for the physically larger target (TVSH prediction: effect of physical size, no effect of illusory size). (3a) There is an effect of the Ebbinghaus illusion on grasping (IEH); (3b) This effect is equally strong in grasping and in perceptual measures; (3c) Across participants, the illusion effects in grasping and in perceptual measures are correlated.

2. Methods

2.1. Participants

Participants were recruited by labs from the following institutions: Università di Parma (Dipartimento di Neuroscienze – NB), University of Aberdeen (School of Psychology – CH), University of Hamburg (Department of General Psychology – KKK, VHF), University of Erlangen-Nuremberg (Department of Neurology – TS). Participants were right-handed (Edinburgh Handedness Inventory, Oldfield, 1971 – L.Q.>+47, decile R.1 or higher), had normal or corrected-to-normal eyesight, and had no history of neurological disorders. Participants' rights were protected according to the 1964 Declaration of Helsinki, and written consent was required from all participants. Ethical approval was obtained from local ethics committees.

To determine the appropriate sample size, we conducted a power analysis for an illusion display effect between two conditions as to be tested in the obstacle avoidance hypothesis (*large-near* > *small-far*; obstacle avoidance) and the IEH (*large-near* < *small-far*; illusion effect). We aggregated illusion display effects and standard deviations from previous studies (using data from a total of 6 studies and 146 participants) weighted by the number of participants to estimate Cohen's d (Cohen, 1988) by the formula $d = IE/SD$: $d = 1.38 \text{ mm}/1.90 \text{ mm} = .73$ (Table 1).

Since a larger distance between target and context circles might cause the target to appear smaller (Girgus, Coren, & Agdern, 1972; Roberts et al., 2005), we might expect the illusion effect in the non-traditional conditions (i.e., the difference SF–LN) to be smaller than the effect in the traditional conditions (i.e., the difference SN–LF). Considering this and some possible inter-lab variability, expecting the same effect as found in previous studies may be an overestimation. Hence, we think it is reasonable to base our calculations on an effect 70% as large as the original one of $d = .73$, as has been done in a previous power analysis for the same effect (Franz et al., 2003). Doing so would give us an effect of $d = .51$. This is close to the smallest illusion display effect observed in previous studies ($d = .50$ – Haffenden & Goodale, 1998).

We decided to aim for at least $1-\beta = 80\%$ power for each lab to ensure that data can be interpreted separately, as well as to account for possible systematic variations between labs. With an alpha-level of $\alpha = .05$, this resulted in a desired sample size of $N = 33$ for each lab. The total of $N = 132$ for all labs combined would enable us to detect an effect of $d = .28$ with $\alpha = .05$ and $\beta = .10$. To make counterbalancing easier, we tested $N = 36$ participants per lab, for a total of $N = 144$ participants. This ensured that if an illusion effect on grasping exists, we should

Table 1 – Illusion display effects on grasping found in earlier studies, as summarised by Franz and Gegenfurtner (2008).

Study	Aglioti et al. (1995)	Haffenden and Goodale (1998)	Pavani et al. (1999)	Franz et al. (2000)	Haffenden et al. (2001)	Franz et al. (2003)
Effect	1.6 mm	1.0 mm	.95 mm	1.47 mm	1.4 mm	1.55 mm
SEM	.36 mm	.47 mm	.24 mm	.38 mm	.64 mm	.26 mm
SD	1.35 mm	1.99 mm	1.00 mm	1.94 mm	2.71 mm	1.87 mm
N	14	18	18	26	27	52
d	1.19	.50	.95	.76	.52	.83

be able to detect it. The power analysis was conducted using the function *t*-test for difference from a constant of the program G*Power 3.1.7 (Faul, Erdfelder, Buchner, & Lang, 2009).

2.2. Stimuli

We used four different versions of the Ebbinghaus illusion. These differed in the distance between target and context circles (“near” and “far”), and size of context circles (“small” and “large”). The four resulting versions can be seen in Fig. 3. In the “near” conditions, the inner diameter of the annulus (i.e., the distance from middle point of the target circle to closest point of the context circles) was 38 mm. In the “far” conditions, the inner diameter of the annulus was 54 mm. These distances are identical to those used by Haffenden et al. (2001). In the “small” conditions, context circles were 10 mm in diameter. In the “large” conditions, context circles were 54 mm in diameter. Target discs were white plastic discs of 3 mm height and 28, 30 and 32 mm diameter. These sizes have also been used by Haffenden et al. (2001), although it should be noted that those experiments also used target discs of 31 mm diameter, which we omitted for symmetry and parsimony. This resulted in distances between the central targets and the annuli of the context circles of 3, 4 and 5 mm in the “near” conditions, 11, 12 and 13 mm in the “far” conditions. Details about measurements and distances are summarised in Table 2. Note that these values apply only for the physically-matched condition, as target sizes for the perceptually-matched condition were determined for each participant separately (see 2.4: Procedure).

2.3. Apparatus

Participants sat comfortably on a chair in front of a table. Their head was at a height of 50 cm above the table to keep the viewing distance constant and the viewing angle at about 80–90°. They were wearing PLATO liquid crystal shutter glasses (Translucent Technologies, Toronto, Ontario, Canada – Milgram, 1987) to control target visibility.

The stimulus set-up consisted of a piece of paper (A4 sized) with the context circles printed on it, laid flat on the table, such that participants viewed the targets from almost directly above (80–90°). In the physically-matched conditions, white PVC discs of 3 mm height and 28, 30 and 32 mm diameter, with a 1 mm black line drawn around the circumference, were used as target stimuli and were positioned in the middle of the context circles. These are the exact specifications of stimuli used by Haffenden and colleagues (Haffenden & Goodale,

1998; Haffenden et al., 2001–31 mm stimuli omitted), as well as Pavani and colleagues (Pavani et al., 1999–31 mm stimuli omitted, 28 mm added). In the size matching task, white circles ranging from 23 mm to 37 mm in diameter (.5 mm steps, 29 circles total) with a 1 mm line around the circumference, printed on a sheet of paper in ascending order, were used as a graded series of comparison stimuli in the size matching task. Pilot testing showed most responses to fall within this range, see Appendix B. For the perceptually-matched condition, two of 15 different discs of sizes ranging from 28 mm to 32 mm in steps of .25 mm were used.

The starting position for the participants' response hand was on the table, 20 cm from the target. For a schematic depiction of the experimental set-up, see Fig. 4. Three markers were attached to participants' right wrist, thumb, and index finger (Fig. 4a). The trajectories of the digits were recorded using appropriate motion tracking systems (see Table 3).

2.4. Procedure

There was a grasping task and three perceptual tasks: size matching, open-loop manual estimation, and closed-loop manual estimation. These tasks were presented in separate blocks. For the perceptual tasks, there were 54 trials each: 36 in the physically-matched condition (4 illusion conditions*3 target sizes*3 repetitions presented in random order) and 18 in the perceptually-matched condition. In the first perceptual task (size matching), two perceptually-matched configurations were created (SFx and LFy). They were tested the same number of times as configurations SF and LF in the physically-matched condition, which gave us an equally precise estimate of the illusion display effect. In grasping, the participants completed 90 trials (60 physically-matched: 4*3*5 + 30 perceptually-matched: 2*3*5), resulting in a total of 252 trials per participant.

The size matching task was always the first task. First, we determined for both a SF and a LF configuration which target sizes were required to create a perceptual match with a reference circle of 30.5 mm diameter (presented on a A4 sheet of paper). For both SF and LF, a 1-up, 1-down staircase procedure was conducted where participants had to indicate whether the target disc appeared to be smaller or bigger than the reference circle, using steps of .25 mm. The discs of corresponding target size were then used to create what we call the perceptually-matched SFx and LFy configuration. It is to be expected that the SFx and LFy vary between participants and that the difference between those two configurations reflects the extent to which the context influences the perceived size.

Table 2 – Sizes of and distances between the stimuli.

	Diameter of target (mm)	Number of context circles	Diameter of context circles (mm)	Inner diameter of annulus (mm)	Min. distance target – annulus (mm)
Small, near	28, 30, 32	11	10	38	5, 4, 3
Small, far	28, 30, 32	16	10	54	13, 12, 11
Large, near	28, 30, 32	5	54	38	5, 4, 3
Large, far	28, 30, 32	5	54	54	13, 12, 11

Note: Inner diameter is the diameter through the points of the context circles closest to the target. In the perceptually-matched condition, diameter of the target and minimal distance target – annulus may be different.

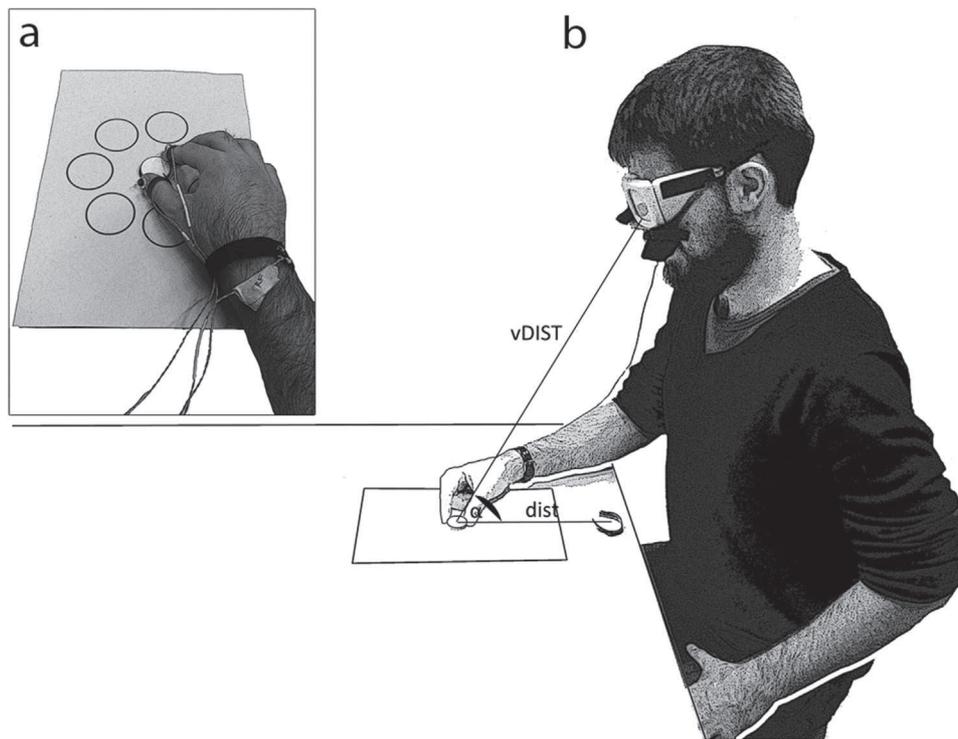


Fig. 4 – (a) Hand with three markers attached to thumb, index finger, and wrist. (b) Experimental set-up with a viewing distance ($vDIST$) of app. 50 cm, viewing angle (α) of app. 80–90°, distance ($dist$) of 20 cm between starting point and stimulus. The participant is sitting comfortably, so as not to fatigue over a large number of trials, and wearing LCD goggles with cloth blinders attached to the bottom to prevent any view below the goggles.

This means we can use this difference as a measure of the individual illusion effect. For the second component of the size-matching task, participants were presented with a target stimulus surrounded by one of the four illusion contexts (LF, LN, SF, SN), as well as an A4 sheet of paper containing a graded series of comparison circles. This was located 20 cm to the left of the target. Participants were asked to indicate verbally which comparison stimulus they perceive as equal in size to the target disc. This was done for all physically-matched (disc sizes 28, 30, 32 mm) and perceptually-matched configurations.

The *open-loop manual estimation task* started with participants resting their right hand at a starting position on the table. When they saw the stimulus, they were asked to lift their right hand and indicate the size of the target stimulus with their right thumb and index finger. They were asked to press a response button with the index finger of the left hand when they felt satisfied with their estimation. The shutter glasses closed when the right thumb or index finger had

moved 20 mm from their starting position to suppress vision. Trials that took longer than 2500 msec from opening of the shutter glasses to pressing the response button, or ended with the button pressed while the participant's thumb and index finger were still moving at more than 30 mm/sec relative to each other, were counted as errors and repeated at a random position within the same block. After this, participants were asked to grasp the target disc and lay it on the table next to the stimulus set-up. This was done to provide the same haptic feedback as in the grasping trials, as was proposed by [Haffenden and Goodale \(1998\)](#). The shutter glasses opened after the experimenter had prepared the next trial. In the *closed-loop manual estimation task*, participants were asked to indicate the size of the target stimulus in the same way as in open-loop manual estimation, except that participants had full view of their hand and of the target throughout the trial.

In the *grasping task*, participants were asked to grasp the target disc with their right hand and lay it on the table next to

Table 3 – Motion tracking systems used by each lab, including basic specifications.

Lab	System	Type	Sampling rate (Hz)	Spatial resolution (mm)
Parma (NB)	SMART System (BTS Bioengineering, Milan, Italy),	Optical (infrared)	120	.3
	Qualisys ProReflex MCU1000 (Qualisys AB, Gothenburg, Sweden)	Optical (retro-reflective)	240	.4
Aberdeen (CH)	Optotrak 3020 (Northern Digital, Waterloo, Canada)	Optical (infrared)	200	.01
Erlangen (TS)	Zebris CMS-70 (Zebris medical GmbH, Isny, Germany)	Acoustic	50	.1
Hamburg (KKK, VHF)	Optotrak Certus (Northern Digital, Waterloo, Canada)	Optical (infrared)	200	.01

the stimulus set-up. The grasping task was performed under open-loop conditions, as was proposed by Post and Welch (1996) and as has been done in most previous experiments (de Grave et al., 2005; Franz et al., 2003, 2000; Glover & Dixon, 2002; Haffenden & Goodale, 1998; Haffenden et al., 2001; Pavani et al., 1999; for a comparison of open-loop to closed-loop grasping in the Ebbinghaus illusion see: Franz et al., 2005). The shutter glasses closed when the right thumb or index finger had moved 20 mm away from the starting position. Trials ended when the participant's thumb or index finger touched the target object.

These four blocks were conducted for each participant. Before each block, participants were asked to perform 5 pseudo-random practice trials. The order of blocks was counterbalanced between participants, with size matching always being the first task, so that each of the 6 (3!) possible task orders was used six times per lab.

2.5. Data analysis

As dependent variables, we used the *diameter* of the selected circle in the graded series for the size matching task, the indicated *distance* between thumb and index finger markers for the manual estimation tasks, and the MGA for the grasping task. For each measure, we eliminated outliers that were more than 2 SD above or below the participant's mean for each condition.

For each dependent variable, a repeated measures analysis of variance (ANOVA) with the factors “target disc size” (three levels – 28 mm, 30 mm, 32 mm) and “context circle type” (four levels – “small-near”, “small-far”, “large-near”, “large-far”) was computed. We used t-tests to compare conditions separately, correcting for multiple comparisons by applying a Holm-Bonferroni correction (Holm, 1979). Wherever t-tests were used, we also calculated Bayes factors (e.g., Dienes, 2011) and denote evidence according to the thresholds proposed by Jeffreys (1961). We used the following prior distributions: in the *physically-matched* condition, the prior distribution for the effect of the illusion on grasping for H1 (prediction of the IEH) was a normal distribution with expected value and SD given by the mean and the SEM of the grasp effect predicted from the measured illusion effect in *size-matching*, taking into account the necessary slope corrections, thereby corresponding to equal effects of the illusion on perception and grasping ($\text{predictedGraspEffect} = \text{graspSlope} * \text{perceptualIllusion} / \text{perceptualSlope}$). The H0-prior (prediction of the TVSH) was a point-hypothesis at 0, corresponding to no illusion effect in grasping ($\text{predictedGraspEffect} = 0$). For the *perceptually-matched* condition, the H1-prior (prediction of the TVSH) for the effect on grasping was a normal distribution, with the expected value and the SD specified by the mean and the SEM of the predicted effect in grasping based on the physical difference alone, without any illusion effect ($\text{predictedGraspEffect} = \text{graspSlope} * \text{physicalDifference}$). As H0-prior (prediction of the IEH), we used a normal distribution with the expected value and the SD given by the mean and the SEM of the residual perceived differences in size between the two perceptually-matched stimuli as measured in the *size-matching* task ($\text{predictedGraspEffect} = \text{graspSlope} * \text{residualPerceptualDifference} / \text{perceptualSlope}$). Note that if our perceptual matching

procedure worked, mean and SEM should be close to 0. For the Fisher-z-transform of the *correlation between grasping and perceptual measures*, we used as H1-prior (prediction of IEH) a normal-distribution with the expected value and the SD corresponding to the z-transformed maximal expected correlation and its SEM as given by standard BC_a bootstrap (Efron & Tibshirani, 1993). As H0-prior (prediction of the TVSH), we used a point-hypothesis at 0, corresponding to no correlation between perceptual effects and grasp effects of the illusion. This allowed us to gather evidence in favour of the null-hypothesis in instances in which one theory predicts an effect and the other one does not. The use of Bayes factors, along with high statistical power and a setup in which both competing theories are tested as H0 as well as H1, makes it easy to argue for the null-hypothesis, should we obtain non-significant results.

To compare the illusion display effects between dependent variables, we needed to calculate a corrected illusion effect for each variable (Franz, 2003) to adjust the illusion effect for different slopes between size and the outcome measure. To make illusion effects comparable to other studies, we used the formula employed among others by Bruno and Franz (2009) for illusion effects as a percentage of the actual size:

$$i_{\text{corr}} = i_{\text{raw}} / s * 100 / t,$$

with i_{corr} = corrected illusion effect, i_{raw} = mean raw illusion effect, i.e., mean difference between responses of two conditions, s = slope, t = target size. Standard errors were calculated using a Taylor-approximation (Franz et al., 2009):

$$SE_{i_{\text{corr}}} = i_{\text{raw}} / s * \sqrt{\sigma_s^2 / s^2 + \sigma_i^2 / i_{\text{raw}}^2 - 2\sigma_{is} / (i_{\text{raw}} * s)} * 100 / t,$$

where $SE_{i_{\text{corr}}}$ stands for standard error of corrected illusion effect, with i_{raw} = mean illusion effect, s = mean slope, σ_s = slope S.E.M., σ_i = illusion S.E.M., and σ_{is} = illusion effect-slope covariance. For details on this formula, see Franz et al. (2005) and Franz (2007). This procedure requires the slopes to be significantly different from 0, which we can reasonably expect them to be.

Illusion effects were calculated as the difference between two conditions. The three effects that are of the most interest to us were the traditional illusion effect (*small-near vs large-far*), the distance-matched illusion effects (*small-far vs large-far* and *small-near vs large-near*), and the critical test condition for the obstacle avoidance effect (*large-near vs small-far*).

To test the across-subject correlations between the illusion effect on MGA and on perceptual measures, correlations were computed between each perceptual measure and grasping. These correlations were then compared to the upper bound of the correlation predicted by the IEH and to 0 (as predicted by TVSH) by submitting the Fisher-z-transformed correlations to t-tests and calculating Bayes factors in the same fashion as described above.

For the expected correlation between effects of the illusion on perception and grasping, we can employ a formula from classical test theory. We are interested in correlating two latent variables (the “true” illusions in grasping and perception). This is analogous to the question in classical test theory of how well a “true” test value and a “true” value of an external criterion will correlate (external validity). In classical test

theory, an upper bound for the measured correlation of a test-score with an external criterion is given by:

$$r_{T,Tc} = r_{tc} / \sqrt{r_{tt} * r_{cc}},$$

with $r_{T,Tc}$ = the “true” correlation between latent variables. In classical test theory, these are the “true” test value and the “true” value of an external criterion. In our case, these are the “true” illusions in grasping and perception. r_{tc} = the measured validity of the test (classical test theory: measured correlation between test score and external criterion; here: measured correlation between grasp illusion and perceptual illusion), r_{tt} = the reliability of the test (here: reliability of grasp illusion), r_{cc} = the reliability of the criterion (here: reliability of perceptual illusion). If the grasp illusion were perfectly based on the perceptual illusion $r_{T,Tc}$ would be 1. Solving the equation for r_{tc} , gives:

$$r_{tc} = r_{T,Tc} * \sqrt{r_{tt} * r_{cc}} = \sqrt{r_{tt} * r_{cc}}.$$

This is the maximal correlation we can expect between the measured illusions in grasping and perceptual tasks, given their reliabilities. Because this prediction is based on strong assumptions, we call it the maximal expected correlation. This is the correlation that a strong version of the IEH would predict. The strong version of the TVSH would predict a correlation of 0. We also conducted a power analysis based on data from a previous experiment (Franz et al., 2003) that gave us split-half reliabilities of .22 and .47 for grasping and size adjustment, respectively, resulting in a maximal expected correlation of .32. With $\alpha = .05$ and $N = 144$, we would have 98% power to detect this effect.

Our hypotheses, in statistically testable terms, were as follows: in the physically-matched condition, we tested for a significant main effect of the factor *illusion condition* on MGA and all three perceptual measures. (1) Furthermore, we examined whether the conditions *large-near* produce a larger MGA than the conditions *small-far* (obstacle avoidance). (2) In the perceptually-matched condition, we examined whether we would find a difference between SFx and LFy in MGA, and ME. (3a) We also tested whether the corrected illusion effects in grasping differed significantly from 0 and (3b) from the corrected illusion effects in any of the three perceptual measures, as well as (3c) whether there is a correlation between illusion effects in grasping and perceptual measures.

For all of these comparisons, paired-sample t-tests were employed, as well as a Bayesian equivalent, i.e., Bayes factors for the null-hypothesis versus an alternative hypothesis. We report Bayes factors and *p*-values as exact values when above .001, and use a significance level of $\alpha = .05$ for all analyses. 95% confidence intervals are reported where applicable; means are reported including the appropriate standard error as $M \pm SEM$.

3. Implementation of preregistered protocol

The introduction and methods section of the present study were reviewed and accepted in-principle as a registered report in September 2014 and were not modified after that (allowing

for minor language adjustments). We collected all data after in-principle acceptance and finished data collection in May 2015.

3.1. Data collection: deviations from preregistered protocol

During testing, the SMART system in Parma had technical difficulties and had to be replaced with a Qualisys system (Table 3). Because of this, the editor agreed to extend the time frame for submission from 10 to 12 months. There were also a few minor inconsistencies in the experimental procedures between labs: (a) In the ME-tasks in Erlangen, participants did not record their ME by pressing a button (as in the other labs), but by keeping their fingers still and indicating verbally to the experimenter that they were showing the perceived size. This meant that the pre-registered time limit of 2500 msec was sometimes exceeded. Both procedures are common practice. (b) Three participants had a slightly smaller handedness score than pre-specified (but were still classified as right-handed, $LQ > 24$ instead of $LQ > 47$). Otherwise, we fully adhered to the registered protocol.

3.2. Post-hoc design critique: would a dual-illusion display be a better test of TVSH predictions?

After we submitted our data in phase 2 of this registered report, a reviewer worried that presenting observers with only one Ebbinghaus figure at a time may not be a fair test of the TVSH. The original studies of Aglioti et al. (1995) and Haffenden and Goodale (1998) used a dual-illusion display, showing two Ebbinghaus figures side-by-side, as is often used in textbooks to demonstrate the illusion. In contrast, our design used only one Ebbinghaus figure at a time, thereby employing a single-illusion display that has typically been used in perceptual research (e.g., Coren & Enns, 1993; Coren & Girgus, 1972; Girgus et al., 1972). We chose a single illusion design at phase 1 because it represents, in our opinion, the optimal choice for testing our hypotheses. All studies (independent of whether they use single- or dual-illusion displays) have to ensure that the task demands are as similar as possible in all conditions. However, when a dual-illusion display is used, the magnitude of the illusion depends on whether the targets in the two Ebbinghaus figures are compared to each other (direct-comparison condition) or whether they are successively and separately compared to a neutral disc (separate comparison). Specifically, in a direct comparison the effect is app. 50% larger than the sum of the effects in two separate comparisons (Foster & Franz, 2014; Franz et al., 2000). This raises an obvious problem: in the perceptual task, participants can compare two discs, whereas in the grasping task, they typically grasp only one target. In other words, when using a dual-illusion display there is a fundamental mismatch of task demands between the perception and action conditions, leading to an underestimation of the action effect relative to the perceptual effect. This has been known for some time (Franz et al., 2000; Pavani et al., 1999). In consequence, it is common practice to use single-illusion displays in research on the TVSH, also by proponents of the TVSH (e.g., Haffenden et al., 2001; for related work see Dewar & Carey, 2006; Foster & Franz, 2014; Foster, Kleinholdermann, Leifheit, & Franz, 2012; for further discussion of the issue of task demand mismatches in perception and action, see Bruno, 2016; Schenk et al., 2011).

4. Results

4.1. Frequentist analyses, Bayesian analyses, and open data

Results of traditional frequentist tests are reported as usual and accompanied (where appropriate) by corresponding Bayes-factors. For the logic of Bayes factors, see Section 2.5 and Dienes (2011). In essence, the Bayes factor indicates the relative likelihoods of two competing hypotheses, which we stated for all our tests in Section 2.5, thus giving us a continuous measure of how strongly either hypothesis is favoured. The evidence for the H1 always equals $1/(\text{evidence for H0})$. Bayes factors may be interpreted following the guidelines proposed by Jeffreys (1961), such that we can speak of strong evidence for H0 for Bayes factors smaller than 1/10, substantial evidence for H0 for Bayes factors between 1/10 and 1/3, inconclusive results for Bayes factors between 1/3 and 3, substantial evidence for H1 for Bayes factors between 3 and 10, and strong evidence for H1 for Bayes factors above 10.

We first report the results of the pre-registered analyses. Then, in Section 4.7, we report results of post-hoc analyses that were not pre-registered. In some cases it seemed easier for the reader that we also include post-hoc analyses before Section 4.7. These are clearly marked as not pre-registered analyses.

All data, analyses and materials for this study can be downloaded via Mendeley Data at <http://dx.doi.org/10.17632/4676n2pdrf.3>. After results were submitted in the stage 2 registered report, the study was reviewed by the same anonymous reviewers as in stage 1. Below, we report minor changes and issues related the design of the study that surfaced after in-principle acceptance.

4.2. Participants

We invited $N = 160$ participants to the laboratory, 16 of which were not included in the data analysis: 9 due to technical errors (recording did not produce analysable data or could not be finished), 3 due to experimenter errors (the experimenter followed the wrong protocol), 4 because they were not unambiguously right-handed (negative LQ or left-handed by self-report; two of these were not tested further but had been given an ID, two were tested because their handedness inventories were evaluated after testing). Thus, we obtained and included the data from $N = 144$ right-handed participants, $N = 36$ in each lab. The order of blocks was counterbalanced between these participants.

4.3. Overall data

Mean responses for all tasks and conditions are depicted in Fig. 5. We will discuss the physically-matched condition and the perceptually-matched condition successively. In both conditions, participants completed the tasks: grasping, classic perception (size matching), closed-loop ME, and open-loop ME as outlined in Section 2.4. For brevity, we will sometimes talk about perceptual tasks in general, which comprises classic perception as well as ME (because the TVSH assumes this to be a perceptual task).

4.4. Physically-matched conditions

The physically-matched conditions consisted of three objects (discs of 28, 30, 32 mm diameter), presented within four context circle types (LF, LN, SF, SN; Table 2). We submitted the results of each task to a 3 (target size) $\times 4$ (context circle type) ANOVA. Results show that both factors affected all tasks (Table 4). In some tasks, there was also a significant interaction between target size and context circle type. Since such small modulations of the context circle type effect are not unusual (Franz et al., 2000) and do not change the overall pattern of results (Fig. 5), these interactions will not be discussed further. Importantly, we found a main effect of context circle type in grasping, meaning that the MGA in grasping was affected by the illusion configuration. This is to be expected if we assume that grasping follows the perceived size (IEH), but needs to be explained by some other mechanism like obstacle avoidance (Haffenden & Goodale, 2000; Haffenden et al., 2001) if we assume that grasping is immune to illusions (TVSH).

To investigate these effects in more detail, we calculated contrasts between specific illusion configurations in each task. Most relevant are the contrasts SN–LF (the traditional Ebbinghaus illusion contrast), as well as SF–LF and SN–LN (the distance-adjusted conditions which should ameliorate obstacle avoidance effects). If the effect of the illusion configuration on MGA is indeed caused by obstacle avoidance, then the TVSH predicts no difference in the adjusted contrasts, while the IEH predicts a difference. Results from our study are depicted in Fig. 6a.

The strongest test of obstacle avoidance is comparing the configurations LN and SF. Here, the IEH and the obstacle avoidance hypothesis make opposite predictions (Fig. 3). We found a larger MGA for the SF condition than for the LN condition [$t(143) = 2.68, p = .008$; Fig. 7], which is consistent with the IEH but not with obstacle avoidance.

Next, we compared the size of the illusion effects between measures. For this, we calculated slope-corrected illusion effects (Fig. 6c) and compared grasping to the perceptual measures (Table 5).

Results show that all but one t -test indicate similar corrected illusion effects for grasping and the different perceptual measures (all $p > .15$). Only one t -test is significant (grasping vs closed-loop ME, SF–LF: $p = .044$; Table 5, row 3). However, this difference is not significant after applying the Bonferroni-Holm correction.⁵ Such an alpha-correction is needed if we wish to interpret the fact that only one out of nine t -tests is significant as evidence that there is an effect. Instead, it seems that the closed-loop ME simply showed an unusually large illusion effect, as is also suggested by the fact that the same contrast is also significantly different when classic perception is compared to closed-loop ME [$t(143) = 2.62, p = .010$, see also Table A.1 in the Appendix; this

⁵ In general, the Bonferroni-Holm correction is less conservative than the classic Bonferroni correction (Holm, 1979). In our case, however, both lead to the same result: in Bonferroni-Holm, the divisor of the alpha level is initially the same as in the Bonferroni correction. This divisor then decreases by one each time a significant result is found at the current alpha level. Thus, the corrections are equivalent in our case of only one significant result.

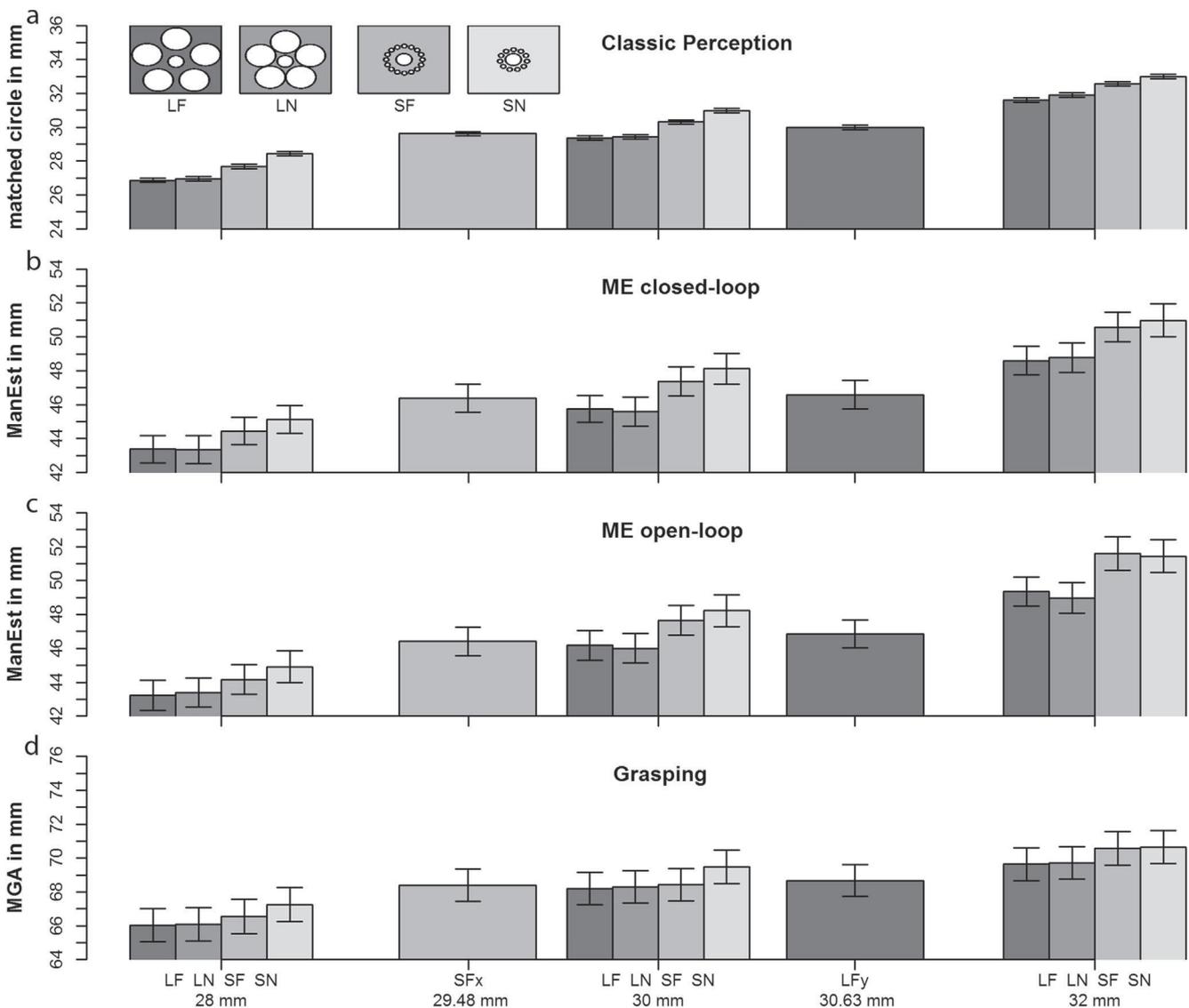


Fig. 5 – Mean responses for each object size and context circle type. **a:** Classic perception task, **b:** closed-loop ME, **c:** open-loop ME, **d:** grasping. The relative width of the bars corresponds to the number of trials in each configuration. Error bars indicate between-subjects SEM. These SEM contain between-subjects variance and can therefore not be used to interpret differences between conditions (because conditions were varied within-subjects). See the following figures for error bars that allow such interpretations (cf. Franz & Loftus, 2012; Loftus & Masson, 1994).

analysis was not pre-registered]. This interpretation is also consistent with the Bayesian analysis because all Bayes factors strongly support the H1 (illusion effects in grasping are comparable to illusion effects in other measures; Table 5).

Finally, we tested whether there is a correlation between illusion effects in grasping and perception. According to the IEH, participants with a relatively large perceptual illusion should also have a large illusion effect in grasping. The TVSH on the other hand predicts no correlation. The main problem when testing for such a correlation is that grasping and ME are relatively noisy measures, such that *a-priori* the correlation must be small, even if grasping and perception were based on perfectly identical size representations and noise is only generated when creating the actual response. Small correlations require very large sample sizes to be detected reliably (e.g., a correlation of $r = .20$ would require $N = 314$ participants

to achieve 95% power). To estimate a lower limit for a meaningful sample size, we used a formula from classic test theory to calculate the maximal theoretically possible correlation given the reliabilities of the measures (Section 2.4). Our sample is large enough to at least detect the maximum possible correlation with sufficient power, while the usual smaller sample sizes would not be able to detect even this upper limit of the correlation with sufficient power.

Table 7 shows the reliabilities and correlations between all illusion contrasts. As expected, the reliabilities are relatively small for grasping and ME, because these measures are affected by noise generated during hand and finger movement. Classic perception is not affected by such movement noise and therefore has considerably larger reliabilities. Given the small reliabilities of grasping and ME, the correlations are also small. Out of 9 correlations, 5 were significantly

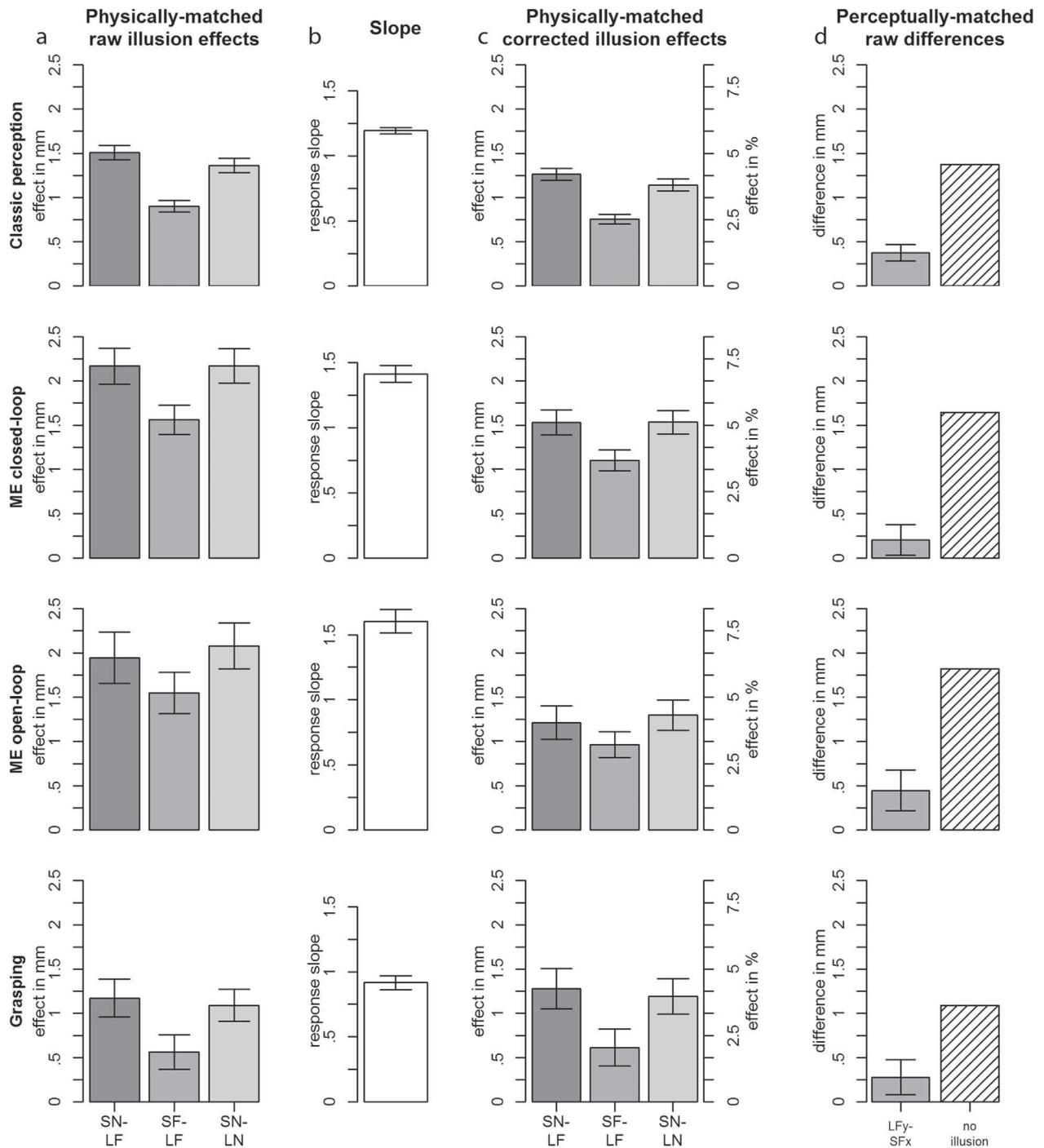


Fig. 6 – Mean illusion effects in the physically-matched conditions and differences between the perceptually-matched conditions. **a:** Raw illusion effects for the traditional contrast (SN–LF) and the adjusted contrasts (SF–LF and LN–SN). **b:** Slopes of the response functions. **c:** Corrected illusion effects (calculated by dividing each raw illusion effect by the corresponding slope). Results show similar corrected illusion effects in all tasks. This is consistent with the IEH but not with the TVSH. **d:** Differences between the perceptually-matched conditions (LFy–SFx). If a response followed perceived size (as determined by our nulling procedure and as predicted by the IEH), the difference should be zero. If a response followed physical size (as predicted by the TVSH), the difference should be equal to the hatched bars (this prediction is calculated by multiplying the physical difference between LFy and SFx by the slope of each response). Results show similar small effects in all tasks (indicating that nulling did not work perfectly). These small effects clearly differed from the no-illusion predictions, indicating that all tasks (including grasping) follow perceived size and not physical size. Error bars depict the within-subjects SEM. Because these SEM are for within-subject differences, they do not contain between-subjects variance and are therefore consistent with the results of a t-test against zero (cf. Franz & Loftus, 2012).

Table 4 – ANOVA results for all tasks in the physically-matched condition.

Task	Main effect				Interaction	
	Context circle type		Object size		Context circle type x object size	
	<i>F</i> (3, 429)	<i>p</i>	<i>F</i> (2, 286)	<i>p</i>	<i>F</i> (6, 858)	<i>p</i>
Grasping	17.10	<.001**	217.71	<.001**	1.36	.227 n.s.
Perception	218.52	<.001**	2304.89	<.001**	4.87	<.001**
ME CL	76.99	<.001**	401.86	<.001**	1.81	.094 n.s.
ME OL	36.84	<.001**	259.47	<.001**	2.53	.019*

Note: n.s. indicates non-significant, * indicates $p < .05$, ** indicates $p < .001$. Statistically significant p -values are given in italics.

different from zero, indicating a relationship between grasping and the perceptual measures. This is a pattern we would expect given a small effect size and, accordingly, relatively low statistical power: if we assume the factual correlation to be $r = .2$, with $N = 144$ we achieved a power of 68% for each test of a correlation against 0. This would be a small effect size according to Cohen (1988) and similar to most correlations we found.

In our pre-registered Bayesian analysis, we contrasted the hypothesis that there is no correlation (H_0) with the hypothesis that the correlation is equal to the theoretical upper bound (i.e., the maximal expected correlation; H_1). This gives a somewhat mixed result, with 5 Bayes factors supporting the H_0 , 2 supporting the H_1 , and 2 being inconclusive (BF_n in Table 7). However, after pre-registration, we learned that this analysis is problematic and will discuss a more appropriate analysis in Section 4.7.2.

To summarise, we found illusion effects on grasping in all relevant contrasts, including contrasts where the distance of context elements was matched, so that the obstacle avoidance hypothesis as suggested by Haffenden et al. (2001) cannot explain these effects. The illusion effects are of similar size as in the perceptual tasks (classic perception, closed-loop ME, and open-loop ME) and most effects correlate significantly between grasping and the perceptual tasks.

4.5. Perceptually-matched conditions

For the perceptually-matched condition, we used two staircase procedures to determine a pair of target discs for each participant such that the one presented within the SF configuration and the one within the LF configuration would be perceived as equal in size. We called these discs SFx and LFy, respectively. Since the IEH assumes that grasping follows perception, it is now the IEH that predicts a null-difference in grasping between SFx and LFy (H_0), while the TVSH assumes that grasping follows physical size and that therefore the two discs should be grasped with different MGAs (H_1).

The LFy disc had an average diameter of 30.63 mm (± 0.06 mm) and the SFx of 29.48 mm (± 0.08 mm), such that the LFy disc was on average 3.91% larger than the SFx disc (Fig. 5). As specified in Section 2.4, the two discs were included in all perceptual tasks to confirm whether they were in fact perceived as being equally large. In the grasping task, these discs were used to detect influences of physical size on MGA that cannot be explained by perceived size. We found a difference in perceived size in the classic perception task [$t(143) = 3.99$, $p < .001$], indicating that the physically larger LFy was also perceived to be slightly larger (Fig. 6d). The same was true in all other tasks, although these differences were not significantly different from zero [open-loop ME: $t(143) = 1.95$, $p = .053$; closed-loop ME $t(143) = 1.18$, $p = .242$; grasping: $t(143) = 1.42$, $p = .158$]. Importantly, the MGA in grasping did not differ between LFy and SFx.

Using the same slope correction as in the physically-matched conditions (Figs. 6d and 8), we found the difference between LFy and SFx in MGA to be $1.01\% \pm .71\%$ of the mean object size, which may be interpreted as the effect of physical size on grasping that is not explained by perceived size as measured by our staircase. Note that the observed difference is in the same direction as in the perceptual tasks. Thus, the remaining perceptual difference may still explain some of the difference in MGA, which makes

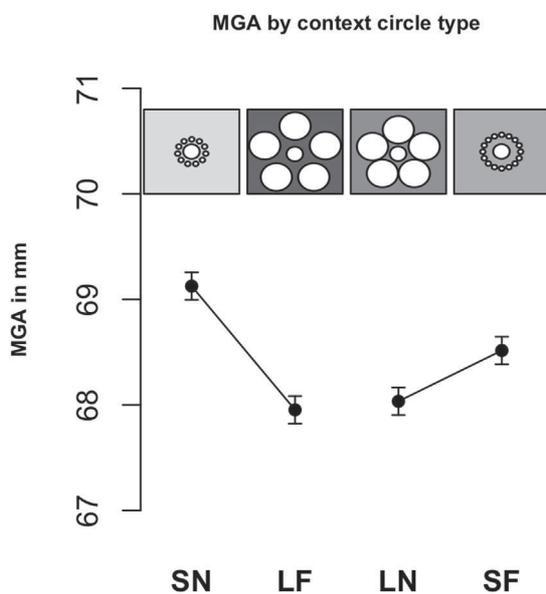


Fig. 7 – Mean grasping responses for the physically-matched conditions. Results for SN and LF replicate the literature, LN and SF are the adjusted configurations testing the obstacle-avoidance hypothesis. Results follow the predictions of the IEH (Fig. 3b) but not the predictions of the TVSH (Fig. 3c). Error bars indicate within-SEM for the pooled difference between context circle types and can therefore be used to interpret differences between conditions (Franz & Loftus, 2012; Loftus & Masson, 1994).

Table 5 – Comparisons between corrected illusion effects in grasping and in the perceptual tasks.

Comparison	SN-LF			SF-LF			SN-LN		
	t(143)	<i>p</i>	BF	t(143)	<i>p</i>	BF	t(143)	<i>p</i>	BF
Grasping versus 0	5.63	<.001**	-	2.95	.004*	–	5.96	<.001**	–
Perception versus grasping	–0.07	.942	>1000	0.65	.515	51.7	–0.24	.813	>1000
		<i>n.s.</i>			<i>n.s.</i>			<i>n.s.</i>	
ME CL versus grasping	0.94	.347	>1000	2.03	.044*	8.4	1.42	.157	>1000
		<i>n.s.</i>			<i>n.s.</i>			<i>n.s.</i>	
ME OL versus grasping	–0.22	.823	>1000	1.38	.169	22.1	0.40	.689	>1000
		<i>n.s.</i>			<i>n.s.</i>			<i>n.s.</i>	

Note: The contrasts SN–LF (traditional Ebbinghaus contrast), SF–LF and SN–LN (adjusted contrasts with controlled context circle distance) are tested for a difference against 0 in grasping (top row) and for a difference between each task and grasping (rows 2 to 4). We used slope-corrected illusion effects for the comparisons between tasks. Bayes factors compare the hypotheses of *illusion in grasping* = 0 (H0) versus *illusion in grasping* = *illusion in perceptual task* (H1). CL and OL are used to abbreviate closed-loop and open-loop, respectively. *n.s.* indicates non-significant, * indicates $p < .05$, ** indicates $p < .001$. Statistically significant *p*-values, as well as Bayes factors smaller than 1/3 or larger than 3, are given in italics.

this a slight overestimate for the effect of physical size. Importantly, not only did the difference in MGA not differ significantly from 0, the corrected residual difference was also almost exactly the same as the LFy–SFx differences found in the classic perceptual task ($1.05\% \pm .26\%$) and in ME (open-loop: $.93\% \pm .47\%$, closed-loop: $.48\% \pm .41\%$).

These results are supported by Bayes factors: we compared the hypothesis that MGA follows perceived size (H0) to the hypothesis that MGA follows physical size (H1). As perceived size, we used the results of our classic perceptual task as well as the results of closed-loop ME and open-loop ME. All Bayes factors decisively supported the H0 (grasping vs classic perception task BF: 1/554, grasping vs open-loop ME BF: 1/490, grasping vs closed-loop ME BF: 1/421), indicating that grasping followed perceived size.

To summarise, the results in the perceptually-matched conditions suggest that grasping follows perception and not physical size. This is consistent with our results in the physically-matched conditions.

4.6. Additional analyses

Our large sample size allowed us to run further pre-registered analyses that were more exploratory in nature and concerned general properties of our measures. Firstly, we analysed response slopes (Section 4.6.1), which are the basis for the corrected illusion effects. Secondly, we assessed grasping kinematics, testing the predictions of another theory of obstacle avoidance in grasping the Ebbinghaus illusion (Section 4.6.2).

Table 6 – Grasp parameters in the physically-matched condition by illusion condition and object size.

Context circle type	MGA in mm	MT in msec	MGA time in msec	Relative MGA time in %
LF	67.97 ± .95	958 ± 45	691 ± 34	75.50 ± 1.09
LN	68.03 ± .95	951 ± 44	692 ± 34	75.83 ± 1.07
SF	68.53 ± .98	956 ± 46	686 ± 33	75.34 ± 1.09
SN	69.12 ± .97	975 ± 47	690 ± 34	74.67 ± 1.15
Object size in mm				
28	66.48 ± .98	961 ± 46	690 ± 33	75.62 ± 1.11
30	68.60 ± .95	953 ± 45	688 ± 34	75.53 ± 1.06
32	70.16 ± .96	967 ± 46	690 ± 34	74.82 ± 1.09

Note: Between-subject standard errors are given for each cell.

4.6.1. Response slopes in different tasks

Our correction method takes into account the slopes of the response-functions. Based on previous studies, we had anticipated that grasping would show a response slope slightly smaller than 1 (e.g., Smeets & Brenner, 1999, report .82), while closed-loop ME (e.g., Dewar & Carey, 2006) and classic perception (e.g., Franz et al., 2000) should show a slope close to 1, and open-loop ME a slope larger than 1 (e.g., Haffenden & Goodale, 1998; Haffenden et al., 2001). In our data (Fig. 6b), we found slopes of $.92 \pm .05$ for grasping and $1.19 \pm .02$ for classic perception, which are both very similar to previous results (Franz et al., 2000; Smeets & Brenner, 1999). For closed-loop ME we observed a slope of $1.41 \pm .06$ and for open-loop ME a slope of $1.60 \pm .09$. Contrary to our expectations, the two ME slopes are numerically quite similar, although statistically they differ significantly [$t(143) = 2.36, p = .020$]. This is due to a larger than expected slope in closed-loop ME, while open-loop ME behaved roughly as we expected. Therefore, further research is needed to elucidate the reason for the relatively small slope in closed-loop ME in studies like Dewar and Carey (2006). See also Foster et al. (2012) for a further discussion of that study.

4.6.2. Grasping kinematics

In each grasping trial, we computed the time between the start of the movement and the occurrence of the MGA (MGA time), as well as between the start of the movement and touching the target disc (movement time, MT). MGA was reached on average at 75.35% of the total movement duration. This is consistent with classic studies on grasping (e.g., Jeannerod, 1984 reported the typical time of MGA to be between 74% and 81% in his participants, in one case earlier).

Analysing the grasping kinematics allowed us to test an idea put forward by Smeets et al. (2003). Based on their grasping model (Smeets & Brenner, 1999), they suggested that grasping kinematics can be used to detect more general obstacle avoidance mechanisms than those proposed by Haffenden et al. (2001). The main idea was that the cause of an increase of MGA might either be different contact points on the object (caused by a physical or illusory change of object size) or a different approach of the objects (possibly caused by obstacle avoidance mechanisms). While both effects could increase the MGA in similar ways, their influence on the relative timing of the MGA would be in opposite directions,

Table 7 – Full correlation table between illusion effects, all measures.

SN–LF	Grasping	Classic perception	ME closed-loop	ME open-loop
Grasping	Rel. = .31	$r = .18$ $p = .031^*$ MEC = .62 $BF_n = 1/41$ $BF_u = 3.0$	$r = .30$ $p < .001^{**}$ MEC = .51 $BF_n = 103$ $BF_u = 355$	$r = .11$ $p = .170$ MEC = .52 $BF_n = 1/112$ $BF_u = 1/1.2$
Classic perception		Rel. = .71	$r = .20$ $p = .014^*$ MEC = .69 $BF_n < 1/1000$ $BF_u = 5.0$	$r = .34$ $p < .001^{**}$ MEC = .68 $BF_n = 1.7$ $BF_u > 1000$
ME closed-loop			Rel. = .39	$r = .27$ $p = .001^*$ MEC = .57 $BF_n = 1/1.6$ $BF_u = 66$
ME open-loop SF–LF				Rel. = .40
Grasping	Rel. = .32	$r = .18$ $p = .027^*$ MEC = .61 $BF_n = 1/64$ $BF_u = 3.4$	$r = .19$ $p = .026^*$ MEC = .43 $BF_n = 1.5$ $BF_u = 5.6$	$r = .15$ $p = .073$ MEC = .41 $BF_n = 1/1.2$ $BF_u = 2.4$
Classic perception		Rel. = .62	$r = .32$ $p < .001^{**}$ MEC = .54 $BF_n = 282$ $BF_u = 671$	$r = .23$ $p = .006^*$ MEC = .51 $BF_n = 6.9$ $BF_u = 18$
ME closed-loop			Rel. = .23	$r = .25$ $p = .002^*$ MEC = .36 $BF_n = 42$ $BF_u = 61$
ME open-loop SN–LN				Rel. = .21
Grasping	Rel. = .27	$r = -.04$ $p = .640$ MEC = .58 $BF_n < 1/1000$ $BF_u = 1/8.7$	$r = .26$ $p = .002^*$ MEC = .48 $BF_n = 16$ $BF_u = 66$	$r = .06$ $p = .490$ MEC = .44 $BF_n = 1/78$ $BF_u = 1/2.4$
Classic perception		Rel. = .65	$r = .13$ $p = .120$ MEC = .64 $BF_n < 1/1000$ $BF_u = 1/1.2$	$r = .39$ $p < .001^{**}$ MEC = .58 $BF_n > 1000$ $BF_u > 1000$
ME closed-loop			Rel. = .37	$r = .27$ $p = .001^*$ MEC = .48 $BF_n = 30$ $BF_u = 82$
ME open-loop				Rel. = .30

Note: MEC stands for “maximal expected correlation”, as described in Section 2.5. Rel. stands for ‘reliability’, r denote a correlation. P-values are given for a t-test of each correlation against 0; Bayes factor BF_n compares the hypotheses $correlation = 0$ (H0) versus $correlation = MEC$ (H1); Bayes factor BF_u is a non-preregistered analysis that compares the hypotheses $correlation = 0$ (H0) versus $correlation$ is positive (i.e., between zero and MEC, H1). n.s. indicates non-significant, * indicates $p < .05$, ** indicates $p < .001$. Statistically significant p-values, as well as Bayes factors smaller than 1/3 or larger than 3, are given in italics.

such that obstacle avoidance mechanisms would lead to an earlier MGA, while a larger object would result in a later MGA. Such opposite effects on the timing of MGA might serve as evidence for obstacle avoidance mechanisms, although the expected difference is small (Smeets & Brenner, 1999; Smeets

et al., 2003) and it is unclear whether these obstacle avoidance mechanisms would be comparable to those proposed by Haffenden et al. (2001). In their study, Smeets et al. (2003) did not find any such effects and argued that the expected effects are too small to be detected with the sample size and power of

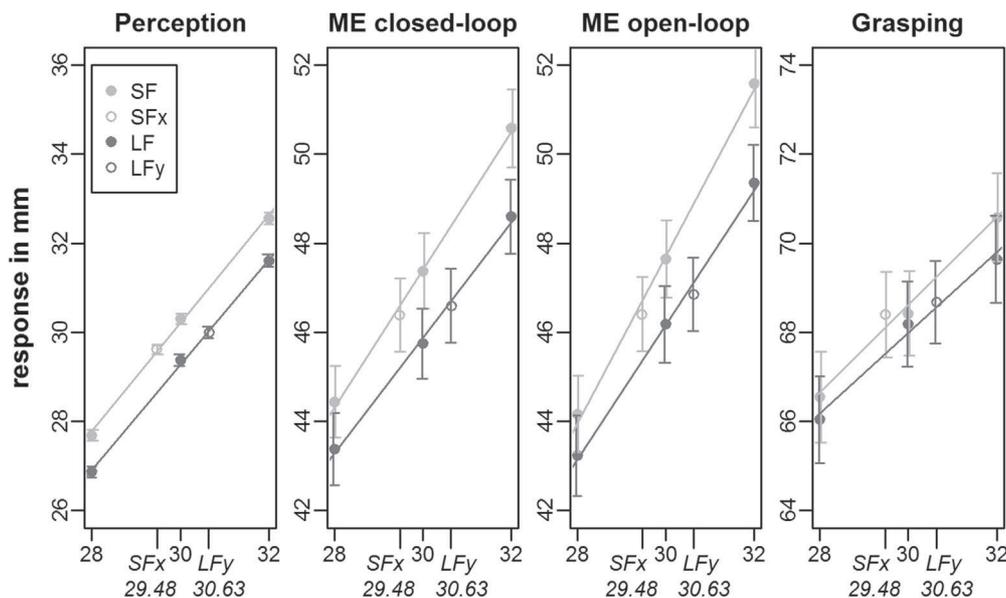


Fig. 8 – Mean responses to configurations LF and SF for physically-matched and perceptually-matched conditions. The x-axis indicates object size in mm. SFx and LFy are the data from the perceptually-matched condition (depicted at the means of the corresponding matched sizes, i.e., at 29.48 mm and 30.63 mm, respectively). The responses in the perceptually-matched condition are fully consistent with the responses in the physically-matched conditions. This indicates that there is no qualitative difference between perceptually-matched and physically-matched conditions. The regression lines were obtained from the physically-matched condition. Error bars indicate between-subjects SEM.

their study (p. 319). Given our very large sample size, we were in a better position to test this idea.

We tested whether an increase of MGA due to a change of physical size had opposite effects on MGA time than an increase due to the illusion configuration. A 3 (object size) × 4 (context circle type) ANOVA with relative MGA time as the dependent variable revealed main effects of both factors: object size [$F(2, 286) = 3.57, p = .029$] and context circle type [$F(3, 429) = 2.87, p = .036$]. However, the interaction was not significant [$F(6, 858) = 1.80, p = .096$], thereby indicating that we did not find the predicted opposite effects. Instead, we found that no matter if MGA increased as a result of an increase in object size or because of a change of the illusion configuration, the MGA always occurred slightly earlier (Table 6). This is surprising, given that previous research has generally found larger MGAs to occur later (Jeannerod, 1984; Smeets & Brenner, 1999), but the finding remains that actual size and illusory size did not impact the timing of the MGA differently.

In short, we found no evidence for general obstacle-avoidance mechanisms as suggested by the grasping model of Smeets and Brenner (1999). This is consistent with our overall conclusions that the effects of Ebbinghaus illusion displays on grasping cannot be explained by obstacle avoidance (Haffenden & Goodale, 2000; Haffenden et al., 2001). However, our study was not designed to test Smeets and Brenner's model (1999). Therefore, our results do not constitute evidence against it. Also, even if grasping perfectly followed the perceptual effect of the Ebbinghaus illusion, this does not necessarily contradict the Smeets and Brenner (1999) model, as it is possible that the Ebbinghaus illusion affects

grasp position (the central variable in this model) in a similar way as object size (the central variable in more traditional accounts of grasping), such that the model could still be consistent with our results and conclusions.

4.7. Post-hoc analyses

4.7.1. Comparing raw illusion effects

As has been laid out in Sections 1.2 and 2.5, we consider it necessary to correct for the slope of the response function before we compare illusion effects obtained from different tasks. However, it may be interesting to also analyse the data without these corrections, especially since some researchers are sceptical of this procedure (Goodale, 2014; Westwood & Goodale, 2011). Therefore, we also compared raw illusion effects for grasping and all perceptual tasks (cf. Table A.2 in the Appendix).

In the physically-matched conditions, we found that for all relevant illusion contrasts (SN–LF, SF–LF and SN–LN), there was no significant difference between the illusion effects in grasping and classic perception (all $p > .08$), while all comparisons of illusion effects in grasping and ME were significant (all $p < .03$). This is fully consistent with the literature, as among studies that did not apply slope-correction, those that compared grasping to ME (e.g., Haffenden & Goodale, 1998) found significant differences between the measures, while those that compared grasping to classic perception tasks (Franz et al., 2001; Pavani et al., 1999) did not.

Interestingly, all ME versus classic perception comparisons (SN–LF, SF–LF and SN–LN) also yielded significant differences

(all $p < .002$), except for one [open-loop ME – classic perception, SN–LF: $t(143) = 1.62, p = .108$]. Almost all of these apparent differences between classic perception and ME disappear when slope-correction is applied (see [Tables A.1 and A.2 in the Appendix](#)). This emphasises the importance of the correction, as even tasks that are unambiguously considered to be perceptual by the TVSH do not produce coherent results without correction (for a similar argument, see [Hesse, Franz, & Schenk, 2016](#)).

In the *perceptually-matched conditions*, we found no differences between the LFy–SFx contrasts [grasping – open-loop ME: $t(143) = .61, p = .540$; grasping – closed-loop ME: $t(143) = .31, p = .760$; grasping – classic perception: $t(143) = .52, p = .606$] or between LFy–SFx and 0. This is confirmed by the Bayes factor for the comparison of H0: LFy–SFx = 0 versus H1: LFy–SFx = physical difference (uncorrected), which also decisively supported the H0 (BF = 1/1988). These results are fully consistent with the results we found with slope correction (which is not surprising, as the perceptually-matched conditions would not require a slope correction if the perceptual match was perfect).

4.7.2. A better Bayesian analysis for the correlations between measures

For the correlations between dependent measures (Section 4.4), we had pre-registered a Bayesian analysis that used a normal-distribution centred at the maximal expected correlation as the H1-prior. However, in the meantime we learned that in a situation where one expects the effect to be larger than 0 but smaller than an upper bound, it is more appropriate to specify a uniform distribution from 0 to the upper bound as the H1 ([Dienes, 2008](#), chap. 4). Because the maximal expected correlation constitutes an upper bound (e.g., [Nunnally, 1967](#); [Vul, Harris, Winkielman, & Pashler, 2009](#)), we recalculated the Bayesian analysis using a uniform distribution between 0 and the maximal expected correlation as prior, thereby contrasting the hypothesis that there is no correlation (H0) with the hypothesis that the correlation is between 0 and the maximal expected correlation (H1). For the most interesting correlations between grasping and the perceptual measures we found support for the H1 in 5 cases, support for the H0 in only one case and inconclusive data in 3 cases, (see BFu in [Table 7](#)). This is consistent with the frequentist analyses (see p -values in [Table 7](#) and Section 4.4). Both results suggest that the correlations between grasping and the perceptual measures are statistically reliable, as predicted by the IEH.

5. Discussion

We tested whether there is an effect of the Ebbinghaus illusion on grasping using a paradigm that accounted for the alternative explanation of obstacle avoidance. We took great care to consider all methodological criticism raised in previous studies. To this end, we replicated and extended two studies that had previously reported inconsistent results on grasping the Ebbinghaus illusion, and specifically on obstacle avoidance ([Franz et al., 2003](#); [Haffenden et al., 2001](#)). Also, we created a symmetric situation with regard to the problem of “proving the null-hypothesis”: in addition to calculating Bayes factors, we employed both a standard physically-matched

design where we manipulated the perceptual context of the target discs (IEH predicts a difference in grasping between physically-matched but perceptually different discs), and a perceptually-matched design similar to that used by [Aglioti et al. \(1995\)](#), (TVSH predicts a difference in grasping between perceptually-matched but physically different discs). The experiment was run in four labs ([Table 3](#)) to achieve more statistical power and to strengthen the generalisability of our results. Together, these factors allow us to draw strong conclusions from our results.

The main reason why we focussed on the Ebbinghaus illusion are the many possible confounds and non-obvious methodological issues described in previous sections. While some work on other illusions has reported dissociations between grasping and perception (e.g., [Ganel et al., 2008](#); [Stöttinger et al., 2012](#)), the Ebbinghaus illusion is by far the most studied paradigm. In consequence, the discussion has advanced to a point where potential confounds related to this paradigm have been identified and can thus be avoided. To the best of current knowledge, we conducted a confound-free test of whether or not grasping is affected by visual illusions in a similar way as perception.

5.1. Physically-matched conditions: grasping and perception are affected by the illusion

Our results clearly show that there is an illusion effect on grasping. In all labs, having discs surrounded by small context circles (thus appearing larger in size) consistently caused a larger MGA than having the same discs surrounded by large context circles (thus appearing smaller in size).

Importantly, the effect on MGA persisted not only for the SN–LF comparison, where an effect has been frequently reported (for reviews, see [Franz & Gegenfurtner, 2008](#); [Schenk et al., 2011](#); [Schenk & McIntosh, 2010](#)), but also for comparisons in which the context–element distance was equal for small and large context circles, the SF–LF and SN–LN comparisons ([Fig. 6](#) and [Table 5](#)). Hence, our study yielded results similar to those reported by [Franz et al. \(2003\)](#), but is in contrast to the findings reported by [Haffenden et al. \(2001\)](#). Since the distance between the context circles and target discs is equal, these illusion effects can only be explained by the difference in context circle size, thus matching the predictions of the IEH, but not those of the obstacle avoidance hypothesis. The key assumption of the obstacle avoidance hypothesis is that participants fit their fingers between target and context elements in the far conditions and grasp around the entire stimulus display in the near conditions ([Fig. 2](#)). Thus, finding a difference in MGA between configurations using the same context circle distance (SF and LF, SN and LN) as we did is incompatible with the obstacle avoidance hypothesis. An even stronger demonstration that obstacle avoidance cannot explain these illusion effects is obtained by comparing the SF and LN conditions ([Fig. 7](#)). The perceived size of the disc in SF is larger than in LN, which should result in a larger MGA in the SF condition according to the IEH, while the obstacle avoidance account would predict the opposite, a larger MGA in the LN condition ([Fig. 3](#)).

We also found, consistent with the IEH, that illusion effects in perception and in grasping tend to correlate ([Table 7](#)). The correlations are small, and only 5 of the 9 tested correlations are significantly different from zero. However, this is to be expected

when correlating two measures with relatively low reliability. To reliably detect such correlations requires very large sample sizes, even larger than the already unusually large sample size employed in our study. Therefore, we interpret these results as consistent with the notion that participants who displayed a large perceptual illusion effect also tended to display a larger illusion effect in grasping. This would be predicted if a common size representation underlies both tasks. A similar result was recently found for perceptual illusions and saccades (Dassonville & Reed, 2015). In addition to MGA, saccades are another prominent action measure that has been frequently used to argue for a functional subdivision between vision-for-action and vision-for-perception (but see Bruno, Knox, & de Grave, 2010; de Brouwer, Smeets, Gutteling, Toni, & Medendorp, 2015).

5.2. Perceptually-matched conditions: physical size does not trump perceived size

In the physically-matched condition, the strong version of the TVSH predicts no illusion effect on grasping (H0), while the IEH predicts an illusion effect (H1). Arguing for the H0 is often seen as problematic (Westwood & Goodale, 2011), especially since some effect of perception on grasping has been demonstrated in many paradigms (Bruno & Franz, 2009; McIntosh & Lashley, 2008) and may be compatible with a weaker version of the TVSH (Goodale, 2008). Therefore, we added the perceptually-matched condition: here, the TVSH predicts a difference in grasping for physically different but perceptually matched discs (H1), while the IEH predicts no difference (H0).

Consistent with the IEH, we did not find a difference (Figs. 6d and 8). As our power-analysis and the Bayes factors reveal, our sample is large enough to interpret these null results as evidence that participants did not scale their grip to the physical size of the discs but to the perceived size, thereby indicating an illusion effect on grasping. Because the distance between context circles and target discs was equal in the perceptually-matched condition, these illusion effects cannot be explained by the obstacle avoidance hypothesis.

Our results also indicate that the matching procedure did not work perfectly, but this is unproblematic for our argument for two reasons: firstly, the deviation from 0 in the classic perceptual task was small. We argue that with a step size of .25 mm, and controlling for the illusion's superadditivity (see Foster & Franz, 2014), our match was close to optimal. As explained in Section 1.6, we did not expect to be able to achieve a perfect match. Secondly, the physically larger object was also perceived to be slightly larger in the classic perceptual task. This means that the physical difference between the two targets was larger than necessary to achieve perceptual equivalence. Consequently, we should have found an even larger difference in grasping than we would have had we been able to create a perfect match. Thus, if anything, the perceptually-matched condition was over-sensitive to detecting a dissociation. The fact that this dissociation was not found suggests that the illusion effect on grasping is sufficiently pronounced to eliminate the physical difference of the target objects. In summary, the results in the perceptually-matched condition are consistent with the results of the physically-matched conditions: both indicate that the Ebbinghaus illusion affects grasping.

5.3. Is there no effect of obstacle avoidance at all?

For a reader with a background in motor control, it might seem implausible to argue that obstacle avoidance has no effect on grasping. In fact, it is well known that distractors can affect movements (e.g., Tipper, Lortie, & Baylis, 1992). However, what we tested and argue against is only one very specific obstacle avoidance hypothesis, namely the notion that the context circles produce distinct grasping behaviour identical to the perceptual illusion in the "classic" illusion display (SN-LF) as used by Aglioti et al. (1995) and many studies after that. This specific obstacle-avoidance hypothesis assumes that in the far condition participants aim to fit their grasping fingers between target and context whereas in the near condition the fingers do not fit in this space and therefore grasp larger. Haffenden and Goodale (2000) and Haffenden et al. (2001) proposed this obstacle avoidance mechanism in order to reconcile the existence of an effect of the Ebbinghaus illusion on grasping in the traditional display with their notion that grasping is immune to the illusion. They argued that the observed illusion effects on grasping in those studies were methodological artefacts due to imperfect stimuli. They suggested that if better stimuli were used – such as stimuli with equated distance of the context elements – the Ebbinghaus illusion would not affect grasping. We tested this claim and can safely refute it.

Note that for our claim it is not necessary that the context elements have no obstacle-like effects on grasping at all. For example, de Grave et al. (2005) found (small) effects of rotating Ebbinghaus displays on grasping parameters other than MGA. What we do claim is that the context elements do not affect MGA in a way that mimics the perceptual illusion effect. Even with our very large sample, we did not find an obstacle avoidance effect on MGA. Thus, it seems unlikely that we have missed an effect large enough to be reliably detected by studies with much smaller samples. Any obstacle effects of the context circles on the MGA, if they exist, would be too small by far to explain the illusion effects that were found in grasping.

6. Conclusion

In summary, we can draw the following conclusions: there is no doubt that there is an effect of the Ebbinghaus illusion on grasping. This effect correlates with the illusion effect on classic perceptual measures as well as on manual estimation. Crucially, this effect cannot be explained as an artefact of obstacle avoidance. A dissociation between vision-for-perception and vision-for-action when grasping the Ebbinghaus illusion, as suggested by the TVSH, is not supported.

Acknowledgements

We thank Brian Roberts and Mike Harris for responding to our questions regarding their paper; Zoltan Dienes for advice on Bayes factors; Denise Fischer, Melanie Römer, Ioana Stanciu, Aleksandra Romanczuk, Stefano Uccelli, Nuria Martos Sánchez, and Rosa María Beño Ruiz de la Sierra for help collecting data; Eva Viviani for managing data collection in

Parma. We thank Maurizio Gentilucci for letting us use his lab, and the Centro Intradipartimentale Mente e Cervello (CIMeC), University of Trento, and especially Francesco Pavani for lending us his motion tracking equipment. We thank Rachel Foster for proofreading. KKK was supported by a Ph.D. scholarship as part of a grant to VHF within the International Graduate Research Training Group on Cross-Modal Interaction in Natural and Artificial Cognitive Systems (CINACS; DFG IKG-1247) and TS by a grant (DFG – SCHE 735/3-1); both from the German Research Council.

Appendix A. Supplementary results tables

process (and will not be reported in detail here). Second, we wished to examine how large the illusion effects and variation between responses would be, so that we would be able to create a graded series of comparison stimuli that would not result in floor or ceiling effects.

In this task, 8 different Ebbinghaus illusion displays were displayed to the participants: SN, SF, LN, LF as described in Section 1.6, and 4 versions of LVF, each with a different annulus diameter (67, 82, 96 and 110 mm). Target circles were 28, 30, and 32 mm in diameter. Each of the resulting 24 conditions was presented 6 times to each participant, resulting in a total of 144 trials per participant. The task was to determine which one of 8 comparison circles was equal in size to the

Table A.1 – Corrected illusion effects tested against each other, all measures.

Comparison		SN–LF		SF–LF		SN–LN	
		t(143)	<i>p</i>	t(143)	<i>p</i>	t(143)	<i>p</i>
Grasping	versus perception	-.07	.942 n.s.	.65	.515 n.s.	-.24	.813 n.s.
	versus ME CL	.94	.347 n.s.	2.03	.044*	1.42	.157 n.s.
	versus ME OL	-.22	.823 n.s.	1.38	.169 n.s.	.40	.689 n.s.
Perception	versus ME CL	1.71	.089 n.s.	2.62	.010*	2.61	.010*
	versus ME OL	-.24	.810 n.s.	1.35	.179 n.s.	.85	.398 n.s.
ME CL	versus ME OL	1.34	.181 n.s.	.73	.469 n.s.	1.09	.276 n.s.

Note: n.s. indicates non-significant, * indicates $p < .05$, ** indicates $p < .001$. Statistically significant p -values are given in italics.

Table A.2 – Uncorrected illusion effects tested against each other, all measures.

Comparison		SN–LF		SF–LF		SN–LN	
		t(143)	<i>p</i>	t(143)	<i>p</i>	t(143)	<i>p</i>
Grasping	versus perception	1.55	.123 n.s.	1.76	.081 n.s.	1.35	.180 n.s.
	versus ME CL	3.92	<.001**	4.27	<.001**	4.66	<.001**
	versus ME OL	2.26	.025*	3.57	<.001**	3.18	.002*
Perception	versus ME CL	3.26	.001**	4.15	<.001**	3.98	<.001**
	versus ME OL	1.61	.108 n.s.	2.85	.005*	2.97	.003*
ME CL	versus ME OL	.71	.476 n.s.	.05	.959 n.s.	.31	.756 n.s.

Note: n.s. indicates non-significant, * indicates $p < .05$, ** indicates $p < .001$. Statistically significant p -values are given in italics.

Appendix B. Pilot data

We tested 4 participants (mean age 33.5 years) on a simple perceptual judgement task to examine two issues: First, we considered including a “large-very far” (LVF) condition as an extra test of obstacle avoidance and wanted to gauge the perceptual illusion effect with different target-annulus distances. This condition was discarded during the review

target circle and to press the corresponding number on the numpad of a standard German QWERTZ-keyboard. The comparison circles were displayed on the left side of the screen, sorted by size, ascending, in steps of 1.136 mm (4 pixels). The sizes were pseudo-randomised, but always chosen such that the smallest comparison circle was at least 8 pixels (2.272 mm) smaller, and the largest comparison circle at least 8 pixels larger than the target. The specifications and mean illusion effects of interest can be found in Table B1.

Table B1 – Conditions and corresponding mean illusion effects in our perceptual pilot data.

Condition	Number of context circles	Diameter of context circles (in mm)	Inner diameter of annulus (in mm)	Mean illusion effect \pm SD (in mm)
SN	11	10	38	1.42 \pm .255
SF	16	10	54	.71 \pm .586
LN	5	54	38	-1.18 \pm .626
LF	5	54	54	-1.23 \pm .673

The observed illusion effects are in the expected range, with 97.1% of the responses being within 12 pixels (3.41 mm) of the actual size. The remaining 2.9% of all responses were within 16 pixels (4.54 mm) of the actual size. Based on these results, we felt confident that our comparison stimuli ranging from 5 mm smaller than the smallest target to 5 mm larger than the largest target would produce no floor or ceiling effects.

REFERENCES

- Aglioti, S., DeSouza, J. F. X., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology*, 5(6), 679–685.
- de Brouwer, A. J., Smeets, J. B. J., Gutteling, T. P., Toni, I., & Medendorp, W. P. (2015). The Müller-Lyer illusion affects visuomotor updating in the dorsal visual stream. *Neuropsychologia*, 77, 119–127. <http://dx.doi.org/10.1016/j.neuropsychologia.2015.08.012>.
- Bruno, N. (2016). Visual illusions in action. In A. Shapiro, & D. Todorovic (Eds.), *The Oxford Compendium of visual illusions*. Oxford, UK: Oxford University Press.
- Bruno, N., & Franz, V. H. (2009). When is grasping affected by the Müller-Lyer illusion? A quantitative review. *Neuropsychologia*, 47, 1421–1433.
- Bruno, N., Knox, P. C., & de Grave, D. D. J. (2010). A meta-analysis of the effect of the Müller-Lyer illusion on saccadic eye movements: no general support for a dissociation of perception and oculomotor action. *Vision Research*, 50(24), 2671–2682. <http://dx.doi.org/10.1016/j.visres.2010.09.016>.
- Carey, D. P. (2001). Do action systems resist visual illusions? *Trends in Cognitive Sciences*, 5(3), 109–113.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Psychology Press.
- Coren, S., & Enns, J. T. (1993). Size contrast as a function of conceptual similarity between test and inducers. *Perception & Psychophysics*, 54(5), 579–588.
- Coren, S., & Girgus, J. S. (1972). A comparison of five methods of illusion measurement. *Behavior Research Methods & Instrumentation*, 4(5), 240–244. <http://dx.doi.org/10.3758/BF03210006>.
- Dassonville, P., & Reed, S. A. (2015). The Two-Wrongs model explains perception-action dissociations for illusions driven by distortions of the egocentric reference frame. *Frontiers in Human Neuroscience*, 9(March), 1–16. <http://dx.doi.org/10.3389/fnhum.2015.00140>.
- Dewar, M. T., & Carey, D. P. (2006). Visuomotor “immunity” to perceptual illusion: a mismatch of attentional demands cannot explain the perception-action dissociation. *Neuropsychologia*, 44(8), 1501–1508. <http://dx.doi.org/10.1016/j.neuropsychologia.2005.11.010>.
- Dienes, Z. (2008). *Understanding psychology as a science. An introduction to scientific and statistical inference*. New York: Palgrave Macmillan.
- Dienes, Z. (2011). Bayesian versus Orthodox statistics: which side are you on? *Perspectives on Psychological Science*, 6(3), 274–290.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior Research Methods*, 41, 1149–1160.
- Foster, R. M., & Franz, V. H. (2014). Superadditivity of the Ebbinghaus and Müller-Lyer illusions depends on the method of comparison used. *Perception*, 43(8), 783–795. <http://dx.doi.org/10.1068/p7802>.
- Foster, R. M., Kleinholdermann, U., Leifheit, S., & Franz, V. H. (2012). Does bimanual grasping of the Müller-Lyer illusion provide evidence for a functional segregation of dorsal and ventral streams? *Neuropsychologia*, 50, 3392–3402.
- Franz, V. H. (2003). Manual size estimation: a neuropsychological measure of perception? *Experimental Brain Research*, 151, 471–477.
- Franz, V. H. (2007). Ratios: A short guide to confidence limits and proper use. Retrieved from <http://arxiv.org/pdf/0710.2024v1.pdf>.
- Franz, V. H., Bühlhoff, H. H., & Fahle, M. (2003). Grasp effects of the Ebbinghaus illusion: obstacle avoidance is not the explanation. *Experimental Brain Research*, 149, 470–477.
- Franz, V. H., Fahle, M., Bühlhoff, H. H., & Gegenfurtner, K. R. (2001). Effects of visual illusions on grasping. *Journal of Experimental Psychology: Human Perception and Performance*, 27(5), 1124–1144.
- Franz, V. H., & Gegenfurtner, K. R. (2008). Grasping visual illusions: consistent data and no dissociation. *Cognitive Neuropsychology*, 25, 920–950.
- Franz, V. H., Gegenfurtner, K. R., Bühlhoff, H. H., & Fahle, M. (2000). Grasping visual illusions: no evidence for a dissociation between perception and action. *Psychological Science*, 11(1), 20–25.
- Franz, V. H., Hesse, C., & Kollath, S. (2009). Visual illusions, delayed grasping, and memory: no shift from dorsal to ventral control. *Neuropsychologia*, 47, 1518–1531.
- Franz, V. H., & Loftus, G. R. (2012). Standard errors and confidence intervals in within-subjects designs: generalizing Loftus and Masson (1994) and avoiding the biases of alternative accounts. *Psychonomic Bulletin & Review*, 19, 395–404.
- Franz, V. H., Scharnowski, F., & Gegenfurtner, K. R. (2005). Illusion effects on grasping are temporally constant not dynamic. *Journal of Experimental Psychology: Human Perception and Performance*, 31(6), 1359–1378.
- Ganel, T., Tanzer, M., & Goodale, M. A. (2008). A double dissociation between action and perception in the context of visual illusions: opposite effects of real and illusory size. *Psychological Science*, 19(3), 221–225.
- Girgus, J. S., Coren, S., & Agdern, M. (1972). The interrelationship between the Ebbinghaus and Delboeuf illusions. *Journal of Experimental Psychology*, 95(2), 453–455.
- Glover, S., & Dixon, P. (2002). Dynamic effects of the Ebbinghaus illusion in grasping: support for a planning/control model of action. *Perception & Psychophysics*, 64(2), 266–278.
- Gonzalez, C. L. R., Ganel, T., & Goodale, M. A. (2006). Hemispheric specialization for the visual control of action is independent of handedness. *Journal of Neurophysiology*, 95, 3496–3501.
- Gonzalez, C. L. R., Ganel, T., Whitwell, R. L., Morrissey, B., & Goodale, M. A. (2008). Practice makes perfect, but only with the right hand: sensitivity to perceptual illusions with awkward grasps decreases with practice in the right but not the left hand. *Neuropsychologia*, 46, 624–631.
- Goodale, M. A. (2008). Action without perception in human vision. *Cognitive Neuropsychology*, 25(7–8), 891–919.
- Goodale, M. A. (2014). How (and why) the visual control of action differs from visual perception. *Proceedings of the Royal Society B*, 281, 20140337. <http://dx.doi.org/10.1098/rspb.2014.0337>.
- Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1), 20–25.
- Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, 349, 154–156.
- de Grave, D. D. J., Biegstraaten, M., Smeets, J. B. J., & Brenner, E. (2005). Effects of the Ebbinghaus figure on grasping are not only due to misjudged size. *Experimental Brain Research*, 163, 58–64.
- Haffenden, A. M., & Goodale, M. A. (1998). The effect of pictorial illusion on prehension and perception. *Journal of Cognitive Neuroscience*, 10(1), 122–136.

- Haffenden, A. M., & Goodale, M. A. (2000). Independent effects of pictorial displays on perception and action. *Vision Research*, 40, 1597–1607.
- Haffenden, A. M., Schiff, K. C., & Goodale, M. A. (2001). The dissociation between perception and action in the Ebbinghaus illusion: nonillusory effects of pictorial cues on grasp. *Current Biology*, 11, 177–181.
- Hesse, C., Franz, V. H., & Schenk, T. (2016). Pointing and anti-pointing in Müller-Lyer figures: why illusion effects need to be scaled. *Journal of Experimental Psychology: Human Perception and Performance*, 42(1), 90–102.
- Himmelbach, M., Boehme, R., & Karnath, H.-O. (2012). 20 years later: a second look on DF's motor behaviour. *Neuropsychologia*, 50, 139–144.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2), 65–70.
- Jandó, G., Agostini, T., Galmonte, A., & Bruno, N. (2003). Measuring surface achromatic color: toward a common measure for increments and decrements. *Behavior Research Methods, Instruments & Computers*, 35(1), 70–81.
- Jeannerod, M. (1984). The timing of natural prehension movements. *Journal of Motor Behavior*, 16(3), 235–254.
- Jeffreys, H. (1961). Theory of probability. In N. F. Mott, E. C. Bullard, & D. H. Wilkinson (Eds.), *The international series of monographs on physics* (3rd ed.). Oxford: Oxford University Press.
- Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1, 476–490.
- McIntosh, R. D., & Lashley, G. (2008). Matching boxes: familiar size influences action programming. *Neuropsychologia*, 46, 2441–2444.
- Milgram, P. (1987). A spectacle-mounted liquid-crystal tachistoscope. *Behavior Research Methods, Instruments & Computers*, 19(5), 449–456.
- Milner, A. D., Dijkerman, H. C., Pisella, L., McIntosh, R. D., Tilikete, C., Vighetto, A., et al. (2001). Grasping the past: delay can improve visuomotor performance. *Current Biology*, 11, 1896–1901.
- Milner, A. D., Ganel, T., & Goodale, M. A. (2012). Does grasping in patient D.F. depend on vision? *Trends in Cognitive Sciences*, 16(5), 256–257.
- Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action* (1st ed.). Oxford: Oxford University Press Inc.
- Milner, A. D., & Goodale, M. A. (2006). *The visual brain in action* (2nd ed.). Oxford: Oxford University Press Inc.
- Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, 46, 774–785.
- Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. *Nature Neuroscience*, 9(3), 429–434.
- Nunnally, J. C. (1967). *Psychometric theory*. New York, NY: McGraw-Hill.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9(1), 97–113.
- Pavani, F., Boscagli, I., Benvenuti, F., Rabuffetti, M., & Farnè, A. (1999). Are perception and illusion affected differently by the Titchener circles illusion? *Experimental Brain Research*, 127, 95–101.
- Post, R. B., & Welch, R. B. (1996). Is there dissociation of perceptual and motor responses to figural illusions? *Perception*, 25(5), 569–581.
- Ramachandran, V. S., & Blakeslee, S. (1999). *Phantoms in the Brain: Probing the mysteries of the human mind*. New York, NY: William Morrow.
- Roberts, B., Harris, M. G., & Yates, T. A. (2005). The roles of inducer size and distance in the Ebbinghaus illusion (Titchener circles). *Perception*, 34, 847–856.
- Schenk, T. (2006). An allocentric rather than perceptual deficit in patient D.F. *Nature Neuroscience*, 9(11), 1369–1370.
- Schenk, T. (2010). Visuomotor robustness is based on integration not segregation. *Vision Research*, 50, 2627–2632.
- Schenk, T. (2012). No dissociation between perception and action in patient DF when haptic feedback is withdrawn. *The Journal of Neuroscience*, 32(6), 2013–2017.
- Schenk, T., Franz, V. H., & Bruno, N. (2011). Vision-for-perception and vision-for-action: which model is compatible with the available psychophysical and neuropsychological data? *Vision Research*, 51, 812–818.
- Schenk, T., & McIntosh, R. D. (2010). Do we have independent visual streams for perception and action? *Cognitive Neuroscience*, 1(1), 52–62.
- Smeets, J. B. J., & Brenner, E. (1999). A new view on grasping. *Motor Control*, 3(3), 237–271.
- Smeets, J. B. J., & Brenner, E. (2006). 10 Years of illusions. *Journal of Experimental Psychology: Human Perception and Performance*, 32(6), 1501–1504.
- Smeets, J. B. J., Glover, S., & Brenner, E. (2003). Modeling the time-dependent effect of the Ebbinghaus illusion on grasping. *Spatial Vision*, 16(3–4), 311–324.
- Stevens, S. S. (1959). Cross-modality validation of subjective scales for loudness, vibration, and electric shock. *Journal of Experimental Psychology*, 57(4), 201–209.
- Stöttinger, E., Pfusterschmied, J., Wagner, H., Danckert, J., Anderson, B., & Perner, J. (2012). Getting a grip on illusions: replicating Stöttinger et al. [Exp Brain Res (2010) 202:79–88] results with 3-D objects. *Experimental Brain Research*, 216(1), 155–157. <http://dx.doi.org/10.1007/s00221-011-2912-8>.
- Tipper, S. P., Lortie, C., & Baylis, G. C. (1992). Selective reaching: evidence for action-centered attention. *Journal of Experimental Psychology: Human Perception and Performance*, 18(4), 891–905.
- Vul, E., Harris, C., Winkielman, P., & Pashler, H. (2009). Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Psychological Science*, 4(3), 274–290. <http://dx.doi.org/10.1111/j.1745-6924.2009.01132.x>.
- Weiskrantz, L. (1990). Outlooks for blindsight: explicit methodologies for implicit processes. *Proceedings of the Royal Society B*, 239, 247–278.
- Westwood, D. A., & Goodale, M. A. (2011). Converging evidence for diverging pathways: neuropsychology and psychophysics tell the same story. *Vision Research*, 51, 804–811.
- Whitwell, R. L., Milner, A. D., Cavina-Pratesi, C., Byrne, C. M., & Goodale, M. A. (2014). DF's visual brain in action: the role of tactile cues. *Neuropsychologia*, 55, 41–50. <http://dx.doi.org/10.1016/j.neuropsychologia.2013.11.019>.

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Discussion forum

Do visual illusions affect grasping? Considerable progress in a scientific debate. A reply to Whitwell & Goodale, 2016

Karl K. Kopiske^{a,*}, Nicola Bruno^b, Constanze Hesse^c,
Thomas Schenk^d and Volker H. Franz^e

^a Center for Neuroscience and Cognitive Systems@UniTn, Istituto Italiano di Tecnologia (IIT), Rovereto, TN, Italy

^b Dipartimento di Neuroscienze, Università di Parma, Unità di Psicologia, Parma, Italy

^c University of Aberdeen, School of Psychology, William Guild Building, King's College, Aberdeen, United Kingdom

^d Ludwig-Maximilians Universität München, Department of Psychology, Munich, Germany

^e University of Tübingen, Institute of Computer Sciences, Department of Experimental Cognitive Sciences, Tübingen, Germany

1. Our registered report and the illusion debate

When we set out to perform our preregistered study (Kopiske, Bruno, Hesse, Schenk, & Franz, 2016), our goal was to clarify whether or not grasping is affected by the Ebbinghaus illusion. This seemingly simple question has far-reaching theoretical consequences for our understanding of the functional architecture of the visual brain, and in particular for the two-visual systems hypothesis (TVSH; Milner & Goodale, 1995, 2006).

We preregistered our design before collecting any data, painstakingly trying to avoid any methodological pitfalls that might compromise the interpretation. Two expert reviewers (at least one of them being a strong advocate of the TVSH) provided detailed input for improving our design and we adapted our study accordingly. Only after the design had been approved did we perform our large study with $N = 144$ participants and collected data in parallel in four different labs, intending to provide the best test to-date of whether or

not grasping is affected by visual illusions, as proposed by the TVSH.

However, Whitwell and Goodale (in this issue) argue that our study was methodologically weak and misguided from the outset because we presented only one Ebbinghaus display at a time, while the predictions of the TVSH could only be tested when simultaneously presenting a pair of two Ebbinghaus displays. In consequence, they think we missed our target and failed to contribute anything new. Here, we argue that this is far too grim a view. The methodological critique offered by Whitwell and Goodale (in this issue) is not justified, and the claim that nothing new has been contributed ignores that a de-facto consensus has been reached on a number of facts, as indirectly also acknowledged by Whitwell and Goodale (in this issue). These facts will in the future facilitate the scientific debate by narrowing down the contentious issues in need of clarification. We will first describe this de-facto consensus before we turn our attention to Whitwell and Goodale's (in this issue) main critique.

DOI of original article: <http://dx.doi.org/10.1016/j.cortex.2016.06.019>.

* Corresponding author. Istituto Italiano di Tecnologia (IIT), Center for Neuroscience and Cognitive Systems@UniTn, Corso Bettini, 31, 38068 Rovereto, TN, Italy.

E-mail addresses: Karl.kopiske@iit.it (K.K. Kopiske), nicola.bruno@unipr.it (N. Bruno), c.hesse@abdn.ac.uk (C. Hesse), thomas.schenk@psy.lmu.de (T. Schenk), volker.franz@uni-tuebingen.de (V.H. Franz).

<http://dx.doi.org/10.1016/j.cortex.2016.10.012>

0010-9452/© 2016 Elsevier Ltd. All rights reserved.

2. De-facto consensus: single Ebbinghaus displays affect grasping as well as perception

In contrast to previous papers (e.g., Goodale, 2008, 2011), Whitwell and Goodale (in this issue) no longer question that there is a clear effect of a single Ebbinghaus display on grasping and that this effect is of the same size as the effect on perception. This is substantial progress, such that scientists should be able to close the files on this question.

Whitwell and Goodale (in this issue) also concede that our study rules out obstacle avoidance mechanisms as the reason for the effects of single Ebbinghaus displays on grasping (“We actually have no issue with this aspect of their study”). This too is progress in the scientific debate, and notably so, given that obstacle avoidance has been the most frequent explanation of TVSH-advocates for why the effects of the Ebbinghaus illusion on grasping should not be attributed to the same processes as the effects of the illusion on perception (Goodale, 2008, 2011; Haffenden, Schiff, & Goodale, 2001; Milner & Goodale, 2008). This also has consequences for the interpretation of studies on other illusions. For example, Whitwell, Buckingham, Enns, Chouinard, and Goodale (2016) used this obstacle-avoidance hypothesis as an argument of why unwanted effects of the Ponzo illusion on grasping should be attributed to different processes than the illusion in perception.

3. Theoretical consequences of this consensus

Despite this de-facto consensus, there is disagreement with respect to its theoretical implications. While we have argued that this finding is not consistent with key notions of the TVSH (see: Aglioti, DeSouza, & Goodale, 1995; Milner & Goodale, 2006, p. 242; Goodale & Ganel, 2016), Whitwell and Goodale (in this issue) argue that single Ebbinghaus displays cannot be used at all to test the validity of the TVSH, that therefore our findings are irrelevant for the TVSH, and that the TVSH can only be tested using dual Ebbinghaus displays.

Before discussing Whitwell and Goodale's (in this issue) dual-Ebbinghaus-only-conjecture, let us point out that their argument is inconsistent with earlier papers from the Goodale-group such that it does not strike us as very convincing. TVSH-proponents have themselves used single Ebbinghaus displays (Haffenden et al., 2001) and concluded that their single Ebbinghaus experiments provide “compelling evidence that the size-contrast illusion elicited by the Ebbinghaus display does not affect grasp scaling” (p. 180), a statement echoed by Ganel, Tanzer, and Goodale (2008). Why, if it was a-priori so clear that single Ebbinghaus displays are not appropriate to test the TVSH, were those displays used in those earlier studies with exactly that purpose? This concern has only now been raised by Whitwell and Goodale (in this issue). That is, after our results have clearly shown that there is no dissociation between perception and action with single Ebbinghaus displays.

However, such post-hoc reasoning is scientifically problematic (see, e.g., Kerr, 1998). In fact, precluding post-hoc

reasoning was one of the main reasons to implement pre-registration in Cortex and other journals (Chambers, Dienes, McIntosh, Rotshtein, & Willmes, 2015). Nevertheless, we will consider below the suggested possibility that the TVSH can be meaningfully tested only with dual Ebbinghaus displays, but not with single Ebbinghaus displays.

4. Are dual-Ebbinghaus displays the only valid tests of the TVSH predictions?

Whitwell and Goodale (in this issue) argue that the illusion effects of single Ebbinghaus displays are too small to test the proposed dissociation. However, the size of the illusion effects cannot be the problem because many studies did find effects of single Ebbinghaus displays on grasping as well as on perception. Now, one could argue that the purported differences between illusion effects on grasping and on perception are too small in single Ebbinghaus displays and that those differences only show up reliably in dual Ebbinghaus displays. However, the large sample size and corresponding a-priori power analysis in our registered report (eight times as many participants as in the largest dual display study; Haffenden & Goodale, 1998), as well as using Bayes factors, and a condition with perceptually matched discs designed specifically to be sensitive to small differences, all rule out the size of the effect or of the differences as potential problems.

In consequence, to make the case that our single-Ebbinghaus-display data should be dismissed, Whitwell and Goodale (in this issue) would have to assume that the dissociation between perception and grasping only exists if we use dual Ebbinghaus displays. By this they assume a qualitatively different, new illusion process, which is active only in dual Ebbinghaus displays, and only for this illusion process the purported dissociation between perception and grasping is existent.¹ This would be a completely new assumption, and we are unaware of any evidence that supports it. The assumption would also be inconsistent with the logic of the TVSH: The TVSH assumes that grasping is unaffected by the Ebbinghaus illusion because it is a contextual effect (Milner & Dyde, 2003). Why then should the single Ebbinghaus illusion (which also is a contextual effect) be allowed by the TVSH to affect grasping? Finally, we want to stress that single Ebbinghaus displays have been typically used in classic studies of the perceptual illusion (e.g., in Coren & Enns, 1993; Coren & Girgus, 1972; Girgus, Coren, & Agdern, 1972), so why should they be inappropriate to test for a possible dissociation between perception and grasping?

However, again, it is a logical possibility that for some hitherto unknown reason the dissociation between perception and grasping can only be detected with dual Ebbinghaus displays but not with single Ebbinghaus displays. Therefore,

¹ Note that this process cannot be the superadditivity of the Ebbinghaus illusion (cf. Foster & Franz, 2014; Franz et al., 2000), because superadditivity can be switched on and off in perceptual measures depending on the task demands (cf. Experiment 3 of Franz et al. (2000). If task demands are matched for perceptual measures and grasping there is no difference between illusion effects on perception and grasping; see also our discussion of superadditivity in the next paragraphs.

let us briefly review whether there is empirical evidence for this notion.

As Whitwell and Goodale (in this issue) point out, there are two prominent grasping studies that used dual Ebbinghaus displays: Aglioti et al. (1995) and Haffenden and Goodale (1998). Both have been taken as evidence for a dissociation between grasping and perception. However, in the first study (Aglioti et al., 1995), task demands were not well matched between grasping and perception (Franz, Gegenfurtner, Bühlhoff, & Fahle, 2000; Pavani, Boscagli, Benvenuti, Rabuffetti, & Farnè, 1999): In grasping, participants operated on only one Ebbinghaus display at a time, while in perception they performed a direct comparison between the target discs of the two Ebbinghaus displays, thereby simultaneously operating on both Ebbinghaus displays. This mismatch is known—as also acknowledged by Whitwell and Goodale (in this issue)—to create an increase of the illusion effect of about 50% (Franz et al., 2000, see also Foster & Franz, 2014), which corresponds well to the difference Aglioti et al. (1995) found between perception and grasping. Therefore, Aglioti et al. (1995) cannot be considered strong evidence for the TVSH. This leaves the Haffenden and Goodale (1998) study, which will be discussed in the next section.

5. Is Haffenden and Goodale (1998) the most decisive study?

Whitwell and Goodale (in this issue) suggest that the study by Haffenden and Goodale (1998) is currently the best test of the TVSH. They argue that the problem of mismatched task demands was avoided in that study (despite using a dual illusion display) by using manual size estimation (ME), where participants indicate the size of an object with index finger and thumb. ME is interpreted as a *perceptual* measure in the framework of the TVSH.² Because participants estimated only one of the central discs of the dual Ebbinghaus display at a time (operating on only one disc, just as in grasping), Whitwell and Goodale (in this issue) argue that there was no mismatch of task demands. Furthermore, Whitwell and Goodale (in this issue) present a reanalysis of the data of Haffenden and Goodale (1998), and calculated for the first time the slope-corrected illusion effects for grasping and ME. They demonstrate that even after slope correction, the illusion effects in ME are much bigger than in grasping.

It is commendable that the appropriate quantitative estimates for the illusion effects are now available for Haffenden and Goodale's (1998) study. However, there are problems that

² We will not discuss the question of what exactly ME measures in further detail here. It seems clear, however, that if ME is a perceptual measure, it should yield results consistent with traditional perceptual measures as, e.g., the methods of adjustment or constant stimuli. To our knowledge, the only systematic investigations into this question have been performed by Franz (2003) and Kopiske et al. (2016), who show that ME can respond with quite a different gain (slope) to a variation of physical size than traditional perceptual measures. In these cases, we need to accurately measure and correct for the response-slope, as now seems to be acknowledged by Whitwell and Goodale (in this issue) (but was questioned in earlier publications of this group).

make us reluctant to accept these recalculations as a strong argument for the proposed dissociation between perception and grasping in visual illusions:

Firstly, the study is only one of multiple studies that investigated the predictions of the TVSH for the Ebbinghaus illusion. If the other studies were now essentially to be ignored, this would constitute a strategy that vastly increases the chances of finding support for just about any given hypothesis (see e.g., Wagenmakers, Wetzels, Borsboom, van der Maas, & Kievit, 2012). If, therefore, Haffenden and Goodale's (1998) study should from now on be the most central argument for the proposed dissociation between grasping and perception in the Ebbinghaus illusion, it would need to be replicated and tested. In Kopiske et al. (2016) we did such a replication and test of Haffenden et al. (2001), another study that was considered to be decisive evidence. Haffenden et al.'s (2001) conclusions did not stand the empirical test – as also acknowledged by Whitwell and Goodale (in this issue) (most notably the idea that the effects of a single Ebbinghaus display on grasping are caused by obstacle-avoidance mechanisms independent of perception).

Secondly, a serious problem of the Haffenden and Goodale (1998) study are the discrepant findings obtained for the two perceptual measures Haffenden and Goodale (1998) measured not only ME, but also a standard perceptual size-matching task. In this task, participants directly compared and matched two central discs in the dual Ebbinghaus displays until they perceived these discs to be equal in size. This yielded a perceptual illusion effect of approximately 2.4 mm. In comparison the newly calculated illusion effect in ME is almost twice as big: About 4.7 mm (our Fig. 1 and Figure 2 of Whitwell & Goodale, in this issue).³

This strong inconsistency between the two perceptual measures is even more surprising if we take into account that in ME there is no superadditivity to be expected (as also argued by Whitwell & Goodale, in this issue). This is so, because participants operated on only one of the two Ebbinghaus displays at a time (just as in grasping). In the standard perceptual size-matching task, on the other hand, the illusion effect should be increased by approximately 50% due to the superadditivity induced by the direct comparison of the two illusory displays (as also acknowledged by Whitwell & Goodale, in this issue). If we take into account this mismatch in task demands, we obtain an illusion effect of approximately 1.6 mm for standard perception ($2.4 \cdot 100/150 = 1.6$) as the most appropriate value to be compared to the illusion effect in ME (cf. Fig. 1). This demonstrates that the two measures of perception in Haffenden and Goodale's (1998) study are dramatically different. In contrast, studies that systematically compared ME to standard perceptual measures (Franz, 2003; Kopiske et al., 2016) obtained similar illusion effects for both

³ The slopes of standard perception were not measured in Haffenden and Goodale (1998), therefore it is not possible to slope-correct those standard-perception illusion effects. However, we know that the slope of standard perception is typically close to 1; therefore we can use the uncorrected data as a fairly good approximation and can compare this approximation to the slope-corrected illusion effects of ME.

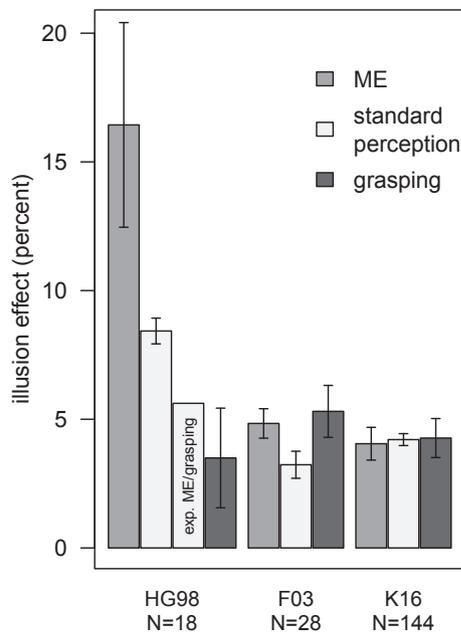


Fig. 1 – Illusion effects in studies comparing grasping (MGA) to manual estimation (ME) as well as a standard perceptual measure. Illusion effects are in percent relative to the physical size of the stimuli; all illusion effects are slope-corrected (Bruno & Franz, 2009; Franz, 2007; Franz, Scharnowski, & Gegenfurtner, 2005; Hesse, Franz, & Schenk, 2016; Kopiske et al., 2016, p. 139); except for standard perception of HG98 (see footnote 3). All studies used roughly similar Ebbinghaus displays (HG98: SN/LF, F03: SN/LN, Kopiske et al., 2016: SN/LF, see Kopiske et al., 2016 for nomenclature). Aggregated data for HG98 were kindly provided by M. Goodale and R. Whitwell (personal communication, July, 29th and Aug, 19th, 2016). Error bars indicate the SEM of the corrected illusion effect, estimated using a Taylor-approximation (cf. Kopiske et al., 2016, p. 139 and Hesse et al., 2016, p. 94 for an equivalent but simplified formula). Note, that Whitwell and Goodale (in this issue) used in their Figure 2 the problematic ‘zero-variance method’, that in general underestimates the size of the SEMs. Although for Whitwell and Goodale (in this issue) this effect is not dramatic, we show here the more appropriate Taylor-approximated SEMs (cf. Franz, 2007; Franz et al., 2005 for a discussion of the zero-variance method).

measures, as long as the slope correction was performed and task demands were matched (cf. Fig. 1).

So what do Haffenden and Goodale's (1998) unusual perceptual illusion effects mean for the comparison to grasping? It is clear that no matter whether we take into account superadditivity or not, the difference between the illusion effects in grasping and standard perception is much smaller than that between standard perception and ME (Fig. 1). Therefore, even this data provides no evidence for a “perceptual cluster” (guided by the ventral stream) versus a “motoric cluster” (guided by the dorsal stream). If anything, Haffenden and Goodale's (1998) data (but not those of the other studies) suggests that grasping and standard perception

are similar but different from ME. Thus, before drawing far-reaching conclusions from this data it will be necessary to clarify why the ME data of this study is so unusual and unexpected – even from the viewpoint of the TVSH.

6. Did we ignore the perceptually-matched condition?

Before closing, we want to discuss a more specific issue: Whitwell and Goodale (in this issue) argue that, historically, we simply ignored the perceptually-matched condition of Aglioti et al. (1995) and Haffenden and Goodale (1998), thereby ignoring a substantial part of the data of those studies. In consequence, it would be no surprise if we came to wrong and biased conclusions. This argument has been brought up repeatedly before and has been responded to (e.g., Franz & Gegenfurtner, 2008). It also seems ironic that it is now raised against Kopiske et al. (2016), a study in which we took great care to laboriously implement such a perceptually-matched condition.

Before describing this condition in Kopiske et al. (2016), let us first comment on the perceptually-matched condition in general: The perceptually-matched condition is a nulling-procedure: A pair of discs is selected that appears perceptually equal in size if one of the disc is surrounded by the enlarging context of the illusion and the other by the shrinking context. If the condition works as intended and perception is equalized, then we can attribute differences in grasping the discs to a different size of the illusion effect between grasping and perception.

However, the perceptually-matched condition has a big disadvantage: Because physical size and illusion are confounded, it is not easy to quantify the size of the illusion effect in grasping. This is a problem if we want to quantitatively compare illusion effects between perception and grasping. Such a quantitative comparison is necessary because studies typically did find at least some illusion effects on grasping (even Aglioti et al., 1995), thereby ruling out ‘strong’ versions of the TVSH that would state complete immunity of grasping to those illusions (as opposed to just a smaller illusion effect in grasping than in perception, as ‘weaker’ versions of the TVSH would state).⁴ Thus, all studies (including Aglioti et al., 1995 and the recalculations in Whitwell & Goodale, in this issue) used the physically-matched conditions to quantify the illusion effect, such that quantitative estimates of the illusion effect are only available for this condition. Note, however, that this is not very critical because there is no reason to assume the illusion effect to be drastically different between

⁴ Another reason for a quantitative comparison is that the task demands in Aglioti et al. (1995) were such that we expect a-priori a larger illusion effect in the perceptual measure than in grasping (because only the perceptual measure employed a direct comparison; due to the superadditivity of the Ebbinghaus illusion, this increases the illusion effect by approximately 50%, cf. Experiment 3 of Franz et al., 2000). Therefore, only a quantitative comparison allows assessing whether the larger illusion effect in the perceptual measure can be explained by this mismatch in task demands (which would be no evidence for the TVSH), or whether it is truly larger (which would be evidence for the TVSH).

perceptually-matched and physically-matched conditions. This is so because (a) the conditions are very similar (the only difference is that one disc has a slightly different size in the perceptually-matched condition to achieve perceptual equivalence), and (b) we explicitly tested for such a difference between perceptually-matched and physically-matched conditions in Kopiske et al. (2016) and found no differences (Figure 8 of Kopiske et al., 2016).

Finally, let us comment on the perceptually-matched condition of our study: We included this condition for many methodological reasons (as detailed in Kopiske et al., 2016) and as suggested by one reviewer. This condition was performed in a much more controlled way than in Aglioti et al. (1995) and in Haffenden and Goodale (1998): (a) The earlier studies selected the pair of matched discs in a pilot phase by the experimenter using trial and error, while in Kopiske et al. (2016) we used a psychophysical constant stimuli method. (b) Previous studies did not quantitatively test whether the matching actually worked or whether there was a residual mismatch of the pair of discs. We tested this laboriously in a second condition. (c) In those earlier studies, participants could only choose between discs that varied in 1 mm steps. This is much too coarse for an illusion effect of, on average, only 2.4 mm (Haffenden & Goodale, 1998). We used step sizes of .25 mm (which is still not perfect, but much better). Given all these advantages, it is quite surprising that Whitwell and Goodale (in this issue) seem to dismiss the relevance of our perceptually-matched condition.

7. Summary and conclusions

Whitwell and Goodale (in this issue) concede that single Ebbinghaus displays seem to affect grasping to a similar degree as perception and that these effects cannot be attributed to non-perceptual, purely motor processes (obstacle avoidance, awkward grasping). However, they argue that a test of the TVSH can only and exclusively be performed using dual Ebbinghaus displays but not with single Ebbinghaus displays. They therefore suggest that Haffenden and Goodale (1998) is the decisive study to test for a dissociation between grasping and perception. However, as we discussed here, this study has serious problems, because the perceptual measures yielded highly inconsistent illusion effects. Future research should first focus on finding consistent perceptual illusion effects in the Haffenden and Goodale (1998) paradigm before these can be meaningfully compared to grasping data.

In contrast, the extensive tests in Kopiske et al. (2016) have demonstrated consistent illusion effects across a wide variety of perceptual measures and also between perception and grasping. The design of Kopiske et al. (2016) was the result of intensive efforts of four independent research groups and end-detail critique by two anonymous expert reviewers. Here we have outlined why we think that Whitwell and Goodale's (in this issue) methodological critique is post-hoc and not convincing, and why we believe that Kopiske et al. (2016) provides a strong and valid test of the claim that certain illusions affect perception more than grasping. The outcome of this test suggests that there is no difference in the effects of the Ebbinghaus illusion on grasping and perception.

REFERENCES

- Aglioti, S., DeSouza, J. F. X., & Goodale, M. A. (1995). Size-contrast illusions deceive the eye but not the hand. *Current Biology*, 5(6), 679–685.
- Bruno, N., & Franz, V. H. (2009). When is grasping affected by the Müller-Lyer illusion? A quantitative review. *Neuropsychologia*, 47, 1421–1433.
- Chambers, C. D., Dienes, Z., McIntosh, R. D., Rotshtein, P., & Willmes, K. (2015). Registered reports: Realigning incentives in scientific publishing. *Cortex*, 66, A1–A2.
- Coren, S., & Enns, J. T. (1993). Size contrast as a function of conceptual similarity between test and inducers. *Perception & Psychophysics*, 54(5), 579–588.
- Coren, S., & Girgus, J. S. (1972). A comparison of five methods of illusion measurement. *Behavior Research Methods & Instrumentation*, 4(5), 240–244. <http://dx.doi.org/10.3758/BF03210006>.
- Foster, R. M., & Franz, V. H. (2014). Superadditivity of the Ebbinghaus and Müller-Lyer illusions depends on the method of comparison used. *Perception*, 43(8), 783–795. <http://dx.doi.org/10.1068/p7802>.
- Franz, V. H. (2003). Manual size estimation: A neuropsychological measure of perception? *Experimental Brain Research*, 151, 471–477.
- Franz, V. H. (2007). Ratios: A short guide to confidence limits and proper use. Retrieved from <http://arxiv.org/pdf/0710.2024v1.pdf>.
- Franz, V. H., & Gegenfurtner, K. R. (2008). Grasping visual illusions: Consistent data and no dissociation. *Cognitive Neuropsychology*, 25, 920–950.
- Franz, V. H., Gegenfurtner, K. R., Bühlhoff, H. H., & Fahle, M. (2000). Grasping visual illusions: No evidence for a dissociation between perception and action. *Psychological Science*, 11(1), 20–25.
- Franz, V. H., Scharnowski, F., & Gegenfurtner, K. R. (2005). Illusion effects on grasping are temporally constant not dynamic. *Journal of Experimental Psychology: Human Perception and Performance*, 31(6), 1359–1378.
- Ganel, T., Tanzer, M., & Goodale, M. A. (2008). A double dissociation between action and perception in the context of visual illusions: Opposite effects of real and illusory size. *Psychological Science*, 19(3), 221–225.
- Girgus, J. S., Coren, S., & Agdern, M. (1972). The interrelationship between the Ebbinghaus and Delboeuf illusions. *Journal of Experimental Psychology*, 95(2), 453–455.
- Goodale, M. A. (2008). Action without perception in human vision. *Cognitive Neuropsychology*, 25(7–8), 891–919.
- Goodale, M. A. (2011). Transforming vision into action. *Vision Research*, 51(13), 1567–1587.
- Goodale, M. A., & Ganel, T. (2016). Different modes of visual organization for perception and action. In J. Wagemans (Ed.), *Oxford handbook of perceptual organization* (pp. 672–690). Oxford: Oxford University Press.
- Haffenden, A. M., & Goodale, M. A. (1998). The effect of pictorial illusion on prehension and perception. *Journal of Cognitive Neuroscience*, 10(1), 122–136.
- Haffenden, A. M., Schiff, K. C., & Goodale, M. A. (2001). The dissociation between perception and action in the Ebbinghaus illusion: Nonillusory effects of pictorial cues on grasp. *Current Biology*, 11, 177–181.
- Hesse, C., Franz, V. H., & Schenk, T. (2016). Pointing and anti-pointing in Müller-Lyer figures: Why illusion effects need to be scaled. *Journal of Experimental Psychology: Human Perception and Performance*, 42(1), 90–102.
- Kerr, N. L. (1998). HARKing: Hypothesizing after results are known. *Personality and Social Psychology Review*, 2(3), 196–217.

- Kopiske, K. K., Bruno, N., Hesse, C., Schenk, T., & Franz, V. H. (2016). The functional subdivision of the visual brain: Is there a real illusion effect on action? A multi-lab replication study. *Cortex*, 79, 130–152. <http://dx.doi.org/10.1016/j.cortex.2016.03.020>.
- Milner, A. D., & Dyde, R. (2003). Why do some perceptual illusions affect visually guided action, when others don't? *Trends in Cognitive Sciences*, 7(1), 10–11.
- Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action* (1st ed.). Oxford: Oxford University Press Inc.
- Milner, A. D., & Goodale, M. A. (2006). *The visual brain in action* (2nd ed.). Oxford: Oxford University Press Inc.
- Milner, A. D., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia*, 46, 774–785.
- Pavani, F., Boscagli, I., Benvenuti, F., Rabuffetti, M., & Farnè, A. (1999). Are perception and illusion affected differently by the Titchener circles illusion? *Experimental Brain Research*, 127, 95–101.
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., van der Maas, H. L. J., & Kievit, R. A. (2012). An agenda for purely confirmatory research. *Perspectives on Psychological Science*, 7, 627–633.
- Whitwell, R. L., Buckingham, G., Enns, J. T., Chouinard, P. A., & Goodale, M. A. (2016). Rapid decrement in the effects of the Ponzo display dissociates action and perception. *Psychonomic Bulletin & Review*, 23(4), 1157–1163. <http://dx.doi.org/10.3758/s13423-015-0975-4>.

Received 26 August 2016

Reviewed 23 September 2016

Revised 13 October 2016

Accepted 14 October 2016

RESEARCH ARTICLE

The SNARC Effect in Chinese Numerals: Do Visual Properties of Characters and Hand Signs Influence Number Processing?

Karl K. Kopiske^{1,2*}, Christian Löwenkamp², Owino Eloka², Florian Schiller³, Chung-Shan Kao², Chaohua Wu⁴, Xiaorong Gao⁴, Volker H. Franz^{2,5}

1 Istituto Italiano di Tecnologia (IIT), Center for Neuroscience and Cognitive Systems@UniTn, Corso Bettini, 31, 38068 Rovereto, TN, Italy, **2** University of Hamburg, Institute of Psychology, Von-Melle-Park 5, 20146 Hamburg, Germany, **3** Justus Liebig University Giessen, Department of General Psychology, Otto-Behaghel-Straße 10, 35394 Giessen, Germany, **4** Tsinghua University, Department of Biomedical Engineering, Beijing 100084, PR China, **5** University of Tübingen, Institute of Computer Science, Department of Experimental Cognitive Science, Sand 13, 72076 Tübingen, Germany

* karl.kopiske@iit.it



OPEN ACCESS

Citation: Kopiske KK, Löwenkamp C, Eloka O, Schiller F, Kao C-S, Wu C, et al. (2016) The SNARC Effect in Chinese Numerals: Do Visual Properties of Characters and Hand Signs Influence Number Processing? PLoS ONE 11(9): e0163897. doi:10.1371/journal.pone.0163897

Editor: Bert De Smedt, Katholieke Universiteit Leuven, BELGIUM

Received: January 11, 2016

Accepted: September 18, 2016

Published: September 29, 2016

Copyright: © 2016 Kopiske et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available via the Open Science Framework at <https://osf.io/pjxgu/>.

Funding: Karl K. Kopiske, Christian Löwenkamp and Chaohua Wu were supported by Ph.D. scholarships as part of grants to Volker H. Franz and Xiaorong Gao within the International Graduate Research Training Group (No. IKG-1247) on Cross-Modal Interaction in Natural and Artificial Cognitive Systems (CINACS – <http://www.cinacs.org>) from the Deutsche Forschungsgemeinschaft and the Chinese Ministry of Education. The funders

Abstract

The SNARC effect refers to an association of numbers and spatial properties of responses that is commonly thought to be amodal and independent of stimulus notation. We tested for a horizontal SNARC effect using Arabic digits, simple-form Chinese characters and Chinese hand signs in participants from Mainland China. We found a horizontal SNARC effect in all notations. This is the first time that a horizontal SNARC effect has been demonstrated in Chinese characters and Chinese hand signs. We tested for the SNARC effect in two experiments (parity judgement and magnitude judgement). The parity judgement task yielded clear, consistent SNARC effects in all notations, whereas results were more mixed in magnitude judgement. Both Chinese characters and Chinese hand signs are represented non-symbolically for low numbers and symbolically for higher numbers, allowing us to contrast within the same notation the effects of heavily learned non-symbolic vs. symbolic representation on the processing of numbers. In addition to finding a horizontal SNARC effect, we also found a robust numerical distance effect in all notations. This is particularly interesting as it persisted when participants reported using purely visual features to solve the task, thereby suggesting that numbers were processed semantically even when the task could be solved without the semantic information.

1 Introduction

1.1 The SNARC effect: Description and models

The SNARC (Spatial Numerical Association of Response Codes) effect refers to a proposed association between numbers and space. This was originally put forward by Dehaene, Bossini and Giraux [1], who found that European participants responded quicker to relatively smaller numbers with a response button on the left than on the right, and vice-versa for relatively large

had no role in study design, data collection or analyses, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

numbers, despite number magnitude being irrelevant for the task. This association is thought to be automatic [2]. The SNARC effect is found in a number of tasks (e.g., judgement of parity [1] or magnitude [3], but even unrelated tasks like finger tapping [4] have been shown to elicit SNARC-like effects). The most common task is a parity judgement task with horizontally separated response buttons: Participants are asked if a number is odd or even, and respond by pressing one of two buttons either on the left or on the right. The mapping of odd and even to buttons on the left and on the right will change during the experiment, so that for each number, left and right responses are given. The variable of interest then is the difference between right-handed and left-handed response times for each number. Statistically speaking, this is sometimes referred to as difference of response times, or dRT [5], where a systematic linear relationship between number magnitude and dRT indicates a SNARC effect.

Several models have been put forward to explain the effect. The most prominent explanation has been the following: during number processing, an internal mental number line is automatically activated. In Western participants, number magnitude will ascend from left to right on this line so that small numbers will be on the left and large numbers on the right [6–9]. The mental number line is mapped into external space; the better mapping location and response location correspond, the faster the response, giving rise to the SNARC effect. In Western participants, for example, small numbers will be mapped to the left of the number line and larger numbers to the right of the number line. This has also been called the direct mapping account [9], since it suggests a direct relationship between numbers and space. However, competing accounts exist. For example, the dual-route explanation [10,11] proposes that the activation of the mental number line is not necessary in numerical cognition and that humans may manipulate numbers without accessing their semantic meaning, purely through automatic activation of a response associated with the word or symbol. According to the dual-route explanation, the SNARC effect can also arise, but only when it is necessary to extract the meaning of a number word, which would then activate the number line. A more general approach than that of a mental number line was put forward by Proctor and Cho [12,13]. They assume that each number and response to some extent has a positive or negative polarity, and that the SNARC is a congruency effect of polarities. For a summary of the ongoing debate about which model best describes the SNARC effect and related findings, see van Dijk, Gevers, Lafosse and Fias [14].

1.2 To what degree does the SNARC effect depend on notation?

A large number of notations and modalities have been found to elicit a SNARC effect or SNARC-like effects. These include, besides Arabic digits, visual and auditory number words, dice patterns [15], letters of the alphabet, months [16], learned magnitude relations [17], German Sign Language hand signs [18], as well as Japanese [19] and Chinese characters, although the strength and direction of a SNARC effect may differ [20]. Thus, it is often claimed that the SNARC effect is amodal and independent of notation [15,21]. However, studies have reported results that the direction of number-space mapping depends on interpretation of the numbers [22], typical context of a symbol [20] or even single number words in a language written in a different direction presented in a previous trial [23], which has prompted others to say that the effect may be more dependent on the particular stimuli and experimental design used. For reviews on this issue, see Gevers and Lammertyn [24] or Wood, Willmes, Nuerk and Fischer [25].

Indeed, where a SNARC effect can and cannot be found has informed several hypotheses about underlying mechanisms of the effect. One example is the common notion that the effect is strongly influenced by reading habit (see e.g. [24,26]). This was already hypothesised in the original paper [1], which reported a strong left to right SNARC effect in French participants,

but a significantly weaker effect in Iranians (whose native language is written right-to-left) living in France, with the strength of the SNARC effect possibly related to the length of time since moving to France. This influence is thought to hold for both reading habit of numbers and of words, as Shaki, Fischer and Petrusic [27] found by testing Israeli (reading habit for words: right to left, numbers: left to right; no clear SNARC effect) and Palestinian participants (words and numbers: right to left; clear right to left SNARC effect). Similar findings have supported this idea, as a vertical SNARC effect (speed-advantage of top-small and large-bottom responses over top-large and small-bottom) has been demonstrated in Taiwanese participants [20], who may read a significant portion of text vertically. A vertical SNARC effect was also demonstrated in Japanese participants [19], although the direction was reversed here, with smaller numbers being responded to faster when the response was on the lower of two vertically separated response buttons. This is interesting since reading habit would be the same in the samples studied by Hung et al. [20] and Ito and Hatta [19], and indeed Ito and Hatta do not cite reading habit as an explanation for their vertical effect at all, but rather explain it in terms of a separate, general association of magnitude and space [19].

This idea of a general magnitude-space association is supported by findings that a vertical SNARC effect can also be found in European and American participants [28–30], although it appears to be less automatically activated than the horizontal SNARC in tasks where the magnitude of the stimulus or the vertical dimension of the response is irrelevant [28,29]. Such associations in a direction that does not correspond with typical reading habit are central evidence for the notion of a general magnitude-space association, although they have also been taken as evidence that the effect may be more dependent on short-term influences as opposed to long-term habits than previously assumed [23].

These associations are also very relevant to the question of which mechanisms might be behind a general association of space and magnitude. In addition to the polarity account of Proctor and Cho [12], it has also been related to grounded cognition by Fischer and Brugger [31], who proposed three levels of stimulus-dependent space-magnitude associations: (a) Grounded representations, in which basic properties of the world determine the associations (e.g., the fact that stacking objects creates increasingly higher piles would contribute to a large-upwards association) (b) embodied cognition, where representations are influenced by associations with body parts, such as hands or fingers (c) situated cognition, with the representation being dependent on the context. Evidence for this view has come for example from Bächtold et al. [22], where conceptualising numbers as a ruler or a clock produced opposite results. Additionally, finger counting habit has also been found to affect the SNARC effect (e.g. [32]; see [33] for a review and [34] for an imaging study reporting consistent modulations of neural activation), in that the strength of the SNARC effect differed between participants that start counting on their left hand and those that start on their right hand. Within such a framework [31], studying effects of numerical cognition and specifically the SNARC effect in hand signs is especially interesting, as the notion of embodied cognition would predict properties of these hand signs to directly influence the spatial representation of magnitude presented through hand signs.

1.3 Our study: Employing Chinese characters to investigate notation dependence

Our study tested whether or not a horizontal SNARC effect can be found for *Arabic digits*, simple-form *Chinese characters*, and *Chinese hand signs* in participants living and raised in Mainland China. The *Arabic digits* notation was included to measure the well-known horizontal SNARC effect as a baseline to compare the other notations to, while each of the other notations allowed us to test specific predictions.

It has been proposed that generally, number processing in Chinese speakers may differ fundamentally from number processing in Western participants [35]. It has also been suggested that the semantic processing of Chinese characters differs from that of *Arabic digits* [35,36], specifically with regard to its temporal properties. More specifically, while Mainland Chinese and Taiwanese participants display a horizontal SNARC effect in Arabic digits [20,37], Taiwanese participants may not display a clear horizontal SNARC effect for *Chinese characters* [20]. The main difference between participants from Taiwan and Mainland China is that whereas horizontal writing was formally introduced in mainland China in 1955 and used henceforth, in Taiwan a similar guideline was introduced for official documents in 2004 (see [38], accessed via [39]), and there remain texts (including e.g. books, textbooks) that are written vertically. Thus, the exposure to vertical text would be greatly different, despite the characters being the same. Finding a clear horizontal SNARC effect for *Chinese characters* in our Mainland Chinese participants would emphasise the importance of reading and writing experience and provide more evidence for it being independent of notation, while the opposite finding would point towards there being an effect of notation.

Chinese hand signs were included to test for the effect in a neither purely symbolic, nor purely non-symbolic notation in which the effect had never been tested before. It has been shown that notations defined by numerosity (e.g. dice patterns, [15]) can evoke a SNARC effect, as can hand signs [18,40]. However, in Western sign languages, numbers tend to be represented by the same number of fingers, making it a non-symbolic representation based on finger numerosity. Chinese hand signs, on the other hand, are represented through finger numerosity for numbers 1..5, and purely symbolically for numbers 6..10 (displayed through different signs using one, two, or three fingers—see Fig 1), which means that one cannot assume that the same effect will necessarily be present. Finding a SNARC effect in this notation, and especially in the higher (symbolically represented) range, would indicate that indeed the magnitude displayed in hand signs can elicit the effect independently of the number of fingers seen. Such a finding would also make this notation potentially a beneficial, confound-free embodied notation to test predictions of embodied cognition with regards to participant counting preferences or body postures. On the other hand, a difference between symbolically and non-symbolically represented numbers would reinforce the notion that hand signs where the number of fingers represents magnitude are susceptible to a confound of magnitude and numerosity.



Fig 1. The stimuli used in our experiments. (a) Arabic digits, (b) simple-form Chinese characters, (c) Chinese hand signs as used in Chinese Sign Language. Stimuli in each column represent identical numbers. Note that the number 5 is omitted in all notations. This enabled us to use it as the standard for the magnitude judgement task. Hand signs images retrieved from https://en.wikipedia.org/wiki/Chinese_number_gestures, created by Wikipedia user Ningling, and used under the terms of the GNU Free Documentation License.

doi:10.1371/journal.pone.0163897.g001

1.4 Our study: Methodology and research questions

We tested the effect in the two most common tasks in the SNARC literature: Parity judgement (i.e., judging if a number is odd or even) and magnitude judgement (i.e., judging if a number is smaller or larger than a standard, in this case 5). Beside the obvious difference between the tasks, they may also differ with regards to the information that is activated, as previous studies have shown that secondary tasks that tax working memory suppress the SNARC effect in *parity* judgement when *verbal* resources are required for the secondary task, whereas the SNARC effect in *magnitude* judgement is suppressed by *visuospatial* secondary tasks [3,41]. If parity judgement and magnitude judgement depend primarily on verbal and visuospatial information, respectively, it would be plausible for magnitude judgement to be more affected by the differences in visual complexity between the notations. We also tested for a numeric distance effect in the magnitude judgement task (i.e., faster response times when the numerical distance between the stimulus and the standard is larger), a common marker of semantic processing [7,42], to test whether the magnitude judgement task was indeed executed based not on purely visual features but based on magnitude information, which would be a prerequisite to interpret the results from it as indicative of a space-number association.

In summary, we tested for the existence of a horizontal SNARC effect in native Chinese participants living in Beijing using three different notations and two tasks. These notations and tasks allowed us to (a) investigate the processing of a mixed symbolic and non-symbolic (numerosity-based), embodied notation (b) separate reading habit and notation for Chinese characters by comparing our results to those of a study conducted with Taiwanese participants, and (c) test if our results would persist under different task demands inherent to parity judgement and magnitude judgement.

2 Experiment 1: Parity judgement

A classic parity judgement paradigm was used in experiment 1, in the three different notations *Arabic numbers*, *Chinese characters*, and *Chinese hand signs*. The goal was to test for the existence of a horizontal SNARC effect in native Chinese speakers who grew up in Mainland China, using the most common task for SNARC experiments.

2.1 Participants

Twenty-six native Chinese speakers (all at least 18 years old, age: $M = 22.5$, $SD = 2.0$), recruited September 2011, participated voluntarily for an agreed pay of 50 RMB. All were right-handed by self-report, 12 were female. All had normal or corrected-to-normal vision. All were students of Tsinghua University, Beijing, and naive to the purpose of the task.

All participants were adults, 18 years or older, recruited from the Tsinghua University Biomedical Engineering department. Participants gave written, informed consent. They were compensated by a previously agreed amount (see specifics for each experiment in the manuscript) that reflected the standard payment in the department. Participants signed their name on the consent sheet, but no identifying information was contained in the experimental data itself, and the consent sheet could not be linked to any data. On the consent sheet, participants also confirmed that they had normal or corrected-to-normal vision and hearing, as well as no attention disorders. Participants that did not meet all of these criteria were not tested, and no data or records of any kind were recorded of them. The information was also not linked to any information in the data, and no further medical information was collected. Consent sheets are being stored in a locked cabinet at the University of Hamburg Psychology department that is only accessible to members of the department.

Both experiments were conducted in accordance with to the 1964 Declaration of Helsinki, and following ethical guidelines of the German Psychological Society (DGPs) and the Professional Association of German Psychologists (BDP) (2005, C.III). The study was conducted within the International Graduate Research Group "Cross-modal Interaction in Natural and Artificial Cognitive Systems" (CINACS) that was reviewed and approved by the German Research Foundation (DFG, project number IGK-1247) which did not require further Institutional Review Board approval. The reviewed description of this research group included response time tasks like the ones conducted for this study.

Our experiments lasted at most 65 minutes, during which participants were allowed to take as many breaks as they wished. They consisted of standard two-alternative-forced-choice response time tasks that required quick button presses in response to numbers displayed on a standard computer screen. For these experiments, no particular risk of harm or stress was apparent to us, other than the possibility that participants may find the monotonous task somewhat tedious. Thus, we did not seek further ethical approval for this particular study. We retroactively asked the Local Ethics Committee of the Faculty for Psychology and Human Movement Science, University of Hamburg to assess whether ethical review would have been necessary. The committee came to the conclusion that this was not the case.

2.2 Stimuli and apparatus

Participants were seated approximately 60 cm in front of a 19" LCD monitor (effective screen diagonal: 48 cm) using a resolution of 1024 * 768 pixels with a refresh rate of 60 Hz and gave responses via a standard QWERTY USB-keyboard. They were presented with the numbers 1 . . 4, 6 . . 9 in three different notations: *Arabic numbers*, *Chinese characters*, and *Chinese number gestures* as used in Chinese Sign Language and displayed with the left hand, see Fig 1. Numbers were presented as 225 pixel * 225 pixel tagged image file (tif) images, centrally displaying Chinese and Arabic characters of font size 60 and images of hands of approximately equal size (app. 24 mm * 24 mm, corresponding to about 2.3 degrees of visual angle). The stimuli were presented in a custom MATLAB program using Psychophysics Toolbox 3 [43].

2.3 Procedure

The experiment was segmented into 6 blocks: Two blocks for each parity mapping, that is, *left = even; right = odd* and *left = odd; right = even*, for each of the three notations *Arabic digits*, *Chinese characters*, and *Chinese hand signs*. Each block consisted of practice trials until 8 correct responses were given, followed by 288 experimental trials (8 numbers * 36 repetitions). This resulted in at least 1776 trials total, and 1728 experimental trials. The order of blocks was counterbalanced between participants, with blocks of the same notation presented consecutively and each participant being assigned to one of 12 groups (6 sequences of blocks * 2 sequences of mapping). The numbers within each block were randomised. Before each block, participants were instructed which of the buttons "s" and "l" on the keyboard was to be pressed for even numbers and which for odd numbers. In each trial, a fixation cross (font size 40) appeared for a mean duration of 500 ms (400 ms + random value from an exponential distribution with mean = 100 ms), followed by the stimulus presented until a response button was pressed, but at most 2000 ms. This was followed by a pause of 250 ms. After every error, the participants saw the word "wrong!" written in red, font size 40, centrally on the screen for 250 ms. In total, the experiment lasted between 50 and 65 minutes.

2.4 Data analysis

A total of four participants had to be excluded from data-analysis: Two because of the number of errors made (more than 2 SD above the mean), one for being unfamiliar with some stimuli, and one because of technical difficulties. This left us with 22 participants for analysis. Response times were trimmed with 200 ms as the lower cut-off and each participant’s median RT for each notation plus three standard deviations as upper cut-offs, respectively, which eliminated 1.7% of trials from analysis.

We ran two main types of analysis: First, we ran a 3 (notation) * 2 (side of response) * 4 (numerical magnitude) * 6 (order) ANOVA over RTs. The factor *order* was a between-subject factor that coded the order of blocks. Numbers were assigned to four magnitude bins (1 and 2, 3 and 4, 6 and 7, 8 and 9) to control for confounding effects like the MARC effect (markedness association of response codes [44]) and to keep the analyses in line with recently proposed methodology [45,46]. Note that markedness, the property of being marked as unique, or uncommon [47], is a linguistic concept in which the English words “odd” and “even” differ, but not the Chinese equivalents. Thus, we did not expect the same mechanism to have an effect here. However, a similar effect would have been plausible: In Chinese, 奇偶, literally “odd even”, means parity. Hence, we tested for an advantage of “left-odd” and “right-even” responses. Greenhouse-Geisser-corrected p-values [48] along with Greenhouse-Geisser epsilon (ϵ_{GG}) are given for factors with more than two levels. Second, we computed response time differences between right-handed responses and left-handed responses (dRTs) for each participant, number and notation, which we then used to compute linear regression slopes of dRT over number and magnitude bin (in *ms per digit* or *ms per bin*, respectively, see Table 1). For each notation, these slopes were then submitted to t-tests against 0 to clarify if the SNARC effect persisted in each notation. Third, we submitted these dRT slopes to a Bayesian analysis that can have some advantages over frequentist statistics (e.g., Dienes [49,50]) and allowed us to disambiguate whether non-significant results when testing for a SNARC effect, or for a difference between SNARC effects, should be interpreted as evidence for the absence of an effect or as a consequence of inconclusive data. We followed the guidelines proposed by Jeffreys [51] for the interpretation of Bayes factors (BFs). In short, throughout this paper we refer to BFs below 1/100 as decisive evidence for the H0, BFs below 1/10 as strong evidence for the H0, and BFs below 1/3 as substantial evidence for the H0. BFs between 1/3 and 3 indicate that the data are not sensitive enough to draw strong conclusions. BFs above 3, above 10, and above 100 represent substantial, strong, and decisive evidence in favour of the H1, respectively.

2.5 Results

There was a significant main effect on RTs for *numerical magnitude* ($F(3, 48) = 24.281, p_{GG} < .001, \epsilon_{GG} = .712$) and *notation* ($F(2, 32) = 72.226, p_{GG} < .001, \epsilon_{GG} = .803$). There was no

Table 1. Regression coefficients of dRT over number and dRT over magnitude bins.

Notation	Regression dRT by number	R	Regression dRT by magnitude bins	R
Digits exp. 1	$y = 26.4 - 5.1x \pm 0.9$.101	$y = 31.3 - 12.2x \pm 2.3$.096
Digits exp. 2	$y = 28.9 - 7.5x \pm 1.6$.246	$y = 36.6 - 18.1x \pm 3.9$.239
Characters exp. 1	$y = 16.1 - 3.7x \pm 1.1$.068	$y = 20.2 - 9.0x \pm 2.6$.067
Characters exp. 2	$y = 12.8 - 4.5x \pm 2.3$.080	$y = 18.0 - 11.0x \pm 5.4$.080
Hand signs exp. 1	$y = 14.2 - 3.2x \pm 1.1$.039	$y = 15.7 - 7.0x \pm 2.5$.032
Hand signs exp. 2	$y = 7.8 - 2.6x \pm 2.8$.021	$y = 11.6 - 6.8x \pm 6.8$.023

Note: Mean coefficients of the linear regression are given as y in ms, x in ms/digit and ms/bin, respectively, \pm SEM. R indicates explained variance. Experiment 1: Parity judgement, experiment 2: Magnitude judgement.

doi:10.1371/journal.pone.0163897.t001

main effect of *side* ($F(1, 16) = .496, p = .491$), but a significant interaction *side * numerical magnitude* ($F(3, 48) = 14.076, p_{GG} < .001, \epsilon_{GG} = .598$), indicating a possible SNARC effect. We also observed a significant *notation * numerical magnitude* interaction ($F(6, 96) = 11.556, p_{GG} < .001, \epsilon_{GG} = .564$), as larger numbers were processed slower in all notations, but especially so in *Chinese hand signs* and *Chinese characters*, which may reflect the fact that visual complexity increases in these notations. Crucially, the three-way *notation * numerical magnitude * side* interaction that would have indicated a different SNARC effect depending on the notation was not significant ($F(6, 126) = 1.427, p_{GG} = .232, \epsilon_{GG} = .716$), and no main effect ($F(5, 16) = 0.515, p = .761$) of factor *order* was significant, with the only significant interaction being *order * side* ($F(5, 16) = 6.498, p = .002$; all other interactions n.s., $p > .13$), indicating that participants were faster in the mapping they had learned first. Left-handed and right-handed RTs from this task are plotted in Fig 2.

We then computed dRTs for each number and used these to run regressions of dRTs by number, see Table 1. The resulting slopes were tested for difference against 0 to investigate if a SNARC effect existed in each notation, as negative slopes would indicate smaller dRTs (i.e., relatively faster right-handed responses) the larger the number. They revealed a SNARC effect for *Arabic digits* ($t(21) = -5.461, p < .001$), *Chinese characters* ($t(21) = -3.395, p = .003$), and for *Chinese hand signs* ($t(21) = -2.879, p = .009$). BFs were calculated following the logic proposed by Dienes [49,50], comparing the likelihood of a point-hypothesis at 0 to that of a uniform distribution ranging from 0 to the dRT slope reported in a meta-analysis by Schiller et al. [54] of -5.74 ms/digit. They confirm that the effect in *Arabic digits* ($BF > 1000$; decisive evidence), as well as both *Chinese characters* ($BF = 147.6$; decisive evidence) and *Chinese hand signs* ($BF = 30.2$ strong evidence) are robust effects. Bayes factors testing for a real difference between dRTs obtained from our experiment by comparing likelihoods of a point-0 H0 to an H1 of $\text{slope}(\text{dRT}_{\text{notation1}}) = \text{slope}(\text{dRT}_{\text{notation2}})$ reported strong or decisive evidence for the H1 in all cases (all BFs > 10), indicating that dRT slopes were comparable and the non-significant three-way interaction indeed pointed towards no difference. Testing for a MARC-like effect of an advantage for left-odd and right-even responses revealed no evidence of such an effect, as dRTs did not differ between odd and even numbers ($t(21) = -0.799, p = .433$). For comparability with other studies, we calculated slopes for both dRT by number and dRT by magnitude bin, see Table 1.

2.6 Discussion

Our results from a parity judgement task indicate that there was a clear horizontal SNARC effect in all notations. Importantly, and unlike previous results [20], this includes *Chinese characters*. This could be explained by differences in reading habit between participants from Taiwan [20] and mainland China (this study). But it should also be noted that this is the first time that a SNARC effect could be demonstrated in *Chinese hand signs*. The slope of -5.1 ms/digit for *Arabic digits* compared well to a recent meta-analysis of SNARC parity judgement experiments found in Western participants (-5.74 ms/digit [54]). The effect was not significantly smaller in other notations (all $p > .06$, which according to our Bayes factors was not just a product of insensitive data, although the explained variance varied somewhat, see Table 1. This is supported by the fact that BFs gave strong evidence in favour of a horizontal SNARC effect in all notations.

3 Experiment 2: Magnitude judgement

Experiment 2 used the same procedure as experiment 1, with the main difference that we used a magnitude judgement paradigm. That is, participants judged whether a given number was

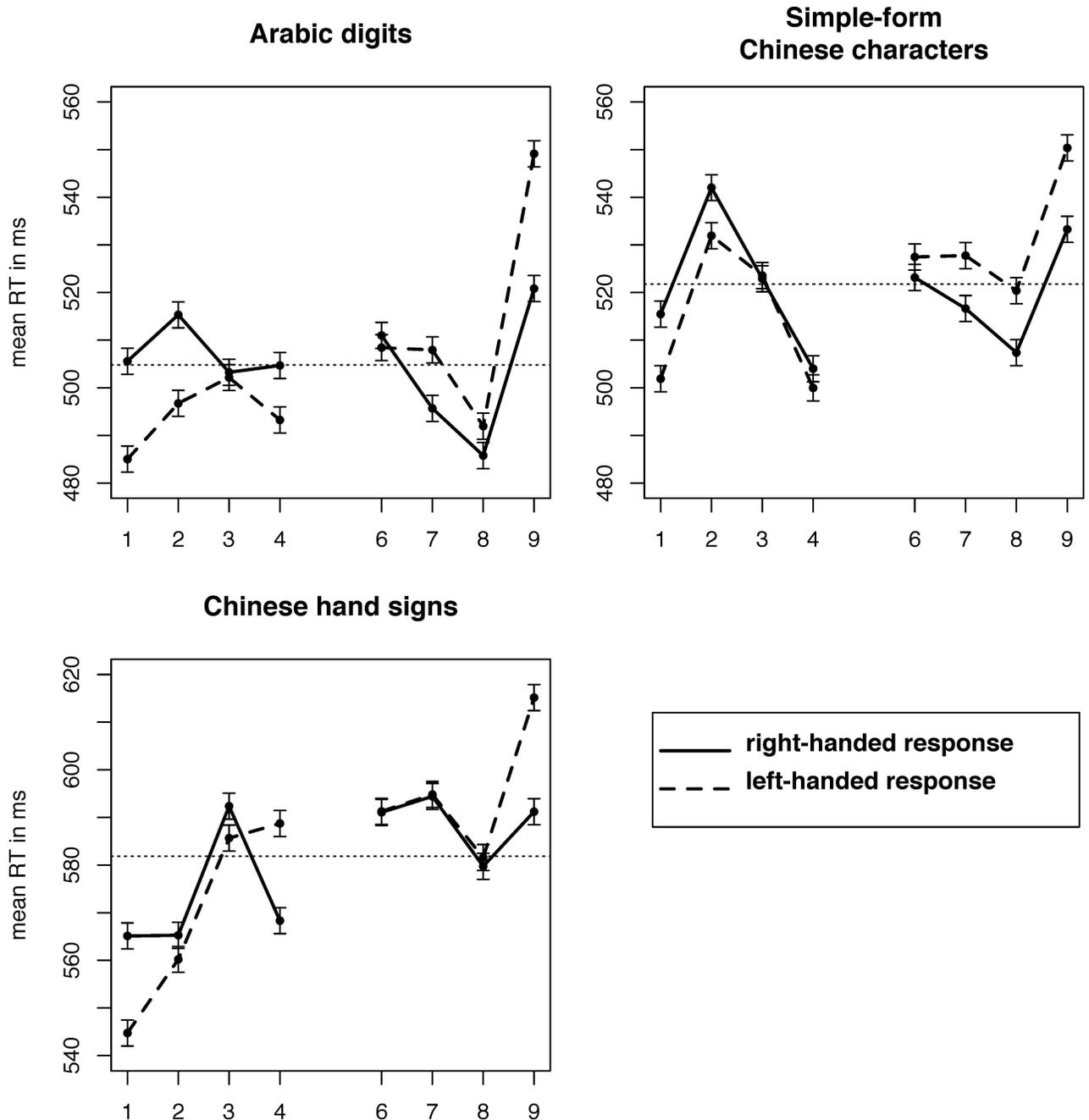


Fig 2. Left-handed and right-handed responses to each number in each notation, parity judgement task. SNARC effect: Right-handed responses slower than left-handed responses for small numbers, faster for large numbers. Error bars indicate within-subject SEMs for each number, pooled across each contrast of numbers [52,53]. Horizontal dashed lines indicate grand means of RTs for each notation.

doi:10.1371/journal.pone.0163897.g002

smaller or larger than 5. We chose to conduct another experiment since some participants told us during experiment 1 that they used visual features of some of these notations to complete the task. Thus, we used a different task with a different grouping of numbers, and additionally presented participants with a self-designed questionnaire following the experiment in which we asked them about any strategies used, allowing us to test if any SNARC effect we would find

would be robust to such strategies that might not rely on processing numerical magnitude (i.e., the main driver of the SNARC effect). Since the stimulus features would still allow completing the magnitude judgement task in Chinese hand signs based primarily on visuospatial features (see Fig 1), and it has been suggested that semantic processing of Chinese characters and Chinese hand signs may occur slightly later in processing than for Arabic digits [35], we also included analyses of numerical distance effects. A numerical distance effect suggests semantic processing of the stimuli [7,42], since it can best be explained by numerical magnitude, which is to say the meaning of a numerical stimulus.

3.1 Participants

Twenty-five native Chinese speakers (at least 18 years old, age: $M = 24.4$, $SD = 2.6$), recruited September and October 2013, participated voluntarily for an agreed pay of 30 RMB. All were right-handed by self-report, 11 were female. All had normal or corrected-to-normal vision. All were university students, 18 years or older, either enrolled or doing project work at Tsinghua University, Beijing and naive to the purpose of the task. Written, informed consent was obtained prior to the experiment from each participant according to the 1964 Declaration of Helsinki. Data were kept anonymously and could not be linked to names on the consent sheets—see section 2.1 for details.

3.2 Stimuli and apparatus

Participants were seated approximately 60 cm in front of an Acer (Acer Inc., New Taipei, Taiwan) laptop computer with a 15.6" flat screen (effective screen diagonal: 40 cm) running at 1024 * 768 pixel with a refresh rate of 60 Hz. Responses were given via an integrated German QWERTZ-keyboard. The stimuli used were identical to those in experiment 1. Due to the different screen, the size was slightly different at app. 20 mm * 20 mm, or approximately 2 degrees of visual angle. The main difference to experiment 1 was that participants now had to decide whether a given number was smaller or larger than 5 (magnitude judgement task). After testing, we administered a self-designed 9-item questionnaire containing mainly multiple-choice questions about their perception of the experiment and, crucially, about whether or not they used strategies other than number processing in any of the conditions. The question on the use of strategies was split between one multiple-choice item asking whether or not strategies were used, and an open-ended question asking participants who answered "yes" what strategies they used.

3.3 Procedure

The experiment was segmented into 6 blocks: Two blocks (*left = small* and *left = large*) for each of the three notations, *Arabic digits*, *Chinese characters*, and *Chinese hand signs*. The blocks, groups, instructions and feedback were analogous to experiment 1. Following the experiment, participants filled out the questionnaire. In total, the experiment lasted between 50 and 65 minutes.

3.4 Data analysis

One participant had to be excluded from analysis due to the number of errors made (more than 2 SD above the mean), leaving 24 participants for analysis. The data were analysed in the same way as in experiment 1 by computing an ANOVA, followed by dRTs and slope analyses over trimmed RTs, see section 2.4 for details. We also computed BFs for comparing a point-hypothesis at 0 to a uniform distribution from 0 to the mean dRT slope for magnitude

judgement experiments (-7.9 ms/digit) reported in a recent meta-analysis [54], see section 2.5. 1.7% of trials were excluded through the trimming procedure. To test for distance effects, we ran another ANOVA with the factors *notation* (3 levels) and *distance* from comparison (i.e., $|x-5|$, 4 levels) over response times, followed by regression analyses of dRT by number for each notation, see section 2.4.

3.5 Results

Again, we found a significant main effect for *numerical magnitude* ($F(3, 54) = 56.607, p_{GG} < .001, \epsilon_{GG} = .761$) and *notation* ($F(2, 36) = 88.489, p_{GG} < .001, \epsilon_{GG} = .932$), and this time also for *side* ($F(1, 18) = 8.159, p = .010$) on RTs. There were also significant interactions *side * magnitude* ($F(3, 54) = 9.586, p_{GG} < .001, \epsilon_{GG} = .521$) and *notation * magnitude* ($F(6, 108) = 15.022, p_{GG} < .001, \epsilon_{GG} = .774$), the former indicating a SNARC effect. Again, no three-way interaction of *side * magnitude * notation* was observed ($F(6, 108) = 1.973, p_{GG} = .130, \epsilon_{GG} = .491$). Similar to Experiment 1, the order of the tasks had hardly any effect on these results: There was no significant main effect of the factor order ($F(5, 18) = 0.597, p = .703$) and only the 3-way interaction of the factors *order * side * magnitude* ($F(15, 54) = 2.681, \epsilon_{GG} = .521, p_{GG} = .026$) was significant. All six other interactions with factor *order* were not significant (all $p > .05$). Future research and replications will need to clarify whether this significant 3-way interaction indicates a modulation of the SNARC effect by order of blocks, or whether this is a false positive due to testing of multiple interactions [55].

As in experiment 1, we computed regressions of dRTs by number (see Table 1). There was a significant SNARC effect for *Arabic digits* ($t(23) = -4.770, p < .001$), a trend for *Chinese characters* ($t(23) = -1.958, p = .062$), and no significant SNARC effect for *Chinese hand signs* ($t(23) = -0.943, p = .356$). Similarly, our Bayesian analysis shows decisive evidence for a SNARC effect in *Arabic digits* ($BF > 1000$), substantial evidence for a SNARC effect in *Chinese characters* ($BF = 4.5$), but inconclusive data regarding the effect in *Chinese hand signs* ($BF = 1.1$). RTs for left-handed and right-handed responses from this task can be seen in. Bayes factors testing for differences between dRTs in the data from this experiment were much less clear in experiment 2, giving substantial evidence for the notations *Arabic digits* and *Chinese characters* being similar ($BF = 3.1$), but showed that the data were in fact insensitive to detect a difference or absence thereof on the other comparisons (*Arabic digits* vs. *Chinese hand signs*: $BF = 0.4$; *Chinese characters* vs. *hand signs*: $BF = 1.1$).

An ANOVA with factors *notation* and *distance* over RTs revealed main effects of *distance* ($F(3, 69) = 93.289, p_{GG} < .001, \epsilon_{GG} = .828$), *notation* ($F(2, 46) = 71.528, p_{GG} < .001, \epsilon_{GG} = .990$), as well as a *notation * distance* interaction ($F(6, 138) = 4.001, p_{GG} = .003, \epsilon_{GG} = .751$). Regression slopes of distance by RT were significant for all notations (*Arabic digits*: $t(23) = -6.823, p < .001$; *Chinese characters*: $t(23) = -16.218, p < .001$; *Chinese hand signs*: $t(23) = -5.674, p < .001$).

Asking participants about their use of strategies other than number processing (see section 3.1 for the motivation, 3.2 for details on the questionnaire) revealed that 8 of 24 participants had used some visual strategy. One strategy was reported by multiple participants: 6 participants stated having categorised the shape of hand gestures by visual features, such as complexity or “straightness” of fingers to decide if numbers were smaller or larger than 5. We re-analysed the notation *Chinese hand signs* separately for participants who used this visual strategy and for those who did not. Grouping participants like this gave us mean regression coefficients of dRT by number of $y = 65.3 - 11.6x$ for participants who reported having used categorization based on visual features, and $y = -11.3 + 0.4x$ for participants who did not. Both of these slopes were not significantly different from 0 (both $p > .19$). However, the Bayesian

analysis revealed substantial evidence for the effect being truly non-existent in participants not using a strategy (BF = 0.3), while the data were insensitive for participants using a visual categorisation strategy (BF = 1.9). Interestingly, the distance effect was present for both groups (visual strategy: -4.3 ms/digit; no visual strategy: -7.2 ms/digit), see Fig 3.

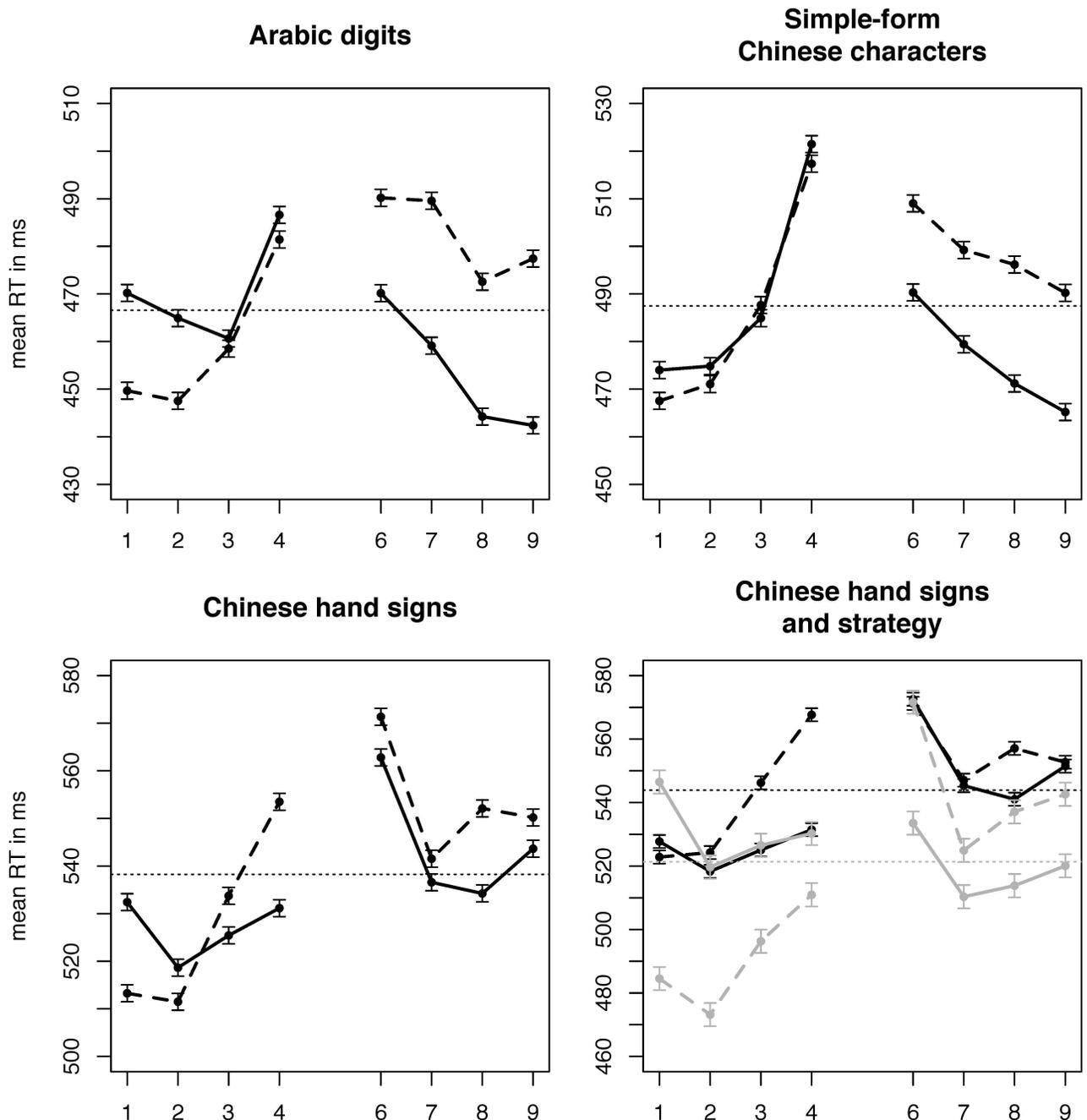


Fig 3. Left-handed and right-handed responses to each number in each notation, magnitude judgement. SNARC effect: Right-handed responses slower than left-handed responses for small numbers, faster for large numbers. **Distance effect:** Increased responses times for numbers closer to the middle. Bottom right: Participants who reported using visual categorisation (per our questionnaire; plotted in grey) vs. those who did not. Note the slightly compressed y-axis in this plot. Error bars indicate within-subject SEMs for each number, pooled across each contrast of numbers [52,53]. Horizontal dashed lines indicate grand means of RTs for each notation.

doi:10.1371/journal.pone.0163897.g003

3.6 Comparing the experiments

Running a 3 (*notation*) * 2 (*side*) * 4 (*numerical magnitude*) * 2 (*experiment*) ANOVA over RTs revealed a significant main effect for experiment. Participants responded faster in magnitude judgement ($F(1, 44) = 6.655, p = .013$). This was the case for all notations, as supported by the fact that there was no *experiment* * *notation* interaction ($F(2, 88) = 0.495, p_{GG} = .609, \epsilon_{GG} = .987$). However, there was no interaction of either *side* * *numerical magnitude* * *experiment* (which would indicate a different SNARC effect for each experiment— $F(3, 132) = 0.460, p_{GG} = .577, \epsilon_{GG} = .499$) or *side* * *numerical magnitude* * *notation* * *experiment* (which would indicate that notational differences in the SNARC effect depend on the type of experiment— $F(6, 264) = 0.442, p_{GG} = .748, \epsilon_{GG} = .568$).

As it has also been suggested that space-number associations become more salient the longer a participant takes to react [56], we conducted an analysis in which we aggregated RTs by latency bins. For this, we vincentized the data, such that for each quantile of the RT-distribution, separate means were computed [57,58]. We applied this procedure by calculating in each experiment vincentized RTs for each experimental condition (i.e., 2 side x 2 number x 3 notation). These vincentized RT were then used to calculate dRTs and the corresponding dRT slopes in the usual fashion (Fig 4). As expected [56], in both experiments and for all notations, dRT slopes decreased by bin (i.e., became more strongly negative, indicating a stronger SNARC effect). A 2 (*experiment*) * 3 (*notation*) * 4 (*bin*) ANOVA over dRT slopes revealed that this effect was statistically significant (main effect of *bin*: $F(2, 88) = 29.177, p_{GG} < .001, \epsilon_{GG} = .557$), in a way that could not be explained by any interaction with another factor (interaction *bin* * *notation*: $F(4, 176) = 2.365, p_{GG} = .087, \epsilon_{GG} = .613$; interaction *bin* * *experiment*: $F(2, 88) = 0.454, p_{GG} = .524, \epsilon_{GG} = .557$; all other interactions also n.s., all $p > .3$).

3.7 Discussion

In the magnitude judgement task, our regression analysis showed a clear horizontal SNARC effect in *Arabic digits* (-7.5 ms/digit; again very similar to the -7.90 ms/digit found for magnitude judgement by a recent meta-analysis [54]) a trend towards a SNARC effect in *Chinese characters*, and no significant SNARC effect in *Chinese hand signs*. Still, while the t-tests failed to reach significance, the ANOVA revealed no significant interaction between the SNARC effect and notation, indicating that these data do not support the non-existence of a horizontal SNARC effect. Bayes factors revealed that our data were too noisy to take this as strong evidence either for or against differences in the strength of the SNARC effect. In fact, even t-tests comparing regression slopes over dRT directly did not provide evidence for a difference in effects (all $p > .10$), and Table 1 shows the slopes to be quite similar. This is consistent with the Bayesian results, which indicate that there was a horizontal SNARC effect in *Chinese characters*, and the data were insensitive in the case of *Chinese hand signs*, rather than there being no effect.

4 General Discussion

For the first time, we demonstrated a horizontal SNARC effect in *Chinese characters* and *Chinese hand signs*. The only other study known to us that used Chinese characters was conducted in Taiwan, and no horizontal SNARC effect was found [20]. However, participants in this study were assumed to have different experiences of reading and writing in Chinese from our participants in Mainland China. Our results thus indicate that not finding a SNARC effect for Chinese characters in Taiwanese participants cannot be attributed to the notation alone. We also found a horizontal SNARC effect for Chinese hand signs, consistent with the idea that the spatial mapping of numbers is independent of notation and an integral part of number

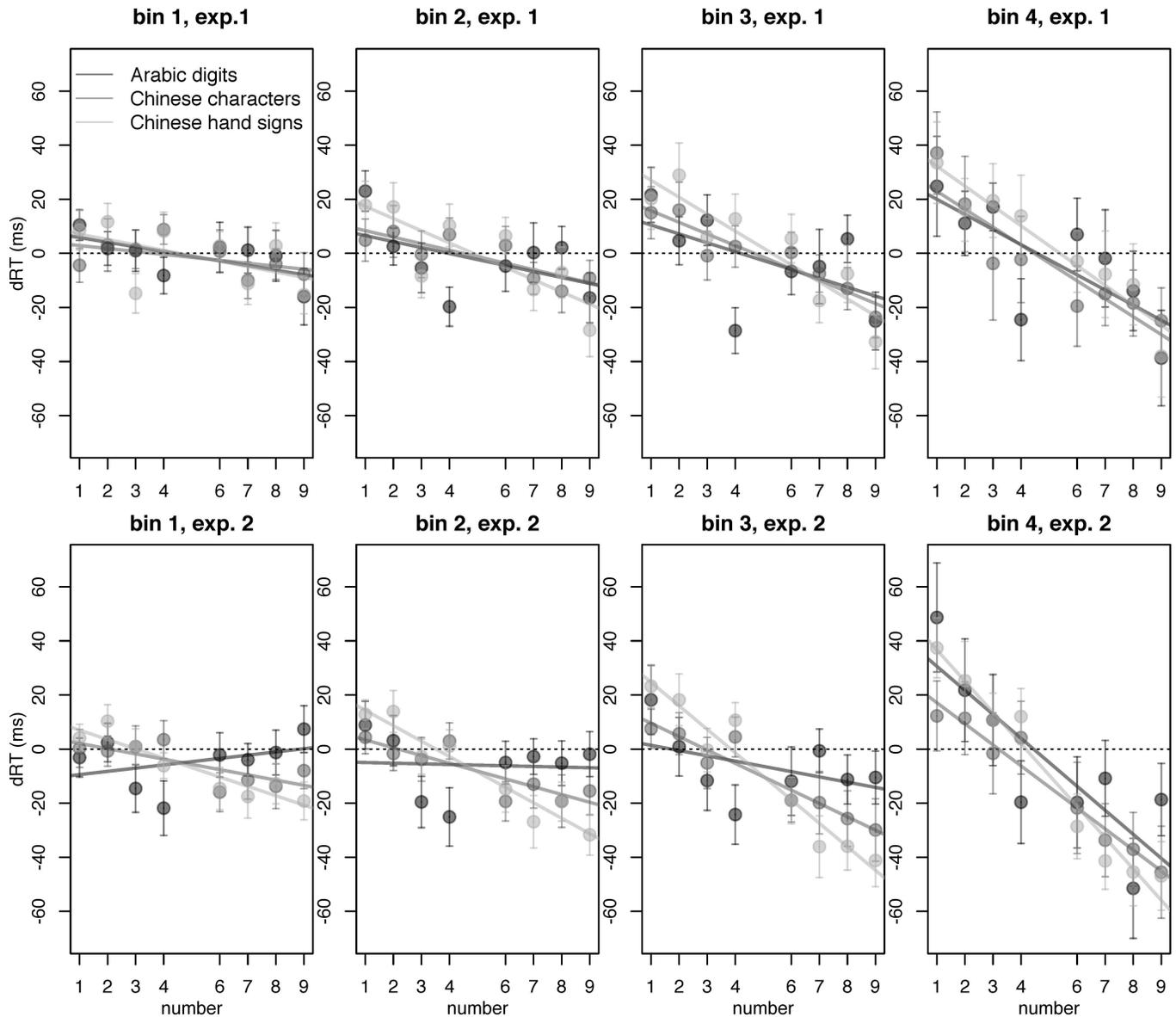


Fig 4. dRTs computed from vincentized RTs. Colours indicate the three notations, Arabic digits, Chinese characters, and Chinese hand signs. Bins in ascending order by RT (i.e., bin 1 contains the fastest responses). Note the more strongly negative dRT slope in bins with slower RTs.

doi:10.1371/journal.pone.0163897.g004

processing [59]. At the same time, it is worth mentioning that some previous studies have proposed a qualitatively different processing of Chinese characters and by Chinese speakers in general [35,36], in which case it is by no means obvious to expect similar effects. The SNARC effect was also less robust in Chinese hand signs than in other notations. Our data allow us only to speculate why this might be the case, but it is possible that this was due to a confounding influence of visual features like finger numerosity, which has been proposed to evoke space-number association [18] and could have interfered with a SNARC effect of number magnitude. Indeed, participants responded faster with the right hand to hand signs for 3 and 4 (represented through relatively high numerosities within our stimulus range). Still, the overall result is that the SNARC effect persists in Chinese characters and hand signs, with minor differences depending on experiment or notation. This emphasises that it may be a promising

tool to further investigate the mechanisms involved in the processing of Chinese numerals, and of numbers by Chinese speakers regardless of notation. Further research may also investigate which resources are used in the processing of these different notations. Finding a SNARC effect in Chinese hand signs makes this notation a potential tool to investigate phenomena of embodied numerical cognition. Our experiments do not allow such tests, as we did not manipulate any features of the hand signs, but there are some predictions of the embodiment theory that could be tested [31]: This theory predicts that finger-counting habits would impact the association of space and number, as should the orientation or posture of the hands. Finally, hand signs above 5 also offer the possibility of testing numerical cognition in hand signs that are not defined by finger numerosities, thus separating embodiment and numerosity.

Considering that the mechanisms giving rise to the SNARC effect likely differ somewhat between magnitude judgement and parity judgement [3,41], it was not clear whether to expect a similar effect in both our experiments. Our analyses revealed no quantifiable effect of the task on the SNARC effect, indicating that task differences may not have a big influence on the size of the effect. However, other properties of the data reflect the differences between the two tasks, as we found that there was vastly more inter-individual variability in the magnitude judgement task, which also led to the fact that SNARC effects in Chinese hand signs and Chinese characters failed to reach significance in magnitude judgement but not in parity judgement. We also observed the typical shape of the SNARC effect with almost constant dRTs on each side of the standard, but a big offset between the two sides (compare the distance between lines in Fig 3 to e.g. Fig 2 of [56]). These dRTs, along with dRTs split by latency quantile, can be seen in Fig 4. We split responses into four quantiles by latency to test for the time course of the SNARC effect in our experiment, as it has been proposed that the magnitude of the effect may increase for slower responses [56]. Splitting responses into bins by latency following the vincentizing procedure proposed by Ratcliff [57], we observed that in both experiments, slower responses showed a markedly stronger SNARC effect. This is consistent with previous results [56] and the time course of several other similar effects [60]. Gevers et al. [56] suggested that this may be due to weak activations taking longer to take effect, so that they would not influence relatively fast responses, which would in consequence show an on average weaker SNARC effect than slower responses. Our experiments did not test this prediction, but our results are very much in line with it.

It may be interesting in future studies to use these two Chinese notations to study the underlying mechanisms of the SNARC effect. Our finding of a robust distance effect across all participants, notations and number ranges in magnitude judgement represents evidence that numerals were indeed processed semantically in all these conditions, compatible with Liu et al.'s [36] view that Chinese numerals are encoded in parallel in multiple modalities. Of course, it is not surprising for Chinese native speakers to automatically process the meaning of hand signs—what is interesting is that processing to this level was fast enough to be detected in our task, as evidenced by the strong numerical distance effect.

This distance effect was present even when participants employed a visual categorisation strategy, and there was no significant speed difference between the two groups ($p = .524$, mean RT without strategy: 544 ms, with strategy: 521 ms). Indeed, this may serve as one possible explanation why somewhat counter intuitively and contrary to what we would have predicted, the SNARC effect was notably stronger instead of weaker in participants using a categorisation strategy vs. those who did not. If semantic processing occurred even when participants employed visual categorization, then it is not surprising to find the usual SNARC effect, in addition to a possible amplification of the typical offset [56] between two sides of a standard in magnitude judgement. In fact, the difference between these two groups was so large that participants who did use a strategy displayed the largest SNARC effect we found in our experiment,

while the participants who did not showed no SNARC effect at all. This is certainly compatible with the idea that in each task, different mechanisms beyond semantic processing (verbal in parity judgement, visuospatial in magnitude judgement) contribute to the effect, although the small number of participants reporting categorisation gives us only rather noisy data on this.

5 Conclusion

We found a horizontal SNARC effect in all three notations. For Chinese characters and Chinese hand signs, this is the first time that we know of that a horizontal SNARC effect has been demonstrated. These effects were slightly smaller than in Arabic digits. The effect persisted in all notations in the parity judgement task, with more mixed results in magnitude judgement. This speaks for the ubiquity of the SNARC effect in number representation and can be used in further research on differences in number processing between Chinese speakers and Western participants.

Acknowledgments

We thank Rachel Foster for proofreading the manuscript. We also thank two anonymous reviewers for their insightful comments that have greatly improved this paper.

Author Contributions

Conceptualization: KKK CL OE FS CSK CW XG VHF.

Data curation: KKK VHF.

Formal analysis: KKK FS VHF.

Funding acquisition: XG VHF.

Investigation: KKK CL.

Methodology: KKK CL OE FS CSK CW XG VHF.

Project administration: KKK CL.

Resources: KKK CL FS CSK CW XG.

Software: KKK FS VHF.

Supervision: XG VHF.

Validation: KKK CL FS.

Visualization: KKK.

Writing – original draft: KKK.

Writing – review & editing: KKK CL OE FS CSK CW XG VHF.

References

1. Dehaene S, Bossini S, Giraux P. The mental representation of parity and number magnitude. *J Exp Psychol Gen.* 1993; 122: 371–396. doi: [10.1037/0096-3445.122.3.371](https://doi.org/10.1037/0096-3445.122.3.371)
2. Fias W, Lauwereyns J, Lammertyn J. Irrelevant digits affect feature-based attention depending on the overlap of neural circuits. *Cogn Brain Res.* 2001; 12: 415–423. doi: [10.1016/s0926-6410\(01\)00078-7](https://doi.org/10.1016/s0926-6410(01)00078-7)
3. van Dijck JP, Gevers W, Fias W. Numbers are associated with different types of spatial information depending on the task. *Cognition.* Elsevier B.V.; 2009; 113: 248–253. doi: [10.1016/j.cognition.2009.08.005](https://doi.org/10.1016/j.cognition.2009.08.005) PMID: [19726035](https://pubmed.ncbi.nlm.nih.gov/19726035/)

4. Plaisier MA, Smeets JBJ. Number magnitude to finger mapping is disembodied and topological. *Exp Brain Res*. 2011; 209: 395–400. doi: [10.1007/s00221-011-2562-x](https://doi.org/10.1007/s00221-011-2562-x) PMID: [21298424](https://pubmed.ncbi.nlm.nih.gov/21298424/)
5. Fias W, Brysbaert M, Geypens F, d'Ydewalle G. The Importance of magnitude information in numerical processing: Evidence from the SNARC effect. *Math Cogn*. 1996; 2: 95–110. doi: [10.1080/135467996387552](https://doi.org/10.1080/135467996387552)
6. Izard V, Dehaene S. Calibrating the mental number line. *Cognition*. 2008; 106: 1221–1247. doi: [10.1016/j.cognition.2007.06.004](https://doi.org/10.1016/j.cognition.2007.06.004) PMID: [17678639](https://pubmed.ncbi.nlm.nih.gov/17678639/)
7. Dehaene S. Varieties of numerical abilities. *Cognition*. 1992; 44: 1–42. doi: [10.1016/0010-0277\(92\)90049-n](https://doi.org/10.1016/0010-0277(92)90049-n) PMID: [1511583](https://pubmed.ncbi.nlm.nih.gov/1511583/)
8. Dehaene S, Piazza M, Pinel P, Cohen L. Three parietal circuits for number processing. *Cogn Neuropsychol*. 2003; 20: 487–506. doi: [10.1080/02643290244000239](https://doi.org/10.1080/02643290244000239) PMID: [20957581](https://pubmed.ncbi.nlm.nih.gov/20957581/)
9. Santens S, Gevers W. The SNARC effect does not imply a mental number line. *Cognition*. 2008; 108: 263–270. doi: [10.1016/j.cognition.2008.01.002](https://doi.org/10.1016/j.cognition.2008.01.002) PMID: [18313655](https://pubmed.ncbi.nlm.nih.gov/18313655/)
10. Fias W. Two routes for the processing of verbal numbers: Evidence from the SNARC effect. *Psychol Res*. 2001; 65: 250–259. doi: [10.1007/s004260100065](https://doi.org/10.1007/s004260100065) PMID: [11789429](https://pubmed.ncbi.nlm.nih.gov/11789429/)
11. Gevers W, Ratinckx E, De Baene W, Fias W. Further evidence that the SNARC effect is processed along a dual-route architecture. *Exp Psychol*. 2006; 53: 58–68. doi: [10.1027/1618-3169.53.1.58](https://doi.org/10.1027/1618-3169.53.1.58) PMID: [16610273](https://pubmed.ncbi.nlm.nih.gov/16610273/)
12. Proctor RW, Cho YS. Polarity correspondence: A general principle for performance of speeded binary classification tasks. *Psychol Bull*. 2006; 132: 416–442. doi: [10.1037/0033-2909.132.3.416](https://doi.org/10.1037/0033-2909.132.3.416) PMID: [16719568](https://pubmed.ncbi.nlm.nih.gov/16719568/)
13. Cho YS, Proctor RW. When is an odd number not odd? Influence of task rule on the MARC effect for numeric classification. *J Exp Psychol Learn Mem Cogn*. 2007; 33: 832–843. doi: [10.1037/0278-7393.33.5.832](https://doi.org/10.1037/0278-7393.33.5.832) PMID: [17723063](https://pubmed.ncbi.nlm.nih.gov/17723063/)
14. van Dijck JP, Gevers W, Lafosse C, Fias W. The heterogenous nature of number-space interactions. *Front Hum Neurosci*. 2012; 5: Article 182. doi: [10.3389/fnhum.2011.00182](https://doi.org/10.3389/fnhum.2011.00182) PMID: [22291629](https://pubmed.ncbi.nlm.nih.gov/22291629/)
15. Nuerk H-C, Wood G, Willmes K. The universal SNARC effect. The association between number magnitude and space is amodal. *Exp Psychol*. 2005; 52: 187–194. doi: [10.1027/1618-3169.52.3.187](https://doi.org/10.1027/1618-3169.52.3.187) PMID: [16076066](https://pubmed.ncbi.nlm.nih.gov/16076066/)
16. Gevers W, Reynvoet B, Fias W. The mental representation of ordinal sequence is spatially organized. *Cognition*. 2003; 87: B87–B95. doi: [10.1016/s0010-0277\(02\)00234-2](https://doi.org/10.1016/s0010-0277(02)00234-2) PMID: [12684205](https://pubmed.ncbi.nlm.nih.gov/12684205/)
17. Shaki S, Petrusic WM, Leth-Steensen C. SNARC effects with numerical and non-numerical symbolic comparative judgments: Instructional and cultural dependencies. *J Exp Psychol Hum Percept Perform*. 2012; 38: 515–530. doi: [10.1037/a0026729](https://doi.org/10.1037/a0026729) PMID: [22288694](https://pubmed.ncbi.nlm.nih.gov/22288694/)
18. Iversen W, Nuerk H-C, Jäger L, Willmes K. The influence of an external symbol system on number parity representation, or What's odd about 6? *Psychon Bull Rev*. 2006; 13: 730–736. doi: [10.3758/bf03193988](https://doi.org/10.3758/bf03193988) PMID: [17201377](https://pubmed.ncbi.nlm.nih.gov/17201377/)
19. Ito Y, Hatta T. Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Mem Cognit*. 2004; 32: 662–673. doi: [10.3758/bf03195857](https://doi.org/10.3758/bf03195857) PMID: [15478760](https://pubmed.ncbi.nlm.nih.gov/15478760/)
20. Hung Y, Hung DL, Tzeng OJL, Wu DH. Flexible spatial mapping of different notations of numbers in Chinese readers. *Cognition*. 2008; 106: 1441–1450. doi: [10.1016/j.cognition.2007.04.017](https://doi.org/10.1016/j.cognition.2007.04.017) PMID: [17572403](https://pubmed.ncbi.nlm.nih.gov/17572403/)
21. Dehaene S, Molko N, Cohen L, Wilson AJ. Arithmetic and the brain. *Curr Opin Neurobiol*. 2004; 14: 218–224. PMID: [15082328](https://pubmed.ncbi.nlm.nih.gov/15082328/)
22. Bächtold D, Baumüller M, Brugger P. Stimulus-response compatibility in representational space. *Neuropsychologia*. 1998; 36: 731–735. doi: [10.1016/s0028-3932\(98\)00002-5](https://doi.org/10.1016/s0028-3932(98)00002-5) PMID: [9751438](https://pubmed.ncbi.nlm.nih.gov/9751438/)
23. Fischer MH, Shaki S, Cruise A. It takes just one word to quash a SNARC. *Exp Psychol*. 2009; 56: 361–366. doi: [10.1027/1618-3169.56.5.361](https://doi.org/10.1027/1618-3169.56.5.361) PMID: [19447752](https://pubmed.ncbi.nlm.nih.gov/19447752/)
24. Gevers W, Lammertyn J. The hunt for SNARC. *Psychol Sci*. 2005; 47: 10–21.
25. Wood G, Willmes K, Nuerk H-C, Fischer MH. On the cognitive link between space and number: a meta-analysis of the SNARC effect. *Psychol Sci Q*. 2008; 50: 489–525.
26. Göbel S. Up or down? Reading direction influences vertical counting direction in the horizontal plane—A cross-cultural comparison. *Front Psychol*. 2015; 6: 1–12. doi: [10.3389/fpsyg.2015.00228](https://doi.org/10.3389/fpsyg.2015.00228) PMID: [25852583](https://pubmed.ncbi.nlm.nih.gov/25852583/)
27. Shaki S, Fischer MH, Petrusic WM. Reading habits for both words and numbers contribute to the SNARC effect. *Psychon Bull Rev*. 2009; 16: 328–331. doi: [10.3758/PBR.16.2.328](https://doi.org/10.3758/PBR.16.2.328) PMID: [19293102](https://pubmed.ncbi.nlm.nih.gov/19293102/)

28. Gevers W, Lammertyn J, Notebaert W, Verguts T, Fias W. Automatic response activation of implicit spatial information: Evidence from the SNARC effect. *Acta Psychol.* 2006; 122: 221–233. doi: [10.1016/j.actpsy.2005.11.004](https://doi.org/10.1016/j.actpsy.2005.11.004) PMID: [16423320](https://pubmed.ncbi.nlm.nih.gov/16423320/)
29. Holmes KJ, Lourenco SF. Orienting numbers in mental space: Horizontal organization trumps vertical. *Q J Exp Psychol.* 2012; 65: 1044–1051. doi: [10.1080/17470218.2012.685079](https://doi.org/10.1080/17470218.2012.685079) PMID: [22630356](https://pubmed.ncbi.nlm.nih.gov/22630356/)
30. Schwarz W, Keus IM. Moving the eyes along the mental number line: Comparing SNARC effects with saccadic and manual responses. *Percept Psychophys.* 2004; 66: 651–664. doi: [10.3758/bf03194909](https://doi.org/10.3758/bf03194909) PMID: [15311664](https://pubmed.ncbi.nlm.nih.gov/15311664/)
31. Fischer MH, Brugger P. When digits help digits: Spatial-numerical associations point to finger counting as prime example of embodied cognition. *Front Psychol.* 2011; 2: 1–7. doi: [10.3389/fpsyg.2011.00260](https://doi.org/10.3389/fpsyg.2011.00260) PMID: [22028696](https://pubmed.ncbi.nlm.nih.gov/22028696/)
32. Fischer MH. Finger counting habits modulate spatial-numerical associations. *Cortex.* 2008; 44: 386–392. doi: [10.1016/j.cortex.2007.08.004](https://doi.org/10.1016/j.cortex.2007.08.004) PMID: [18387569](https://pubmed.ncbi.nlm.nih.gov/18387569/)
33. Previtali P, Rinaldi L, Girelli L. Nature or nurture in finger counting: A review of the determinants of the direction of number-finger mapping. *Front Psychol.* 2011; 2: Article 363. doi: [10.3389/fpsyg.2011.00363](https://doi.org/10.3389/fpsyg.2011.00363) PMID: [22319502](https://pubmed.ncbi.nlm.nih.gov/22319502/)
34. Tschentscher N, Hauk O, Fischer MH, Pulvermüller F. You can count on the motor cortex: Finger counting habits modulate motor cortex activation evoked by numbers. *Neuroimage.* 2012; 59: 3139–3148. doi: [10.1016/j.neuroimage.2011.11.037](https://doi.org/10.1016/j.neuroimage.2011.11.037) PMID: [22133748](https://pubmed.ncbi.nlm.nih.gov/22133748/)
35. Cao B, Li F, Li H. Notation-dependent processing of numerical magnitude: Electrophysiological evidence from Chinese numerals. *Biol Psychol.* 2010; 83: 47–55. doi: [10.1016/j.biopsycho.2009.10.003](https://doi.org/10.1016/j.biopsycho.2009.10.003) PMID: [19854237](https://pubmed.ncbi.nlm.nih.gov/19854237/)
36. Liu C, Tang H, Luo Y, Mai X. Multi-representation of symbolic and nonsymbolic numerical magnitude in Chinese number processing. *PLoS One.* 2011; 6: e19373. doi: [10.1371/journal.pone.0019373](https://doi.org/10.1371/journal.pone.0019373) PMID: [21541303](https://pubmed.ncbi.nlm.nih.gov/21541303/)
37. Liu C, Mai X, Fu X. The spatial numerical association of response codes effect of number processing in different attention conditions. *Acta Psychol Sin.* 2004; 36: 671–680.
38. Laws and Regulations Database of the Republic of China [Internet]. [cited 5 May 2016]. Available: <http://law.moj.gov.tw/eng/LawClass/LawSearchNo.aspx?PC=A0030133&DF=&SNo=8>
39. Annotated Republic of China Laws/Central Regulation Standard Act/Article 8 [Internet]. [cited 5 May 2016]. Available: https://en.wikibooks.org/wiki/Annotated_Republic_of_China_Laws/Central_Regulation_Standard_Act/Article_8
40. Bull R, Blatto-Valle G, Fabich M. Subitizing, magnitude representation, and magnitude retrieval in deaf and hearing adults. *J Deaf Stud Deaf Educ.* 2006; 11: 289–302. doi: [10.1093/deafed/enj038](https://doi.org/10.1093/deafed/enj038) PMID: [16601145](https://pubmed.ncbi.nlm.nih.gov/16601145/)
41. Herrera A, Macizo P, Semenza C. The role of working memory in the association between number magnitude and space. *Acta Psychol (Amst).* 2008; 128: 225–237. doi: [10.1016/j.actpsy.2008.01.002](https://doi.org/10.1016/j.actpsy.2008.01.002) PMID: [18295166](https://pubmed.ncbi.nlm.nih.gov/18295166/)
42. Moyer RS, Landauer TK. Time required for judgements of numerical inequality. *Nature.* 1967; 215: 1519–1520. doi: [10.1038/2151519a0](https://doi.org/10.1038/2151519a0) PMID: [6052760](https://pubmed.ncbi.nlm.nih.gov/6052760/)
43. Kleiner M, Brainard D, Pelli D. What's new in Psychtoolbox-3? *Perception* 36 ECPV. 2007.
44. Nuerk H-C, Iversen W, Willmes K. Notational modulation of the SNARC and the MARC (linguistic markedness of response codes) effect. *Q J Exp Psychol A Hum Exp Psychol.* 2004; 57: 835–863. doi: [10.1080/02724980343000512](https://doi.org/10.1080/02724980343000512) PMID: [15204120](https://pubmed.ncbi.nlm.nih.gov/15204120/)
45. Tzelgov J, Zohar-Shai B, Nuerk H-C. On defining quantifying and measuring the SNARC effect. *Front Psychol.* 2013; 4: 302. doi: [10.3389/fpsyg.2013.00302](https://doi.org/10.3389/fpsyg.2013.00302) PMID: [23750147](https://pubmed.ncbi.nlm.nih.gov/23750147/)
46. Pinhas M, Tzelgov J, Ganor-Stern D. Estimating linear effects in ANOVA designs: The easy way. *Behav Res Methods.* 2012; 44: 788–794. doi: [10.3758/s13428-011-0172-y](https://doi.org/10.3758/s13428-011-0172-y) PMID: [22101656](https://pubmed.ncbi.nlm.nih.gov/22101656/)
47. Merriam-Webster. Markedness [Internet]. [cited 14 May 2016]. Available: <http://www.merriam-webster.com/dictionary/marked>
48. Greenhouse SW, Geisser S. On methods in the analysis of profile data. *Psychometrika.* 1959; 24: 95–112. doi: [10.1007/bf02289823](https://doi.org/10.1007/bf02289823)
49. Dienes Z. *Understanding psychology as a science. An introduction to scientific and statistical inference.* New York: Palgrave Macmillan; 2008.
50. Dienes Z. Bayesian versus Orthodox statistics: Which side are you on? *Perspect Psychol Sci.* 2011; 6: 274–290. doi: [10.1177/1745691611406920](https://doi.org/10.1177/1745691611406920) PMID: [26168518](https://pubmed.ncbi.nlm.nih.gov/26168518/)
51. Jeffreys H. *Theory of probability.* 3rd ed. Mott NF, Bullard EC, Wilkinson DH, editors. The international series of monographs on physics. Oxford: Oxford University Press; 1961.

52. Franz VH, Loftus GR. Standard errors and confidence intervals in within-subjects designs: Generalizing Loftus and Masson (1994) and avoiding the biases of alternative accounts. *Psychon Bull Rev.* 2012; 19: 395–404. doi: [10.3758/s13423-012-0230-1](https://doi.org/10.3758/s13423-012-0230-1) PMID: [22441956](https://pubmed.ncbi.nlm.nih.gov/22441956/)
53. Loftus GR, Masson MEJ. Using confidence intervals in within-subject designs. *Psychon Bull Rev.* 1994; 1: 476–490. doi: [10.3758/BF03210951](https://doi.org/10.3758/BF03210951) PMID: [24203555](https://pubmed.ncbi.nlm.nih.gov/24203555/)
54. Schiller F, Eloka O, Franz VH. Using key distance to evaluate theories on the SNARC. *Perception.* 2016; 45: 196–221. doi: [10.1177/0301006615616754](https://doi.org/10.1177/0301006615616754) PMID: [26822868](https://pubmed.ncbi.nlm.nih.gov/26822868/)
55. Cramer AOJ, van Ravenzwaaij D, Matzke D, Steingroever H, Wetzels R, Grasman RPPP, et al. Hidden multiplicity in multiway ANOVA: Prevalence, consequences, and remedies. *Psychon Bull Rev.* 2016; 23: 640–647. doi: [10.3758/s13423-015-0913-5](https://doi.org/10.3758/s13423-015-0913-5) PMID: [26374437](https://pubmed.ncbi.nlm.nih.gov/26374437/)
56. Gevers W, Verguts T, Reynvoet B, Caessens B, Fias W. Numbers and space: a computational model of the SNARC effect. *J Exp Psychol Hum Percept Perform.* 2006; 32: 32–44. doi: [10.1037/0096-1523.32.1.32](https://doi.org/10.1037/0096-1523.32.1.32) PMID: [16478324](https://pubmed.ncbi.nlm.nih.gov/16478324/)
57. Ratcliff R. Group reaction time distributions and an analysis of distribution statistics. *Psychol Bull.* 1979; 86: 446–461. doi: [10.1037/0033-2909.86.3.446](https://doi.org/10.1037/0033-2909.86.3.446) PMID: [451109](https://pubmed.ncbi.nlm.nih.gov/451109/)
58. Vincent SB. The function of the viborissae in the behavior of the white rat. *Behav Monogr.* 1912; 1: No. 5.
59. Fischer MH, Shaki S. Two steps to space for numbers. *Front Psychol.* 2015; 6: 00612. doi: [10.3389/fpsyg.2015.00612](https://doi.org/10.3389/fpsyg.2015.00612) PMID: [26029145](https://pubmed.ncbi.nlm.nih.gov/26029145/)
60. Proctor RW, Miles JD, Baroni G. Reaction time distribution analysis of spatial correspondence effects. *Psychon Bull Rev.* 2011; 18: 242–266. doi: [10.3758/s13423-011-0053-5](https://doi.org/10.3758/s13423-011-0053-5) PMID: [21327376](https://pubmed.ncbi.nlm.nih.gov/21327376/)

Comparing symbolic and nonsymbolic number lines: Consistent effects of notation across output measures

Karl K. Kopiske^a, Volker H. Franz^b

a: Istituto Italiano di Tecnologia, Center for Neuroscience and Cognitive Systems@UniTn. Corso Bettini, 31, 38068 Rovereto, TN, Italy.

b: University of Tübingen, Department of Computer Science, Experimental Cognitive Sciences, Sand 13, 72076 Tübingen, Germany.

corresponding author:

Karl K. Kopiske

email:

karl.kopiske@iit.it

address:

Center for Neuroscience and Cognitive Systems@UniTn,

Istituto Italiano di Tecnologia,

Corso Bettini, 31,

38068 Rovereto, TN,

Italy.

Running head: Symbolic and nonsymbolic number lines

Word count: 7121 (excl. title page, abstract, tables, references, figures, footnotes; overall: 8930)

1 ABSTRACT – The mental number line (MNL) is a popular metaphor for magnitude
2 representation in numerical cognition and is often assumed to represent numerical input for
3 manipulation whenever a transformation between magnitude formats is required. The shape of
4 the MNL has been much investigated and has frequently been reported as being nonlinear. We
5 investigated whether this shape reflects a phenomenon of the mapping from stimulus to internal
6 magnitude representation or of the mapping from internal representation to response. In five
7 experiments (total N = 66), participants adjusted the location of magnitudes on a ruler-like
8 response bar. Magnitudes were either presented symbolically to the participants (i.e., as arabic
9 number), or nonsymbolically (i.e., as clouds of dots). Responses to symbolic stimuli were
10 linear, while responses to nonsymbolic stimuli were power-shaped. This suggests that the
11 nonlinearity is due to the mapping from stimulus to internal representation. We also investigated
12 whether the nonlinearity could be explained by effects of previous trials, but such effects were
13 (a) not strong enough to explain the nonlinear responses, and (b) existed only between trials of
14 the same input notation, corroborating the notion that the nonlinearity is due to input
15 transformations. Introducing veridical or distorted feedback affected the responses, but only
16 temporarily and responses to nonsymbolic stimuli remained nonlinear. We conclude that the
17 nonlinearity is a phenomenon of the mapping from nonsymbolic input format to internal
18 magnitude representation and that the phenomenon is surprisingly robust to calibration

19

20 **Keywords:** Numerical cognition, nonsymbolic magnitude, number line, calibration

21

1 **1 Introduction**

2 The model of the mental number-line (MNL) for an internal scale of numerical magnitude has
3 been around at least since the 1960's (Moyer & Landauer, 1967), but was in its current form
4 postulated by Dehaene (1992). Dehaene proposed the MNL to be one element of a model of
5 number representation that sought to explain, among other things, the ability of neurological
6 patients to make approximate, but not exact judgements based on simple verbal input, as well as
7 spatial-numerical stimulus-response associations (SNARC - Dehaene, Bossini, & Giraux, 1993).
8 In this view, the MNL determines the internal mapping between numbers and other forms of
9 magnitude. In this article, we use the MNL in a similar sense to refer to the *internal*
10 representation that gives rise to an *observable* response function. Our goal was to investigate
11 encoding mechanisms and representation-to-output transformations that may give rise to
12 particular attributes of this response function, in particular its nonlinear shape (Dehaene, 2003).

13 *1.1 The shape of the MNL, and its relation to non-symbolic number*

14 Dehaene (1992) noted that the mapping of non-symbolic magnitude to symbolic magnitude
15 (e.g., numbers) tends to be non-veridical, showing systematic underestimation for large
16 magnitudes; a finding that caused him to propose a MNL that was compressed and possibly
17 logarithmic in shape: $\text{response}(x) \sim \log(x)$, with x being the numerical magnitude of the stimulus.
18 Two things should be noted with respect to the shape of the MNL.

19 Firstly, the MNL is not the same as this response function. The MNL refers to the internal
20 representation of magnitude, the response function to the output generated by a participant in a
21 task. While the response function might be nonlinear, the internal representation might still be
22 linear. A related point is that the origin of both the shapes of the MNL and the response function
23 has been much debated (Barth & Paladino, 2011; Cantlon, Cordes, Libertus, & Brannon, 2009;

1 Cicchini, Anobile, & Burr, 2014; Cohen & Blanc-Goldhammer, 2011; Dehaene, 2003;
2 Dehaene, Izard, Pica, & Spelke, 2009; Siegler & Opfer, 2003), as has the question whether
3 testing one allows inferences about the other. Drawing conclusions from a response about an
4 internal representation requires knowledge of the mapping between the two, and of potential
5 response biases. For example, it has been argued that in a classic task of locating numbers (or
6 other forms of magnitude) on a horizontal line akin to a ruler, participants actually perform a
7 proportion judgement, which in turn relies heavily on reference points (Barth & Paladino, 2011;
8 but see also Opfer, Siegler, & Young, 2011). At the same time, this ruler-like task is one of the
9 few tasks that allow a non-symbolic output of magnitude, which bypasses a non-symbolic-to-
10 symbolic transformation that is required for other tasks as, for example, verbalising magnitude. A
11 recently proposed potential solution to the problem of proportion judgments is allowing
12 participants to go beyond the presented ruler, thereby effectively allowing a judgement of
13 multiples (Cohen & Blanc-Goldhammer, 2011; Link, Huber, Nuerk, & Moeller, 2014). Another
14 potential confound that may bias responses independently of internal representation is the known
15 tendency towards the mean, be it of a scale or of previous responses. This has recently been
16 brought up as criticism of the notion that a compressed response suggests a compressed
17 representation of magnitude (Cicchini et al., 2014), and has been applied to other judgements for
18 a long time (Haubensak, 1992; Parducci & Perret, 1971).

19 Secondly, the systematic underestimation in numerical estimation for large magnitudes can be
20 explained in several ways that do not assume a logarithmic transformation. A fairly similar view

1 is that of the response being a power function¹ of x , which is indeed what even proponents of
2 an internal logarithmic MNL have argued (Izard & Dehaene, 2008). This will often result in very
3 similar fits in behavioural data (indeed, the models may be virtually indistinguishable unless the
4 number range is extended to include very large numbers, see Opfer et al., 2011), and fit similarly
5 well to the corresponding neural activation (Nieder & Miller, 2003). That said, the two are based
6 on slightly different classic concepts with slightly different implications, as a logarithmic
7 function implies an additive effect when stimulus magnitude is increased by a given factor
8 (Fechner, 1860), while a power function implies a multiplicative effect of the same increase
9 (Stevens, 1957). Finally, in designs employing a bounded response, a linear internal
10 representation may also be compatible with a compressed response function if another
11 assumption is made, that of size-dependent variability (see Weber's Law, Fechner, 1860; such a
12 relationship has also been found for transformations between symbolic and non-symbolic
13 magnitudes, see e.g. Cordes, Gelman, Gallistel, & Whalen, 2001; Dehaene, 1992; Whalen,
14 Gallistel, & Gelman, 1999): If variability increases with the response, a larger tail of the
15 distribution for large responses would be truncated by the bound than for smaller response,
16 resulting in a systematic underestimation of large magnitudes (see e.g. Cantlon et al., 2009).

¹ Such a function has the form $response(x) = b * x^a$ and is often modeled and plotted as $log(response(x)) = a*log(x) + b$, taking advantage of the fact that $a*log(x) = log(x^a)$ and $log(b) + log(x) = log(b*x)$. In other words, a power function can be fitted linearly when a logarithm is applied to both sides of the equation, and therefore looks linear in a log-log plot.

1 *1.2 Our study*

2 Our goals were twofold: Firstly, to investigate which step of an input-output transformation
3 gives the MNL its characteristic nonlinear shape – that is, whether the same shape would be
4 achieved with not only symbolic-output measures, but also different variations of a non-symbolic
5 number estimation task. To do this, we compared number lines obtained from symbolic-to-
6 nonsymbolic, as well as nonsymbolic-to-symbolic transformations, having participants map
7 different types of input to the same output measure, as well as the same input to different output
8 measures.

9 Secondly, we wanted to find out if previous-trials effects and calibration (or a lack thereof)
10 could explain the shape of the responses, and if so if any effects of previous trials and of
11 calibration would persist between trials of different input types, as well as different output
12 measures.

13 To test this, we employed two methods: A nonsymbolic response, in which participants were
14 asked to indicate the magnitude represented by the stimulus on a response bar akin to a ruler
15 (similar to previous studies, e.g. Cicchini et al., 2014; Dehaene, Izard, Spelke, & Pica, 2008;
16 Siegler & Opfer, 2003), as well as a simple numeric response, in which participants typed the
17 corresponding number on a keyboard. The ruler-based task was performed with both Arabic
18 numbers (experiments 1-4) and clouds of dots as nonsymbolic stimulus magnitudes (experiments
19 2-5). We also varied features of the response bar between experiments to rule out alternative
20 explanations (e.g., we employed both numerosities and numbers as endpoints). The number
21 response task was performed only in experiment 5. In this experiment, participants received both
22 veridical as well as perturbed feedback in order to investigate the mechanisms behind calibration
23 of the MNL. For an overview of all conditions and experiments, see Table 1.

1 **Table 1:** Tasks and stimuli used in our experiments.

Experiment	Input	Output	Endpoints	Comment
<i>Exp. 1</i>	symbolic	response bar	symbolic	
<i>Exp. 2a</i>	symbolic	response bar	symbolic	
<i>Exp. 2b</i>	non-symbolic	response bar	symbolic	
<i>Exp. 3a</i>	symbolic	response bar	symbolic	starting position random
<i>Exp. 3b</i>	non-symbolic	response bar	non-symbolic	starting position random
<i>Exp. 4a</i>	symbolic	response bar	symbolic	endpoints & starting position random
<i>Exp. 4b</i>	non-symbolic	response bar	non-symbolic	endpoints & starting position random
<i>Exp. 5a</i>	non-symbolic	response bar	non-symbolic	with feedback; otherwise like 3b
<i>Exp. 5b</i>	non-symbolic	typed number	none	with feedback

2 *Note:* Details about the response bar are described in section 2.1.2. Detail about the stimuli used as non-
3 symbolic endpoints given in section 3.2.2. For each experiment, conditions a and b were presented in a
4 blocked design that included both separate and interleaved blocks.

5

6 For each experiment and each type of input and output, we fit separate linear, logarithmic and
7 power functions to assess the shape of the response function. Our predictions were as follows:
8 We expected the response function for Arabic digits to be almost linear, and the response
9 function for nonsymbolic magnitudes to be better fitted through a logarithmic or power-function
10 model. We expected this relation to hold true for both the ruler-based task (experiments 1 ... 4)
11 and the number-response task (experiment 5). We also expected dependencies between
12 consecutive trials. If these were equally strong between different types as within types, this would
13 speak for a calibration of the mapping from internal magnitude to the number line, or a
14 calibration of nonsymbolic magnitude-to-internal magnitude mapping. These two possibilities
15 were investigated in experiment 5, where the latter would predict a calibration effect between
16 tasks.

1 **2 Experiment 1: Testing the response function**

2 In the ruler-based task employed in all experiments, participants clicked on a response bar
3 displayed horizontally on a computer screen. A similar method has been used in numerous
4 experiments (e.g., Cicchini et al., 2014; Dehaene et al., 2008; Nieder & Miller, 2003). To verify
5 that this method was not in itself susceptible to artefacts, we tested it in experiment 1 with the
6 simplest, most straightforward number line task we could find: Mapping Arabic numbers to a
7 horizontal ruler marked on the left and on the right by Arabic numbers (see Figure 1).

8 *2.1 Methods*

9 *2.1.1 Participants*

10 Six participants (age range: 25 to 39 years, mean age 32.2, 4 female) took part in the
11 experiment. All participants were volunteers taking part without compensation, consisting of
12 graduate students and faculty members of the University of Hamburg Department of Psychology.
13 In this and all following experiments, participants gave written, informed consent and their data
14 were protected according the 1964 Declaration of Helsinki.

15 *2.1.2 Apparatus*

16 Sitting in front a 21'' LCD monitor (effective screen diagonal: 52 cm), participants were
17 presented with a centrally displayed number (font size 60 pixel, app. 2° of visual angle). Below
18 the number, a response bar was displayed (located near the bottom of the screen, centrally on the
19 x-axis, app. 20° of visual angle). In the middle of the response bar was a black block (a square of
20 20 px * 20 px, corresponding to approx. 6 mm * 6 mm). The block could be moved horizontally
21 by the participants using a standard USB mouse. Participants were asked via on-screen

1 instructions to move the block to the location that they perceived as the position the number
2 belonged to and then click the mouse button to register this position (see Figure 1).

3 *2.1.3 Procedure*

4 In each trial, the number was displayed until the participant either performed the mouse click or
5 a maximum of three seconds had elapsed; after which a fixation cross appeared for on average
6 500 ms (inter-stimulus-interval was defined as 400 ms + a pseudo-random value from an
7 exponential distribution with $m = 100$ ms) until the next trial. Trials that were not completed in
8 time were repeated at a random time later during the experiment and a corresponding error
9 message ("Please answer within 3 seconds!") appeared on the screen. At the start of the
10 experiment, a response bar was presented on the screen below the instructions to let participants
11 familiarise themselves with the bar and the adjustment block for as long as they liked, while
12 clicking was disabled. All experiments were implemented in a custom MATLAB program using
13 Psychtoolbox 3 (Kleiner, Brainard, & Pelli, 2007).

14 To be able to verify the results from our models and to enable easier detection of previous-trial
15 effects, the experiment was conducted in a blocked design, with one block containing pseudo-
16 random numbers (in which stimuli were randomly drawn from numbers 1...200, with no
17 duplicates), as well as two oddball blocks. These were included to test specifically to what degree
18 responses would be influenced by the trials directly preceding them, as is often found in repeated-
19 measures designs (Haubensak, 1992) and has been proposed specifically for processing of
20 nonsymbolic number (Cicchini et al., 2014). In these blocks, either were 87.5% of trials high
21 numbers (from top third, 134...200) and 12.5% of trials numbers from the opposite end of the
22 range (1...66), or vice-versa for high and low numbers. All participants started with a random-
23 number block, followed by two oddball blocks. The order of oddball-up and oddball-down blocks

1 was counterbalanced between participants. Each block consisted of 64 trials, resulting in a
2 total of 192 trials for the whole experiment.

3 *2.2 Data analysis*

4 Our main dependent variable was the location of the click on the response bar. This was coded
5 as the relative position on the response bar, with 0 corresponding to a click on the left-most end
6 of the response bar and 1 to a click on the right-most end. Responses given after less than 200 ms
7 were eliminated. Outliers were eliminated according to the following method: For each presented
8 magnitude x we linearly interpolated the ‘expected’ response based on responses to stimuli $[x-10,$
9 $x+10]$ (truncated for responses near the top or bottom end of the stimulus range). We determined
10 the SD of the residuals of the linear interpolation for all responses excluding the response to x .
11 Responses greater than 3 SD removed from the ‘expected’ response for considered outliers. In
12 experiment 1, this applied to 28 trials (2.4% of all trials).

13 Responses, aggregated by numeric values, were fitted to three models: A linear function ($y = a$
14 $+ b*x$), a logarithmic function ($y = a + b*\log(x)$), and a power function ($y = a*x^b$), with x being
15 the numeric value of the stimuli and y the response of the participants. Note that the intercept for
16 the power function was fixed at 0, since (a) we wanted to fit the same number of parameters in all
17 models, and (b) this form is the classic power function common in research on human perception
18 (Stevens, 1957; Teghtsoonian, 1965).

19 Coefficients for the models were fit for each participant. The best-fitting model type, as
20 indicated by the lowest Akaike information criterion (AIC; Akaike, 1974; see Burnham &
21 Anderson, 2004, for guidelines on the interpretation of AIC and Δ AIC), was selected. This model
22 was then used to investigate if previous-trial effects would explain more variance in the data. To
23 do this, we fit a linear, logarithmic or power model to the trial—by-trial data, and compared the

1 fit of this simple model to a model that included the numeric magnitude in the previous trial
2 as an additional predictor. We also conducted a repeated-measures ANOVA with relative error
3 (defined as: $(response(x)-x)/x$) as the dependent variable and the 3-level factor *oddball*
4 ('up'/'down'/'no oddball') to investigate if oddball-trials displayed a systematically different
5 error to other trials, that is, whether responses would be affected by preceding trials. Greenhouse-
6 Geisser correction (Greenhouse & Geisser, 1959) was applied in all situations where sphericity
7 could be violated. Bonferroni-Holm correction was applied to all t-test (Holm, 1979).

8 *2.3 Results and Discussion*

9 We found that the data were fit best by a linear model of $y = 1.07x - 7.90$, explaining 99% of
10 the variance. Including the magnitude presented in the previous trial created a slightly better fit
11 ($\Delta AIC = -2.71$), with the predictor having a negative weight ($b = -0.0139$). Detailed descriptions
12 of all models can be seen in in Table 2, visualisations of the models can be seen in Figure 2.

13 Our ANOVA revealed a main-effect of the factor *oddball* ($F(2, 10) = 28.01, p_{GG} < .001, e_{GG} =$
14 $.83$), although post-hoc t-tests (comparing relative error in oddball-trials to relative error in all
15 range-matched non-oddball trials) indicated that while descriptively, oddball trials where the
16 oddball was smaller produced a smaller response than non-oddball trials and vice-versa for
17 oddball-trials 'upwards', these were not statistically significant differences ($t(5) = -1.228, p =$
18 $.274$ and $t(5) = 1.153, p = .301$, respectively).

19 The main goal of experiment 1 was to ascertain that the mapping of numbers to our response
20 bar was relatively accurate. This was indeed the case: The mapping was almost perfectly linear,
21 and the slope of the model was close to 1. Thus, we felt confident using this condition as a
22 control in the following experiments. With regards to previous-trial effects, a positive coefficient
23 in the model for previous magnitude indicates that relatively larger previous trials would lead to

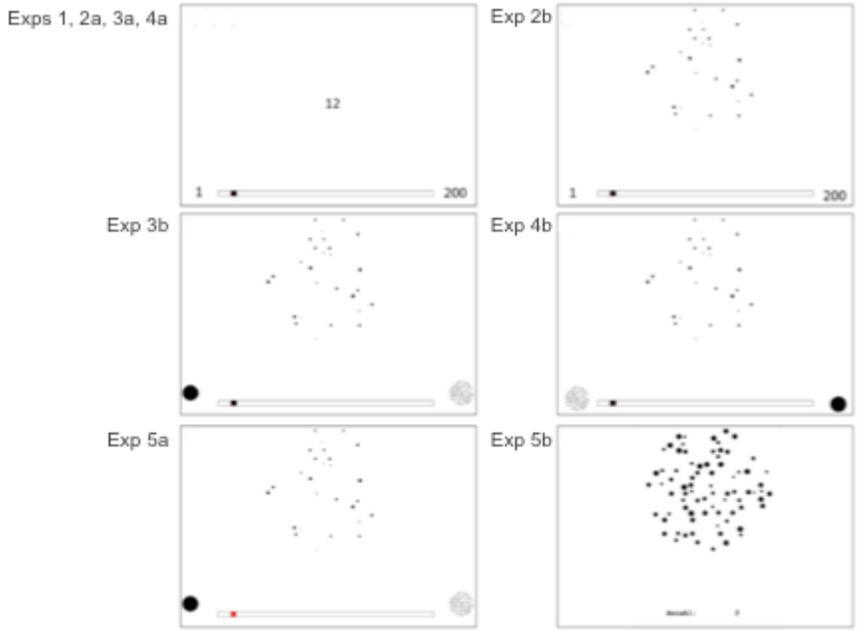
1 somewhat larger responses, while the analysis of oddball trials hinted towards the opposite
2 pattern. These were exploratory analyses, however, and we aimed to test this in the experiments
3 that followed.

4 **3 Experiment 2, 3, and 4: Nonsymbolic magnitude and its relation to symbolic magnitude**

5 Experiments 2, 3, and 4 incorporated not only symbolic, but also nonsymbolic magnitudes as
6 stimuli. That is, half of the trials consisted of participants being presented with Arabic digits and
7 clicking on the respective position on the response bar, the other half consisted of the same task
8 but with participants being presented with clouds of dots instead of digits. They were instructed
9 to assess the number of dots in these clouds, and select the appropriate location on the response
10 bar in the same way as with Arabic digits.

11 The experiments differed only in subtle, but nevertheless important details (see Figure 1, Table
12 1). Experiment 2 employed a response bar with numerical endpoints, for both symbolic and
13 nonsymbolic stimuli. These were presented in a blocked design that included both blocks of one
14 notation only and mixed blocks containing both notations. This gave us a first idea if our method
15 was appropriate for nonsymbolic stimuli, and whether our results would be in line with the
16 literature. In experiment 3, we employed a response bar with endpoints defined by nonsymbolic
17 numerosities for nonsymbolic stimuli (so that the mapping from nonsymbolic magnitude to
18 output did not involve a symbolic notation). In Experiment 4, we used the design of experiment
19 3, but flipped around the response bar in half of the trials, such that the upper end would now be
20 on the left side. We also randomised the starting position of the adjustment block on the response
21 bar, which had previously always been in the middle, to prevent participants from learning to
22 execute movements rather than performing an estimate for each trial. Each participant took part

1 in only one of the experiments 1...4. The purpose of these experiments was to investigate
 2 whether (a) the typical MNL shape could reliably be found in number line tasks with symbolic
 3 and nonsymbolic input, even when controlling for possible confounds, and (b) whether there
 4 would be previous-trial effects within, or even between trial-types.



5
 6 **Figure 1:** Schematic depiction of our experiments. Top left: Experiment 1. Top right: Experiment 2,
 7 nonsymbolic stimuli. Middle left: Experiment 3, nonsymbolic stimuli. Middle right: Experiment 4,
 8 nonsymbolic stimuli. Bottom: Experiment 5, response bar (left) and number-response (right). The red
 9 block indicates the feedback that was given after a trial.

10 *3.1 Experiment 2: Symbolic and nonsymbolic magnitude on a standard response bar*

11 In experiment 2, we introduced nonsymbolic stimuli in addition to symbolic stimuli (numbers).
 12 Thus, we were able to compare the responses in the same task – that is, responses on a ‘ruler-like’
 13 response bar – for symbolic and nonsymbolic stimuli.

1 3.1.1 Methods

2 3.1.1.1 Participants

3 Eight participants (students of University of Hamburg, aged between 19 and 26 years, mean age
4 = 22.9; 6 female) were tested. Each participants received course credit or 8€/hr.

5 3.1.1.2 Apparatus

6 The same setup as in experiment 1 was used. The main difference was that we employed not
7 only symbolic, but also nonsymbolic stimuli. These non-symbolic stimuli were generated using a
8 modified version of a program developed by Gebuis and Reynvoet (2011) that will generate
9 clouds of dots and was designed to keep visual stimulus properties uninformative about the
10 number of dots in a certain design. In our design, keeping visual dimensions completely
11 uninformative about number would have been impossible, since nonsymbolic magnitude is
12 ultimately defined by visual features, and all of our stimuli differed in magnitude. Thus, we
13 settled for a compromise in which the visual features total area of the clouds of dots ($r = .41$),
14 density ($r = .28$), surface are of the dots ($r = .57$) and circumference of the cloud ($r = .77$) were
15 all imperfectly correlated with nonsymbolic magnitude.

16 3.1.1.3 Procedure

17 As in experiment 1, participants indicated the position of a symbolic or nonsymbolic stimulus
18 on a response bar presented horizontally between the numbers 1 and 200 (in all tables and
19 figures, the symbolic condition is indicated as Exp. 2a and the non-symbolic as Exp. 2b; the same
20 nomenclature is used for experiments 3 and 4). Symbolic and non-symbolic stimuli were
21 presented in a counterbalanced blocked design that contained *single-type, random magnitude*
22 blocks, *single-type, oddball* blocks, and *mixed-type, oddball* blocks. Blocks were randomised in
23 the same manner as in experiment 1, with the order of symbolic, nonsymbolic and mixed blocks

1 counterbalanced between participants. The random-number block was always the first for
2 each stimulus type, and the mixed blocks were always at the end of the experiment. In total,
3 participants completed 11 blocks of 64 trials each, for 704 trials overall.

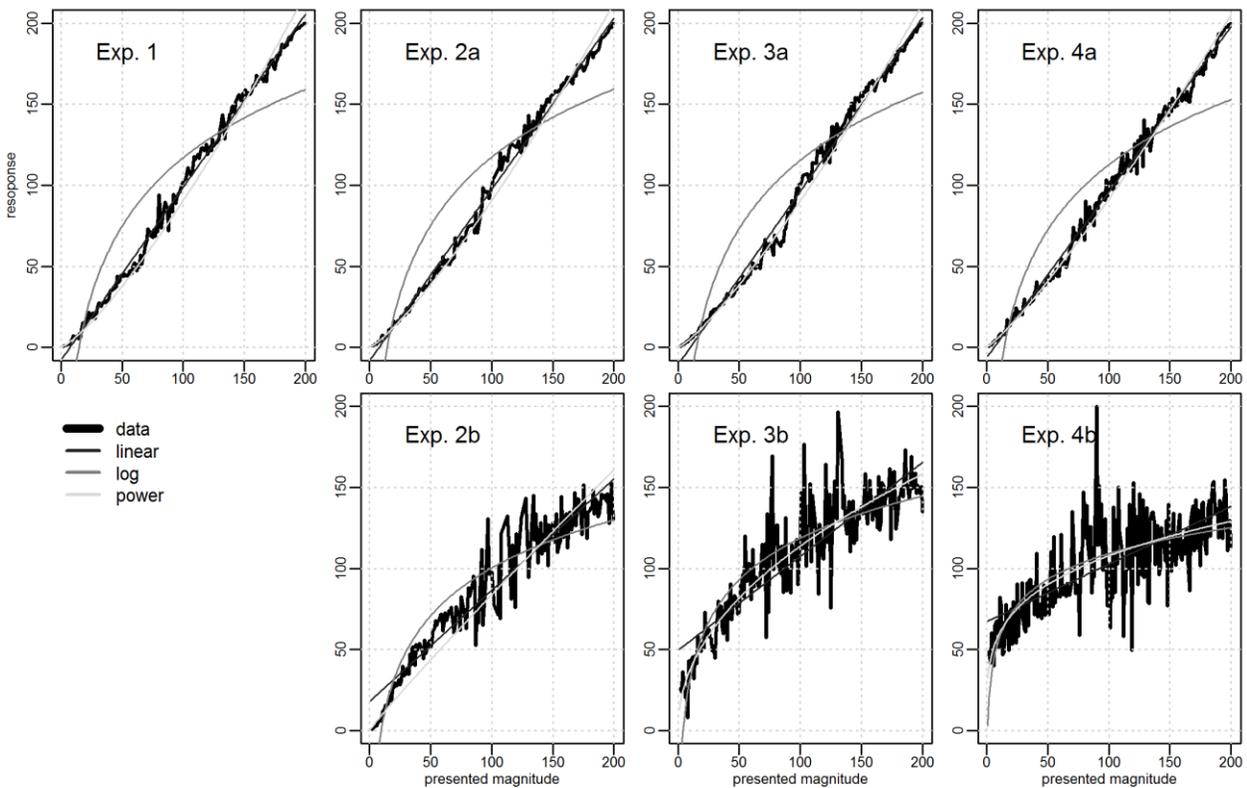
4 *3.1.2 Data analysis*

5 Analyses mirrored those from experiment 1, with the addition of stimulus type as another
6 independent variable. Because of this, we fit the same three models (linear, logarithmic, power)
7 to responses to each of the stimulus types (symbolic and nonsymbolic). To be able to investigate
8 previous-trial effects between stimulus types, we added an interaction term to the previous-trial
9 model that allowed for a differential effect of ‘same type’ or ‘different type’ previous trials. We
10 also included ‘stimulus type’ as an additional 2-level factor in the ANOVA, and conducted
11 separate post-hoc t-tests for each type when testing for oddball-effects.

12 *3.1.3 Results & Discussion*

13 A total of 114 trials (2%) had to be removed, as participants had not given an answer. A further
14 83 trials (1.5%) were excluded as outliers (see section 2.2). Of the remaining data, trials that
15 contained symbolic stimuli and trials that contained nonsymbolic stimuli were each separately
16 fitted to three models (linear, logarithmic, power model; see Figure 2). For the symbolic stimuli,
17 the best fit was again a linear model ($y = 1.06x - 8.23$; $R^2 = .99$). The same was true for the
18 nonsymbolic stimuli, although the model was markedly different ($y = 0.69x + 17.88$) and the fit
19 was not as good ($R^2 = .91$). The data were fitted better when the model included the previous
20 magnitude as a predictor with a positive weight ($\Delta AIC = -7.65$, $b = 0.0344$). Introducing an
21 additional interaction between previous trial magnitude and previous trial type improved the fit
22 marginally ($\Delta AIC = -2.01$) and revealed that the weight for previous trials was somewhat smaller
23 when the trial type was different to the current trial ($b_{diff} = 0.0041$, $b_{same} = 0.0380$).

1 A 2 (*trial-type*: symbolic or nonsymbolic) * 2 (*block-type*: mixed or homogenous) * 3
 2 (*oddball*) factor repeated-measures ANOVA on relative error indicated only one interaction:
 3 *Trial-type* * *oddball* ($F(2, 14) = 19.26, p_{GG} = .003, e_{GG} = .53$). All other effects were non-
 4 significant ($ps > .12$). Post-hoc t-tests gave only tentative evidence, as only comparing ‘upwards’
 5 oddballs with regular trials gave some indication of an effect ($t(7) = -3.82, p = .007$; all other $ps >$
 6 $.2$).



7
 8 **Figure 2:** Data from experiments 1...4, compared to linear, logarithmic, and power models. Left to right:
 9 Experiments 1...4. Top row: Symbolic stimuli. Bottom row: Nonsymbolic stimuli. For details on the
 10 experiments, see Table 1.

11 To summarise, our results from experiment 2 also showed that response-bar responses to
 12 symbolic magnitudes (numbers) were almost perfectly linear and veridical. Responses to
 13 nonsymbolic magnitudes showed the characteristic underestimation for relatively large numbers,
 14 but interestingly still were fit better by a linear function than a logarithmic or power function.

1 Additionally, larger magnitudes displayed in previous trials seemed to correlate with slightly
2 larger responses.

3 *3.2 Experiment 3: Symbolic and nonsymbolic magnitude on a response bar with nonsymbolic* 4 *endpoints*

5 A potential drawback of experiment 2 was its use of numeric endpoints, which means that one
6 could argue that the output measure was in fact not really nonsymbolic. To remedy this, we
7 conducted experiment 3, in which we used nonsymbolic magnitudes as endpoints to the response
8 bar. In all other respects the experiment was identical to Experiment 2.

9 *3.2.1 Methods*

10 Again, we recruited eight participants from the same pool as in experiment 2 (aged between 20
11 and 29 years, mean age = 23.9; 6 female). As in experiment 2, all participants indicated the
12 position of a symbolic or nonsymbolic stimulus on a response bar. The only difference to
13 Experiment 2 being that the endpoints of the response bar for the nonsymbolic stimulus now were
14 a single dot on the left and a cloud of 200 dots on the right instead of Arabic numbers.

15 *3.2.2 Results & Discussion*

16 We removed 76 trials (1.3%) because of a lack of a valid answer, and 73 trials (1.3%) as
17 outliers. The remaining data were modelled in the same fashion as in experiment 2, indicating a
18 linear model as the best, and once again near perfect, fit for symbolic responses ($y = 1.07x -$
19 10.74 ; $R^2 = .99$). Responses to nonsymbolic stimuli were fit best by a power model ($y = 12.66 *$
20 $x^{0.48}$), which explained 82% of the variance. The data were not fit better when previous number
21 was included as a predictor ($\Delta AIC = 0.71$, $b = 0.0183$), but slightly better when accounting for
22 previous number split up by previous trial type ($\Delta AIC = -1.75$, $b_{diff} = -0.0273$, $b_{same} = 0.0241$).
23 The usual $2*2*3$ ANOVA on relative error revealed no interactions between factors (all $p > .13$)

1 and no main effects, either (all $p > .16$). However, when oddball trials were tested against
2 magnitude-matched non-oddball trials, this showed a significant difference between upwards
3 oddball trials with nonsymbolic stimuli ($t(7) = -3.92, p = .006$), while other differences were not
4 significant when correcting for multiple comparisons (downwards oddballs, nonsymbolic: $t(7) =$
5 $2.55, p = .038$; upwards symbolic: $t(7) = -2.61, p = .035$; downwards symbolic: $t(7) = 0.86, p =$
6 $.421$), although they all pointed in the same direction: Oddball trials tended to err more towards
7 the middle than other trials in the same number range, consistent with the fact that previous trials
8 had a positive weight in the model.

9 Again, we found a virtually veridical response function to symbolic magnitudes, with a notable
10 underestimation and previous-trial dependency present for responses to nonsymbolic magnitudes.
11 This was fit best by a logarithmic model. Responses to nonsymbolic stimuli were generally not
12 predicted as well by the actual magnitude presented as in experiment 2 (see Table 2), perhaps
13 indicating that the response-bar task was more difficult with nonsymbolic endpoints. Still, the
14 response function was quite similar to the response function in experiment 2b (see Figure 2).

15 *3.3 Experiment 4: Symbolic and nonsymbolic magnitude with left/right flipped endpoints*

16 In experiment 4, we wanted to preclude that what participants had been giving were stereotyped
17 responses. To prevent them from using such a strategy, we slightly increased the difficulty of the
18 task and introduced another input-response mapping by (a) flipping randomly the response bar in
19 half of the trials, thus displaying the largest magnitude on its left side and the smallest magnitude
20 on the right (b) randomly varying the starting position of the adjustment block in each trial.

21 *3.3.1 Methods*

22 Eight participants from the same pool as in experiments 2 and 3 (aged between 20 and 32 years,
23 mean age = 25.7; 4 female) took part in the experiment. The task was mostly the same as in

1 experiment 3, but in 50% of the trials (randomly distributed within each block), the response
2 bar was flipped, such that the lower end was on the right and the higher end was on the left.
3 Additionally, the starting position for the adjustment block was randomized, such that the block
4 was equally likely to appear anywhere on the response bar at the start of each trial.

5 3.3.2 Results & Discussion

6 We had to remove 38 trials (0.7%) for a lacking valid answer, and 102 trials (1.8%) as outliers.
7 Modelling the responses to symbolic stimuli, the best model was a linear fit of $y = 1.02x - 6.26$,
8 explaining 99% of the variance (Figure 2). The nonsymbolic responses were once again fit best
9 by a power model ($y = 32.50 * x^{0.26}$; $R^2 = .54$). Including the magnitude of the previous trial did
10 not improve the fit ($\Delta AIC = 0.99$), although including an interaction term did (ΔAIC model with
11 interaction vs. simple model: -5.19), indicating that previous trials actually had a negative weight
12 if they were of a different type ($b_{diff} = -0.0777$), and much smaller negative weight when they
13 were of the same type ($b_{same} = -0.0061$). The standard $2*2*3$ repeated-measures ANOVA on
14 relative error revealed a main-effect of *oddball* ($F(2, 14) = 20.20$, $p_{GG} = .003$, $e_{GG} = .51$), but no
15 other main effect (all $p > .6$), with statistically significant interaction of *oddball*trial-type* ($F(1,$
16 $7) = 7.23$, $p = .031$) and the three-way interaction ($F(2, 14) = 7.89$, $p_{GG} = .025$, $e_{GG} = .51$). No t-
17 test comparing oddball-trials to range-matched regular trials indicated any significant difference
18 ($ps > .6$).

19 As was found in previous experiments, responses to symbolic stimuli were fit best by a linear
20 function with a slope close to 1, while responses to nonsymbolic stimuli resembled a logarithmic
21 function. This strengthens the results from experiments 2 and 3, as the results were similar even
22 when the experiment was designed to prevent participants from learning mouse movements as
23 opposed to considering the desired location of the click.

1 **Table 2:** Linear, logarithmic and power models fit to data from all experiments.

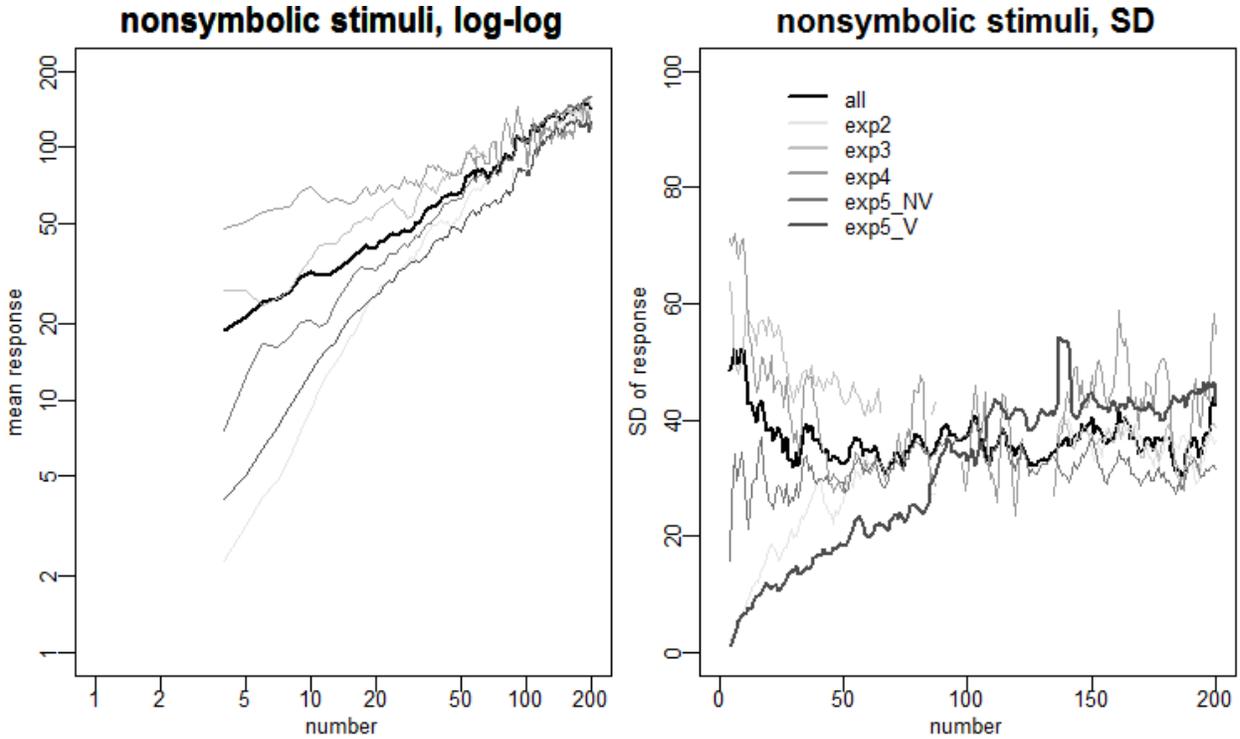
<i>Exp</i>	<i>Condition</i>	<i>lin</i>	<i>log</i>	<i>power</i>	<i>Previous trial weight</i>	<i>Previous trial weight by type</i>
1	Symbolic	<i>1.07x-7.90</i>	60.85*log(x) – 163.28	0.27*x ^{1.27}	<i>B = -0.0139</i>	-
		<i>R² = .99</i>	R ² = .80	R ² = .98	<i>ΔAIC = -2.71</i>	
2	Symbolic	<i>1.06x-8.23</i>	60.23*log(x) – 161.93	0.32*x ^{1.23}	B = 0.0344	<i>B_{same} = 0.0380</i>
		<i>R² = .99</i>	R ² = .81	R ² = .99		<i>B_{diff} = 0.0041</i>
	Nonsymbolic	<i>0.69x+17.88</i>	42.67*log(x) – 96.12	1.10*x ^{0.94}	ΔAIC = -7.65	
		<i>R² = .91</i>	R ² = .87	R ² = .89		<i>ΔAIC = -9.66</i>
3	Symbolic	<i>1.07x-10.74</i>	60.15*log(x)-163.82	0.31*x ^{1.23}	B = 0.0183	<i>B_{same} = 0.241</i>
		<i>R² = .99</i>	R ² = .79	R ² = .99		<i>B_{diff} = -0.0273</i>
	Nonsymbolic	<i>0.58x+49.50</i>	37.34*log(x)-53.17	12.66*x ^{0.48}	ΔAIC = 0.71	
		<i>R² = .77</i>	R ² = .80	R ² = .82		<i>ΔAIC = -1.75</i>
4	Symbolic	<i>1.02x-6.26</i>	58.02*log(x)-154.57	0.43*x ^{1.16}	B = -0.0193	<i>B_{same} = -0.0061</i>
		<i>R² = .99</i>	R ² = .79	R ² = .99	ΔAIC = 0.99	<i>B_{diff} = -0.0777</i>
	Nonsymbolic	<i>0.36x+67.13</i>	23.04*log(x)+3.09	32.50*x ^{0.26}		
		<i>R² = .50</i>	R ² = .52	R ² = .54		<i>ΔAIC = -5.19</i>
5	response bar, pre-FB	<i>0.65x+32.13</i>	41.03*log(x)-80.84	4.69*x ^{0.66}	<i>B = 0.0424</i>	-
		<i>R² = .84</i>	R ² = .78	R ² = .85	<i>ΔAIC = -11.43</i>	
	response bar, FB	<i>0.82x + 17.59</i>	59.92*log(x)-163.15	2.97*x ^{0.77}	<i>B = 0.2398</i>	
		<i>R² = .93</i>	R ² = .88	R ² = .94	<i>ΔAIC = -314.70</i>	
	response bar, post-FB	<i>0.79x+24.85</i>	49.29*log(x)-109.54	1.56*x ^{0.91}	<i>B = 0.0440</i>	
		<i>R² = .93</i>	R ² = .91	R ² = .91	<i>ΔAIC = -26.86</i>	
	Number-response, pre-FB	<i>0.50x+16.13</i>	30.91*log(x)-67.22	1.91*x ^{0.77}	<i>B = 0.0579</i>	
		<i>R² = .88</i>	R ² = .83	R ² = .89	<i>ΔAIC = -35.32</i>	
	Number response, FB	<i>0.84x +8.59</i>	59.77*log(x)-170.18	2.10*x ^{0.83}	<i>B = 0.3258</i>	
		<i>R² = .94</i>	R ² = .86	R ² = .94	<i>ΔAIC = -463.97</i>	
	Number response, post-FB	<i>0.64x+21.63</i>	40.20*log(x)-87.54	2.10*x ^{0.81}	<i>B = 0.0648</i>	
		<i>R² = .89</i>	R ² = .86	R ² = .90	<i>ΔAIC = -39.36</i>	

2 *Note:* Italics indicate the best-fitting model. ΔAIC given relative to the simplest model. FB: feedback, AIC:
 3 Akaike information criterion (Akaike, 1974; Burnham & Anderson, 2004).

4

1 *3.4 Discussion of Experiments 2-4*

2 Two findings appeared robustly in all experiments: Responses to symbolic stimuli had an
3 almost perfectly linear shape, and responses to nonsymbolic stimuli tended to overestimate low
4 magnitudes and underestimate higher magnitudes. The former is readily explained by the
5 proficiency of participants: Despite having limited time available, the symbolic magnitude task
6 was overall not very difficult. Responses to nonsymbolic stimuli, on the other hand, mirror
7 known patterns of nonlinearity (Dehaene, 2003; van Oeffelen & Vos, 1982) that is robust across
8 different variants of the task. Notably, the responses were in all cases fit better by a power
9 function than by a logarithmic function (see Table 2), which is corroborated by the fact that when
10 plotted in a log-log graph, the functions appear roughly (Figure 3). These results are also quite
11 compatible with Dehaene's notion of a logarithmic number line (Dehaene, 1992) with an 'output
12 grid' (Izard & Dehaene, 2008) transformation. It is also compatible with a dynamic encoding
13 mechanism as proposed by Cicchini et al. (2014). Notions such as a linear response function with
14 scalar variability, however, can be dismissed; while the variability does reach a ceiling of sort in
15 response-bar tasks (Figure 3), this model would predict underestimation for higher numbers, but
16 not overestimation for smaller numbers.



1
 2 **Figure 3:** Exploring variability in responses to nonsymbolic stimuli. Left: Responses in a log-log plot. A
 3 power function would be linear in such a plot. Right: Standard deviation by number, for all experiments.
 4 For the explanation of dynamic encoding mechanisms based on previous stimuli to be plausible,
 5 it is a prerequisite that sequential dependencies exist, which was the second question our
 6 experiments sought to answer. Indeed, this was the case, as the fitting of previous-trial models
 7 found a positive relationship between the magnitude of the previous trial and the response in the
 8 current trial, which was consistent with the analysis of oddball trials.

9 **4 Experiment 5: Does calibration of symbolic responses calibrate nonsymbolic responses?**

10 In our final experiment, we sought to explore whether the responses to nonsymbolic stimuli
 11 could be linearized through calibration, and if this could be done for symbolic and nonsymbolic
 12 responses independently. This served the broader purpose of testing if the observed shape of the
 13 responses is better understood as a phenomenon of encoding or of representation-to-response
 14 mapping by comparing sequential effects between stimulus types to calibration effects between

1 tasks. To this end, we conducted an experiment in which participants conducted both a
2 response-bar task like the one described in experiment 3 but with randomised starting positions,
3 and a classic numerosity-judgement task in which they typed in the estimated number of dots for
4 each stimulus. Both tasks included feedback blocks to allow participants to calibrate their
5 responses, with feedback being either (a) veridical, (b) systematically lower, or (c) systematically
6 higher than the actual magnitude.

7 *4.1 Methods*

8 *4.1.1 Participants*

9 Due to the larger number of conditions, we increased the number of participants to a total of 36
10 (mean age = 24 years, age range 18 to 36; 26 female) from the same participant pool as in the
11 experiments 2-4.

12 *4.1.2 Procedure*

13 All participants completed two numerosity-estimation tasks: A response-bar task like the one
14 used in experiment 3, and a simple number-response task, in which participants entered their
15 estimate via a standard computer keyboard. Stimuli were the same nonsymbolic stimuli as in
16 experiments 2, 3, and 4. Participants were allowed 3 seconds to respond in the response-bar task
17 and 4 seconds in the number-response task, as typing the answer was typically somewhat slower
18 than clicking.

19 For both the response-bar task and the number-response task, we included a block where
20 participants received feedback about the correctness of their response. This feedback could be
21 either veridical (reflecting a 1-to-1 mapping of stimulus to the position on the response bar or the
22 number to be entered), or distorted by an amount of 15. This distortion (or lack thereof) was
23 consistent across all trials within one block, but varied independently for the two tasks, and

1 distortions were counterbalanced between participants. The order of tasks was also
2 counterbalanced. Feedback in the number-response task was presented as a red number appearing
3 on the screen once the participant hit the *return* key, to the right of the number the participant had
4 just entered. In the response-bar task, feedback was given through a red square block of the same
5 size as the adjustment block, appearing on the horizontal bar at the location corresponding to the
6 correct magnitude (or the correct magnitude +15 / -15). Participants received 10 practice trials
7 with feedback on a response bar right before the response-bar task, along with instructions on
8 how the feedback worked. In the -15 and +15 feedback blocks, stimuli were restricted so that
9 feedback fell in the 15...85 and 115...185 ranges, such that feedback was never too close to the
10 bounds of the response bar, or on the ‘wrong’ side of the mid-point, so as to not make the
11 manipulation too obvious (see Barth & Paladino, 2011). No oddball blocks were included. Each
12 block consisted of 120 trials, resulting in a total of 720 trials per participant.

13 *4.2 Results*

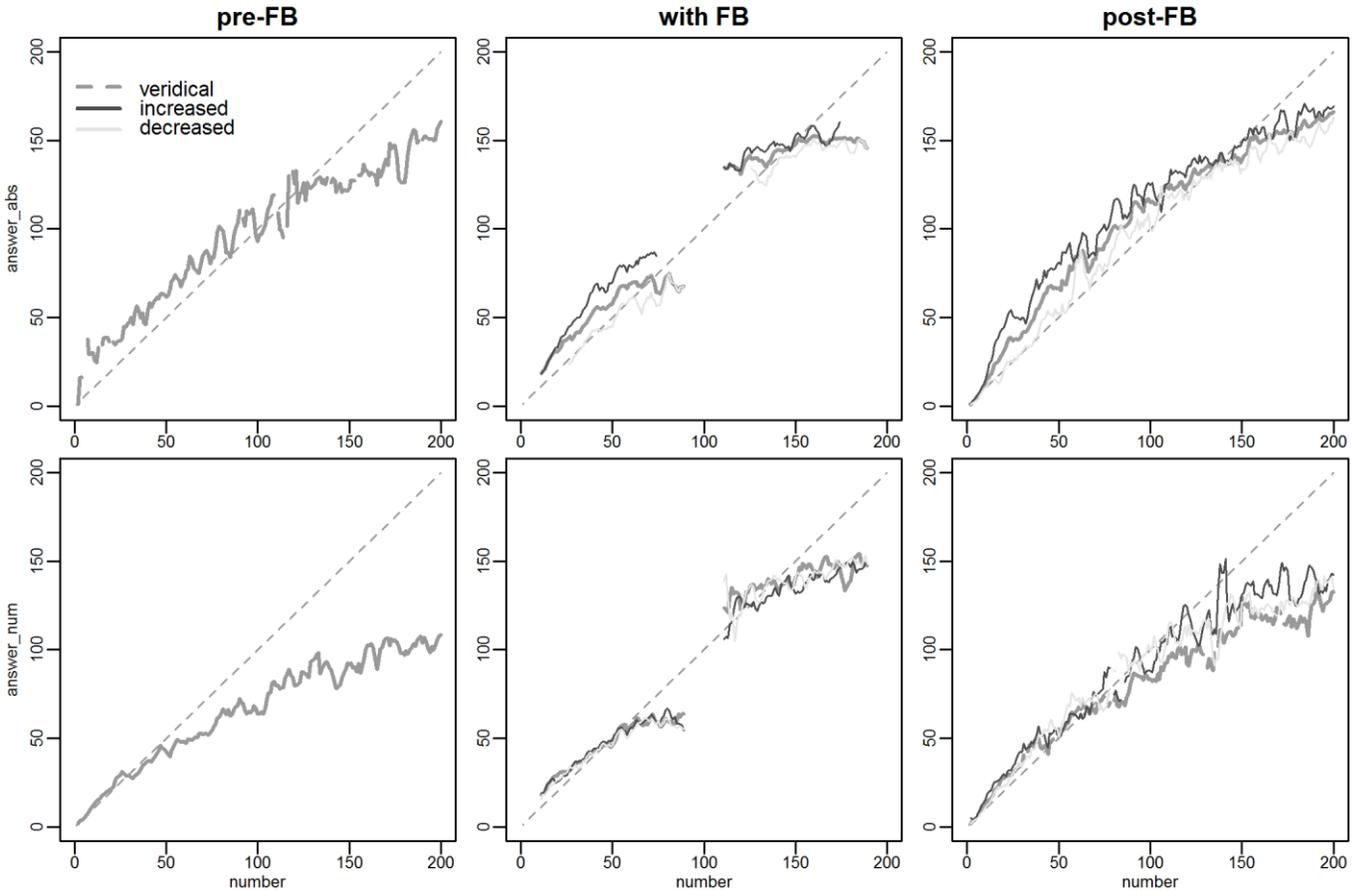
14 Over both tasks combined, 859 trials had to be removed because no valid answer had been
15 given (3.3%). A further 330 trials (1.3%) were excluded as outliers. One participant had to be
16 removed from analysis for not understanding the task. During debriefing, all participants were
17 asked if they had considered the feedback to be accurate. Six out of 36 participants said that they
18 had not, which included one participant who had received veridical feedback. Removing these
19 participants from analysis did not substantially change the results (the following analyses include
20 those participants).

21 Similar to the analysis of the other experiments, we first investigated how to best model the data
22 from each task. We fit separate models for responses given in blocks prior to feedback, during
23 feedback, and after feedback had been presented. In the number-line task, pre-feedback data were

1 fitted best by a power model ($y = 4.69 * x^{0.66}$; $R^2 = .85$), which fit marginally better than a
2 linear model ($R^2 = .85$ to $.84$), with post-feedback data being better fit linearly ($y = 0.79x +$
3 24.85 ; $R^2 = .93$). As was the case in most previous tasks, the data were fit better when including
4 previous trial magnitude as predictor, both before ($\Delta AIC = -11.43$; $b = 0.0424$) and after
5 feedback ($\Delta AIC = -26.86$; $b = 0.0440$). During feedback, a linear function gave the best fit ($y =$
6 $2.97 * x^{0.77}$; $R^2 = .94$), and unsurprisingly there was a very strong previous-trial effect ($\Delta AIC = -$
7 314.70 ; $b = 0.2399$). In all three phases, the differences in goodness-of-fit between the power
8 model and the linear model were marginal, with the logarithmic function doing substantially
9 worse, see Table 2. For the number-response task, the data were fit best by a power function, in
10 both the pre-feedback ($y = 1.91 * x^{0.77}$; $R^2 = .89$) and post-feedback blocks ($y = 2.10 * x^{0.81}$;
11 $R^2 = .90$). Including the previous trial also improved the fit, pre-feedback ($\Delta AIC = -35.32$, $b =$
12 0.0579) and post-feedback ($\Delta AIC = -39.36$, $b = 0.0648$). With feedback, the linear model did best
13 ($y = 8.59 + 0.84x$, $R^2 = .94$), and showed a very strong effect of previous trials ($\Delta AIC = -463.97$;
14 $b = 0.3258$). Again, model fits were virtually equally good for linear and power functions, but
15 worse for logarithmic fits.

16 To investigate further the effects of feedback on relative error, and to see if the feedback in the
17 respective other task mattered at all, we further conducted a mixed ANOVA with between-factors
18 *symbolic feedback* and *nonsymbolic feedback* (4 levels each: +15, -15, 0, none), as well as the
19 within-participant factor *trial-type* ('response-bar response' or 'number response').
20 Unsurprisingly, this revealed a main effect of *trial-type* ($F(1, 27) = 33.24$, $p < .001$) but, perhaps
21 surprisingly, no interaction of *trial-type* and either feedback factor (*trial-type*verbal feedback*:
22 $F(2, 27) = 1.35$, $p = .275$; *trial-type*nonverbal feedback*: $F(2, 27) = 0.30$, $p = .746$), which would
23 have indicated a general impact of feedback on the response. Post-hoc two-sample t-tests
24 revealed that no response was influenced by feedback in the other task ($ps > .24$), but indeed also

1 that there was only very weak evidence, if any, for an effect of same-task feedback (number-
 2 line task, feedback +15 vs. 0: $t(20.20) = 0.92, p = .370$; feedback -15 vs. 0: $t(21.26) = -0.56, p =$
 3 $.580$; verbal task, feedback +15 vs. 0: $t(20.95) = 0.38, p = .706$; feedback -15 vs. 0: $t(15.94) = -$
 4 $2.57, p = .021$). This is supported by inspection, see Figure 4.



5
 6 **Figure 4:** Number lines from experiment 5, by feedback. Black: Veridical feedback; light grey: Feedback
 7 distorted by -15; dark grey: Feedback distorted by +15. Pre-feedback panels show data from all groups.
 8 Dashed lines depict identity. Top row: Response-bar task, bottom row: Typed number response.

9 *4.3 Discussion*

10 We conducted experiment 5 for two main purposes: To be able to compare responses obtained
 11 from different tasks with nonsymbolic stimuli, and to investigate the effect of different types of
 12 feedback on the shape of the response functions. The shape we found for the number-response
 13 tasks was indeed different from what was found in previous experiments, in that the best fitting

1 model was clearly a power function, as has been proposed by several authors (Izard &
2 Dehaene, 2008; Krueger, 1972; Nieder & Miller, 2003). We can also see (Figure 3) that this was
3 the only task where variability increased almost linearly with stimulus magnitude – a typical
4 feature of magnitude estimation. In the response-bar task, we found a similar, albeit somewhat
5 steeper and more linear response function than in previous experiments, even before any
6 feedback had been given. It is possible that the added practice trials had an effect here, although
7 another possibility is that the lack of a nonsymbolic response-bar task played a role.

8 With regards to feedback, we can see that response functions were markedly steeper after
9 feedback was given. Feedback linearized the responses somewhat, although the response function
10 was still quite far from a veridical function, and the effect did not appear to be permanent. Still,
11 response-bar responses were also fit better after feedback was given. This was not true of number
12 responses. . Importantly, feedback effects did not transfer between tasks, indicating that any
13 calibration was task-specific as opposed to a more general calibration.

14 **5 General discussion**

15 Two classic findings were reproduced in our experiments: response-bar responses to symbolic
16 magnitude stimuli exhibited a linear shape in adult participants (Anobile et al., 2012; Siegler &
17 Opfer, 2003), and typed responses to nonsymbolic magnitudes were fit best by a power function
18 (Izard & Dehaene, 2008; Krueger, 1972; Nieder & Miller, 2003). Still, it should be noted that the
19 predictions of these models were remarkably similar, as exponents of the power functions tended
20 to be close to 1 and intercepts of the linear functions close to 0. We also investigated the
21 response-bar responses to nonsymbolic magnitudes, which were best fit by a logarithmic
22 function.

1 What remains to be explained is the cause of the shape of these responses. Our experiments
2 (see Table 1, Figure 1) were designed to investigate three questions: Namely, whether responses
3 would depend on previous trials and whether this could explain the response function shape,
4 whether response functions for different input (in the same task) and output (with the same input
5 type) would differ, and whether giving feedback to calibrate the response would linearize the
6 response function.

7 With regard to the first question of whether these were dynamic effects brought about by the
8 effects of previous trials, we found the strongest effects to be mostly static: While we found
9 previous-trial effects, these were not strong enough to explain much of the variance – and,
10 importantly, not robust to variations (such as flipping the response bar or randomising the starting
11 position in experiment 4) that the shape of the response function was robust to. It is interesting,
12 however, that the previous-trial effect we found virtually disappeared when the current and
13 previous trial were of different types. This touches on another debate: The question whether a
14 single semantic representation is underlying the processing of magnitudes of different kinds
15 (Dehaene, 1992; Walsh, 2003), or whether some magnitudes may be explained as sensory
16 features (Arrighi, Togoli, & Burr, 2014). Our data not speak strongly against either hypothesis,
17 although a strong version of an underlying magnitude representation would probably predict less
18 distinct interaction patterns between trials of different and the same type.

19 Regarding the second question (telling apart the different roles of encoding and output mapping
20 in creating the typical shape of response functions): While a lot of research has focussed on the
21 different properties of symbolic and nonsymbolic magnitude processing, the contribution of
22 output format has not been investigated as much beyond the question of methodological
23 confounds (Barth & Paladino, 2011; Cohen & Blanc-Goldhammer, 2011). Although using a

1 response-bar task may have some problems, it is also quite clear that it provides the
2 possibility of a useful, very direct instrument of measuring responses to symbolic and
3 nonsymbolic stimuli in a comparable way. In this study, we have laid out some key differences
4 between response-bar task and verbal or typed number-response tasks, finding that both the mean
5 responses (see Figure 2) and the measured variability (Figure 3) differ between the two.
6 However, we also found that effects of input type are much more pronounced than effects of the
7 output measure (compare top and bottom row, Figure 2).

8 We also used the two different tasks to investigate whether calibrating one output measure
9 would have any effect on responses in another output measure, which we did in experiment 5,
10 thus trying to answer the third question: Can we linearize responses to nonsymbolic magnitudes
11 through calibration, and if so, what do we calibrate? We found some, but not very strong
12 linearization (Figure 4), and no impact of calibration from one task to the other task. As this was
13 not the case, we can quite confidently draw the conclusion that what is calibrated is not the input
14 mapping from stimulus to internal representation, but the output mapping from internal
15 representation to response. We also see that this is not sufficient to completely linearize the
16 response. Our data also allow us the conclusion that the nonlinear shape is arguably a function of
17 encoding mechanisms, as it is seen in both tasks, and we have demonstrated that the response-bar
18 task in itself does not lead to a nonlinear response function.

19 We should also offer a word of caution on our stimuli. It should be noted that, as mentioned in
20 section 3.1, several sensory stimulus features were somewhat informative about the numerosities
21 presented. However, we do not believe this to be problematic, as none of these explained more
22 than 61% of the variance, whereas all but one of our models explain substantially more variance
23 (see Table 2). Of course, we cannot rule out the possibility that a combination of sensory cues

1 might have been used (see e.g. Gebuis, Gevers, & Cohen Kadosh, 2014). However, the result
2 of such a combination of cues would be a sort of nonsymbolic magnitude, and while the
3 discussion about how to define nonsymbolic magnitude is an interesting one, it is beyond the
4 scope of this paper.

5 We conclude that the nonlinear shape of the number line is largely robust to calibration even
6 through direct, veridical feedback, and that features of both stimuli and output measures
7 contribute to it. Mechanisms seem relatively separate for different types of tasks, as well as
8 different types of magnitude representation. Serial dependencies exist between trials, but are too
9 weak to explain the shape of the responses in such tasks.

10 **Acknowledgements**

11 We thank Alistair Yousefi and Larissa Brockmann for their help collecting data. KKK was
12 supported by a PhD scholarship as part of a grant to VHF by the Deutsche
13 Forschungsgemeinschaft (DFG) within the international graduate research training group on
14 Cross-Model Interaction in Natural and Artificial Cognitive Systems (CINACS; DFG IKG-1247).

15

1 References

- 2 Akaike, H. (1974). A new look at the statistical model identification. In *IEEE Transactions on*
3 *Automatic Control AC-19* (pp. 716–723).
- 4 Anobile, G., Cicchini, G. M., & Burr, D. C. (2012). Linear mapping of numbers onto space
5 requires attention. *Cognition*, *122*(3), 454–459. doi:10.1016/j.cognition.2011.11.006
- 6 Arrighi, R., Togoli, I., & Burr, D. C. (2014). A generalized sense of number. *Proceedings of the*
7 *Royal Society B*, *281*(1797), 20141791. doi:10.1098/rspb.2014.1791
- 8 Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: evidence
9 against a representational shift. *Developmental Science*, *14*(1), 125–135. doi:10.1111/j.1467-
10 7687.2010.00962.x
- 11 Burnham, K. P., & Anderson, R. P. (2004). Multimodel Inference: Understanding AIC and BIC
12 in Model Selection. *Sociological Methods & Research*, *33*(2), 261–304.
13 doi:10.1177/0049124104268644
- 14 Cantlon, J. F., Cordes, S., Libertus, M. E., & Brannon, E. M. (2009). Comment on “Log or
15 linear? Distinct intuitions of the number scale in Western and Amazonian indigene
16 cultures.” *Science*, *323*, 38.
- 17 Cicchini, G. M., Anobile, G., & Burr, D. C. (2014). Compressive mapping of number to space
18 reflects dynamic encoding mechanisms, not static logarithmic transform. *Proceedings of the*
19 *National Academy of Sciences of the United States of America*, *111*(21), 7867–7872.
20 doi:10.1073/pnas.1402785111
- 21 Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded
22 number line tasks. *Psychonomic Bulletin & Review*, *18*(2), 331–338. doi:10.3758/s13423-
23 011-0059-z
- 24 Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish
25 verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin &*
26 *Review*, *8*(4), 698–707. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11848588>
- 27 Dehaene, S. (1992). Varieties of numerical abilities. *Cognition*, *44*, 1–42.
- 28 Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number
29 line. *Trends in Cognitive Sciences*, *7*(4), 145–147. doi:10.1016/S1364-6613(03)00055-X
- 30 Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number
31 magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396.
- 32 Dehaene, S., Izard, V., Pica, P., & Spelke, E. (2009). Response to comment on “Log or linear?
33 Distinct intuitions of the number scale in Western and Amazonian indigene cultures.”
34 *Science*, *323*, 38.
- 35 Dehaene, S., Izard, V., Spelke, E., & Pica, P. (2008). Log or Linear? Distinct Intuitions of the
36 Number Scale in Western Amazonian Indigene Cultures. *Science*, *320*, 1217–1220.
37 doi:10.1126/science.1156540
- 38 Fechner, G. T. (1860). *Elemente der Psychophysik*. Leipzig: Breitkopf und Haertel.

- 1 Gebuis, T., Gevers, W., & Cohen Kadosh, R. (2014). Topographic representation of high-
2 level cognition: Numerosity or sensory processing? *Trends in Cognitive Sciences*, *18*(1), 1-
3 3.
- 4 Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior Research*
5 *Methods*, *43*(4), 981–986. doi:10.3758/s13428-011-0097-5
- 6 Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data.
7 *Psychometrika*, *24*, 95–112.
- 8 Haubensak, G. (1992). The consistency model: A process model for absolute judgements.
9 *Journal of Experimental Psychology: Human Perception and Performance*, *18*(1), 303–309.
- 10 Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal*
11 *of Statistics*, *6*(2), 65–70.
- 12 Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*(3), 1221–
13 1247.
- 14 Kleiner, M., Brainard, D., & Pelli, D. (2007). What’s new in Psychtoolbox-3? In *Perception 36*
15 *ECVP*.
- 16 Krueger, L. E. (1972). Perceived numerosity. *Perception & Psychophysics*, *11*, 5–9.
17 doi:10.3758/BF03212674
- 18 Link, T., Huber, S., Nuerk, H.-C., & Moeller, K. (2014). Unbounding the mental number line -
19 new evidence on children’s spatial representation of numbers. *Frontiers in Psychology*, *4*,
20 Article 1021.
- 21 Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality.
22 *Nature*, *215*, 1519–1520.
- 23 Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: Compressed scaling of
24 numerical information in the primate prefrontal cortex. *Neuron*, *37*(1), 149–157.
25 doi:10.1016/S0896-6273(02)01144-3
- 26 Opfer, J. E., Siegler, R. S., & Young, C. J. (2011). The powers of noise-fitting: Reply to Barth
27 and Paladino. *Developmental Science*, *14*(5), 1194–1204. doi:10.1111/j.1467-
28 7687.2011.01070.x
- 29 Parducci, A., & Perret, L. F. (1971). Category rating scales: Effects of relative spacing and
30 frequency of stimulus values. *Journal of Experimental Psychology*, *89*(2), 427–452.
- 31 Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: evidence for
32 multiple representations of numerical quantity. *Psychological Science*, *14*(3), 237–243.
33 doi:10.1111/1467-9280.02438
- 34 Stevens, S. S. (1957). On the psychophysical law. *Psychological Review*, *64*(3), 153–181.
- 35 Teghtsoonian, M. (1965). The judgment of size. *The American Journal of Psychology*, *78*(3),
36 392–402.
- 37 van Oeffelen, M. P., & Vos, P. G. (1982). A probabilistic model for the discrimination of visual
38 number. *Perception & Psychophysics*, *32*(2), 163–170. doi:10.3758/BF03204275
- 39 Walsh, V. (2003). A theory of magnitude: Common cortical metrics of time, space and quantity.

- 1 *Trends in Cognitive Sciences*, 7(11), 483–488.
- 2 Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The
- 3 psychophysics of number representation. *Psychological Science*, 10(2), 130–137.
- 4

Appendix: Erklärungen zur Dissertation



Erklärung gemäß (bitte Zutreffendes ankreuzen)

§ 4 (1c) der Promotionsordnung

des Instituts für Bewegungswissenschaft der Universität Hamburg vom 18.08.2010

§ 5 (4d) der Promotionsordnung

des Instituts für Psychologie der Universität Hamburg vom 20.08.2003

Hiermit erkläre ich,

_____ (Vorname, Nachname),

dass ich mich an einer anderen Universität oder Fakultät noch keiner Doktorprüfung unterzogen oder mich um Zulassung zu einer Doktorprüfung bemüht habe.

Ort, Datum

Unterschrift



Eidesstattliche Erklärung nach *(bitte Zutreffendes ankreuzen)*

- § 7 (4) der Promotionsordnung des Instituts für Bewegungswissenschaft der Universität Hamburg vom 18.08.2010**
- § 9 (1c und 1d) der Promotionsordnung des Instituts für Psychologie der Universität Hamburg vom 20.08.2003**

Hiermit erkläre ich an Eides statt,

1. dass die von mir vorgelegte Dissertation nicht Gegenstand eines anderen Prüfungsverfahrens gewesen oder in einem solchen Verfahren als ungenügend beurteilt worden ist.
2. dass ich die von mir vorgelegte Dissertation selbst verfasst, keine anderen als die angegebenen Quellen und Hilfsmittel benutzt und keine kommerzielle Promotionsberatung in Anspruch genommen habe. Die wörtlich oder inhaltlich übernommenen Stellen habe ich als solche kenntlich gemacht.

Ort, Datum

Unterschrift