On the response function and range dependence of manual estimation

Karl K. Kopiske [a, c], Fulvio Domini [a, b]

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

b: Department of Cognitive, Linguistic and Psychological Sciences, Brown University, Providence, RI 02912, USA.

c: Cognitive Systems Lab, Institute of Physics, Chemnitz University of Technology, 09126 Chemnitz, Germany.

Corresponding author:

Karl K. Kopiske [a, c]

email:

karl.kopiske@physik.tu-chemnitz.de

address:

Cognitive Systems Lab, Institute of Physics,

Chemnitz University of Technology,

Reichenhainer Str. 70,

09126 Chemnitz, Germany.

Running Head: What is manual estimation?

Manuscript overall word count: 8590 (main text only: 6401).

Abstract

Manual estimates without vision of the hand are thought to constitute a form of cross-modal matching between stimulus size and finger opening. However, few investigations have systematically looked at how manual estimates relate a perceived size to the response across different ranges of stimuli. In two experiments (N=18 and N=14), we sought to map out the response properties for (i) manual estimates of visually presented stimuli as well as (ii) visual estimates of proprioceptive stimuli, and to test whether these properties depend on the range of stimuli. We also looked at whether scalar variability is present in manual estimates, as predicted by Weber's Law for perceptual tasks. We found that manual estimates scale linearly and with a slope of close to 1 with object sizes up to 90 mm, before participants' hand size limited their responses. In contrast, we found a shallower response slope of about 0.7 when participants performed the inverse task, adjusting the size of a visual object to match a not actively chosen, induced finger opening. Our results were mixed with regards to scalar variability in large objects. We saw some indication of a plateau, but no evidence for an effect of mechanical constraints in the range studied (up to 90 mm). Participants also showed a clear tendency to overestimate small differences when a set of objects differed little in size, but not when stimulus differences were more pronounced.

Keywords: manual estimation, sensory matching, sensory magnitude, vision, finger span

THIS PAPER HAS BEEN ACCEPTED FOR PUBLICATION IN *EXPERIMENTAL BRAIN RESEARCH*!

For a full and up-to-date version, please visit http://doi.org/10.1007/s00221-018-5223-5

1. Introduction

Manual estimation (ME), the action of giving an estimate of a perceived size with the thumb and index finger, is a frequently used measure not just in behavioural research, but also in everyday life – think about the last time you described something as ‘about *this* big’ with the corresponding gesture. As a way to measure size perception, ME is intuitive, flexible to execute, and due to its similarity with regards to motor demands potentially a very useful comparison to grasping measures. What is less known are the mechanisms of how it relates size perception to the motor response, and what parameters need to be considered when using it as a perceptual measure.

1.1 ME as cross-modal matching

ME has been likened to a manual read-out, or cross-modal matching (Stevens, 1959) of perceived size (Haffenden, Schiff, & Goodale, 2001; Kopiske, Bruno, Hesse, Schenk, & Franz, 2016), in which the opening of the digits is matched to the size of the stimulus. Most commonly, this stimulus would be presented visually. In one version of the task (‘open-loop’ ME), participants are prevented from seeing their own hand, in which case the match between modalities is that of proprioceptive cues about the hand to a visual percept. This is a classic concept in psychophysics: One sensory magnitude is adjusted until perceived equal in magnitude to another, allowing the researcher to determine a function that relates the two modalities (Stevens, 1946; R. Teghtsoonian, 1971). Such a matching is straightforward when the magnitudes are clearly defined. It is quite clear for example how loudness and vibration can be adjusted to match one another (Stevens, 1959). However, even in a simple ‘open-loop’ ME where finger opening is adjusted based solely on proprioceptive information, such proprioceptive information may well differ depending on how aperture is achieved. That is, it might matter whether an aperture is deliberately chosen or passively induced without being entirely under the participant's control (see e.g. Shadmehr, 2017). In addition, both types of action would depend on information from different types of muscle fibres (i.e., type Ia muscle fibers for a quickly achieved active opening and type II muscle fibers for sensing a slow, induced opening; see Boyd, 1980). Both actively chosen and passively induced apertures can be monitored online via proprioceptive signals, but actively chosen apertures may depend more strongly on movement planning processes (i.e., a type of forward model, Wolpert, Ghahramani, & Jordan, 1995, although more recent work [Gallivan, Logan, Wolpert, & Flanagan, 2016] has pointed out that under conditions of ambiguity, participants may simply have multiple movement plans existing in parallel). It has also not been systematically investigated how the manual estimate relates to the input it is supposed to measure. While several studies have looked at responsiveness (i.e., by how much a response changes given a certain stimulus change, see Franz, 2003; Kopiske, Bruno, et al., 2016) and precision (Bruno, Uccelli, Viviani, & De'Sperati, 2016; Davarpanah Jazi & Heath, 2014; Heath & Manzone, 2017; Kopiske, Gornik, & Franz, 2016) of ME, its response function has never been fully mapped out. In other words, we do not know what goes into 'about *this* big', and how big it really is.

1.2 Range-dependencies in ME

Accounting for the slope, or responsiveness, of a given measure is crucial when comparing it to another measure on the same set of inputs. Since manual estimates are often compared to the maximum grip apertures of grasping movements, several previous studies have investigated the response properties of ME. In many cases, manual estimates tend to overestimate differences such that responses scale with object size with a slope substantially larger than 1 (Franz, 2003; Haffenden et al., 2001; Kopiske, Bruno, et al., 2016). However, this is not always the case, as more recent studies reported slopes much closer to 1 for manual estimates (Bruno et al., 2016; Heath & Manzone, 2017; Kopiske, Gornik, et al., 2016). One potentially important difference between these two groups of studies is the range of object sizes presented. Indeed, it has been shown that larger ranges of input stimuli tend to elicit lower responsiveness in classic matching studies, both between (Poulton, 1967) and within modalities (R. Teghtsoonian, 1973). This is consistent with findings in ME: the latter three studies used stimuli spanning ranges of over 60 mm and found slopes close to 1 (Bruno et al., 2016; Heath & Manzone, 2017; Kopiske, Gornik, et al., 2016), whereas studies investigating much smaller ranges (< 10 mm) found steeper slopes (Franz, 2003; Haffenden et al., 2001; Kopiske, Bruno, et al., 2016). While slopes can be calculated from both types of studies and descriptively seem to differ, this has, to our knowledge, not been investigated.

1.3 Motor constraints in ME

Despite its use as a perceptual measure, a manual estimate ultimately consists of the motor action of moving the thumb and index finger into the correct positions to indicate a given size. This raises the question of how such actions are performed and whether neural computations are based on a size, or perhaps rather positions for each digit (as proposed in the model of grasping by Smeets & Brenner, 1999). It also introduces additional noise into the response as well as potentially systematic distortions, since hand gestures not only have a natural limit based on hand size but may also behave differently when nearing those limits (i.e., for large hand openings), as more force is needed for the action. This in turn may mask some properties of the perceived magnitudes ME is supposed to measure. For example, some controversy (see Ganel, Chajut, & Algom, 2008; Smeets & Brenner, 2008, or more recently Bruno et al., 2016; Heath & Manzone, 2017) has recently ensued about the possibility of the variability of manual estimates not scaling linearly with stimulus size (sometimes referred to as violating Weber's law – Fechner, 1860) given a wide enough range of stimuli. This has been proposed to be due to precisely the above-mentioned motor constraints for large openings (Bruno et al., 2016; Löwenkamp, Gärtner, Haus, & Franz, 2015; but see also Manzone, Davarpanah Jazi, Whitwell, & Heath, 2017; Schenk, Utz, & Hesse, 2017).

1.4 Our study

Our experiment sought to investigate the basic properties of ME with regards to three specific questions. First, is the cross-modal matching of visual size and finger opening dependent on the way the finger opening is achieved, and what is the response function of ME for visual stimuli? Second, do ME's properties depend on the range of objects to be estimated? Third, at what point are those properties affected by motor constraints?

We conducted two experiments to help shed light on these questions. The first experiment was designed to map out the response function of ME, and to compare its properties to a task of matching a visual size to an induced finger opening. Such an 'inverse ME' task could potentially tell us about the mechanisms of ME and answer our question whether it is as simple as 'proprioception matched to vision'. To this end, we had participants perform (1) open-loop ME (that is, ME without vision of the hand while the visual stimulus was presented continuously) on a range of objects that was chosen to exhaust each participant's range of possible finger openings, and (2) adjustment of a visual probe object to match an induced finger opening. Our second experiment tested whether the properties of manual estimates depended on the range of objects presented. Participants estimated a range of object sizes that spanned either 50 mm or only 5 mm. In both experiments, we recorded participants' hand sizes to investigate whether this natural ceiling for finger spans would impact the response function. We also investigated the scaling of variability with object size. This was not the main target of our experiment, but several competing accounts (Bruno et al., 2016; Heath & Manzone, 2017) make clear predictions that could be tested with data obtained from a design like ours.

2. Experiment 1: ME as matching visual size to felt finger span

In our first experiment, we aimed to investigate basic properties of standard open-loop manual estimates. To this end, we explored the response slope, possible biases, as well as the variability of manual estimates. We also measured the same parameters in a task designed to be the inverse of open-loop ME to test whether the matching of visual size to finger opening depends on whether this finger opening is chosen actively by the participant, or passively. To do this, we employed two different tasks where we (i) asked participants to manually estimate different visual sizes, as well as (ii) to adjust visual objects to match a felt finger opening. In addition, we included (iii) a task to measure how manual estimates would behave close to the physical limits imposed by participants' hand span.

2.1 Participants

Nineteen participants participated in the first experiment. One participant could not complete the experiment due to technical problems, leaving us with N=18 participants (17 right-handed, 13 women, age range 20 to 35 years, mean age = 25.4; this included author KKK) as the complete sample in all subsequent analyses. Participants were either volunteers from CNCS@UniTn department, or recruited via an online advert on facebook.com and received 6€ (8€ per hour). All participants gave written, informed consent to participate in the study and have the data collected published in an anonymous format. The project was approved by the life sciences ethics committee at University of Trento (*Comitato Etico per la Sperimentazione con l'Essere Vivente dell'Università degli Studi di Trento*), and participant data were protected according to the 1964 Declaration of Helsinki.

2.2 Stimuli and apparatus

Participants were seated with their head in a chin-rest in front of a semi-transparent mirror, slanted 45° away from the body midline, projecting images from a 19'' CRT monitor (running at 100Hz and 1024*768 px, located to the left of the participant) to a position in front of the participant. The chin rest was located on a table, 50 cm above the table's surface. Infrared-emitting diodes were located on two poles attached to the fingernails of the right thumb and index finger, and one diode was attached to the participant's right wrist. Makers on thumb and index finger were used to measure the distance between these two digits, while data from the wrist marker were recorded to allow us to check the consistency of potentially unusual finger movements, but otherwise not further used. An Optotrak Certus (Northern Digital, Waterloo, Canada) was used to track the position of these diodes at a frequency of 100 Hz. Stimuli were virtual red rectangles of 20 mm width and variable height, presented at a distance of 420 mm from the participant, centrally and at eye height, on an otherwise black screen. The stimuli were rendered in a custom C++ program using OpenGL and the GLUT toolkit. No physical objects were involved, allowing us to easily use small increments of visual size across a large range of stimuli.

2.3 Procedure

Prior to the experiment, each participant's maximal finger span on their right hand was taken by asking participants to separate thumb and index finger as widely as possible and measuring the distance between the inside tip of the thumb and the inside tip of the index finger with a ruler. The result was rounded to the nearest multiple of 5 mm. Next, the right thumb and index finger's fingertip positions relative to the diodes on the poles were calibrated by having the participant lay each finger on a diode attached to a movable platform which was moved to the back of the table after calibration.

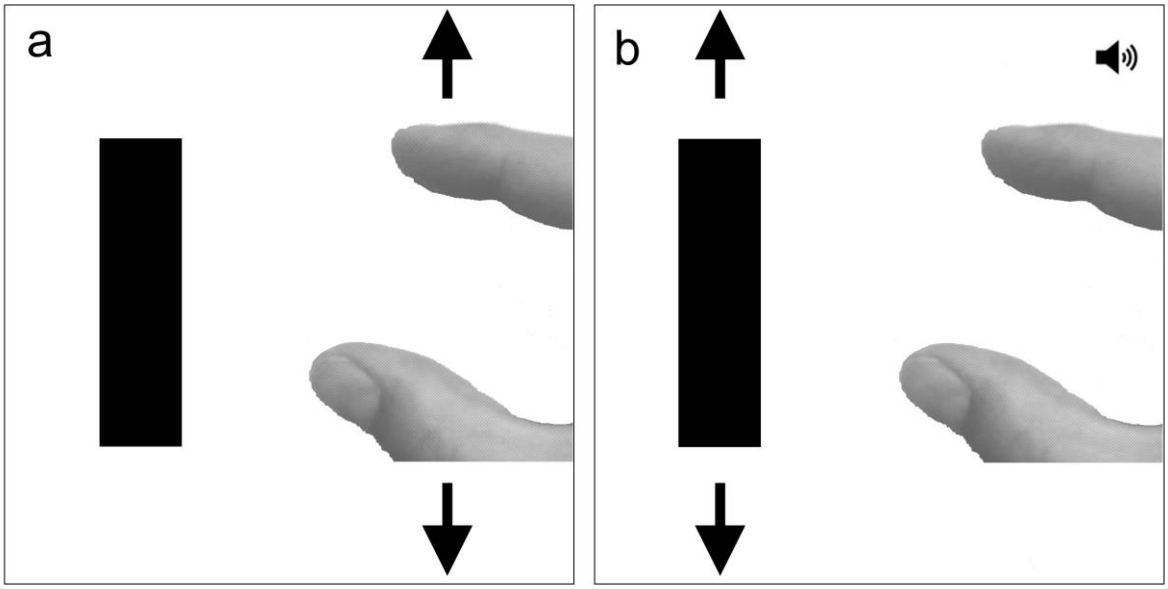

Figure 1: Schematic illustration of the tasks used in experiment 1. (**a**) ME task, where the height of a rectangle is estimated between thumb and index finger, (**b**) the visual adjustment task, where participants opened their fingers until a sound indicated the correct opening and a rectangle's height could be adjusted to match the opening. Infrared-emitting diodes were attached to the thumb and index finger to record the finger opening.

There were three blocks to each experiment, each of which started with 5 practice trials. One block was a standard, 'active' open-loop ME task (figure 1a) with objects ranging between 20 mm and 90 mm in steps of 10 mm. Each object was repeated 5 times, for a total of 40 experimental trials. Each trial started with a beep and the object appearing in the mirror, after which participants would indicate the size with their right hand, lifting it above the table and pressing the space bar with the left hand when they were satisfied with their response. The Euclidean distance between the tip of the thumb and the tip of the index finger at the moment when the space bar was pressed was used as the dependent variable. The object remained visible throughout the trial. This was done without time constraints and with no instructions about how to orient the hand, except pointing out that the back of the hand needed to face the Optotrak (which was located behind and to the right of the participant) for the markers to be visible.

Another block was a visual adjustment task (figure 1b) in which participants were instructed to open their hand until they heard a sound, and then adjust a visually presented object to match the signalled, 'induced' hand opening. The sound appeared when the opening was within 4 mm of the desired aperture and continued while participants tried to maintain the finger opening. Objects were adjusted to be 1 mm larger or smaller, respectively, by pressing the ',' and '.' keys of a standard Italian USB keyboard. Pressing the space bar confirmed the response. The initial height of each object was a random integer between 1 and 100 mm. The same stimulus sizes and number of repetitions as in the ME task (20 mm to 90 mm, 5 repetitions) were used. Of course, this method implies that there was some variability in stimulus magnitude. However, the only way to prevent such variability would have been moving the fingers by applying force externally, which would have given additional haptic input and thus resulted in a poorer match between the information available in this task and ME. The third kind of block was an investigation of the full range of the ME response function where the stimuli were objects up to the maximum opening of the hand. Each object size between 10 mm and the maximum size (step size 10 mm) was repeated twice, for a maximum of 38 trials. In all tasks, stimuli were presented in a randomised order, and trials with missing data in over 20% of frames were marked as invalid and repeated for up to two times at a random time (this concerned 7 ME trials and 26 full-range trials). The order of blocks was counterbalanced between participants.

2.4 Results and Discussion

We excluded outliers when the response in a given trial was 3 or more inter-quartile ranges (IQRs) larger or smaller than the 1st or 3rd quartile, respectively (criteria used in Bruno et al., 2016). This concerned three trials in the adjustment task and one trial in ME, leaving us with 717 adjustment trials, 559 full-range task trials, and 718 ME trials (one trial in each ME and the full-range task was removed due to technical difficulties, i.e., missing frames around the response). All subsequent analyses were conducted on these data. All data processing and analysis was conducted using R (R Core Team, 2015).

Mean trial durations (from the start beep until the ‘space’ bar was pressed) were 3243 ms in the ME task, 3575 ms in the full-range task, and 16360 ms in the adjustment task. An 8 (stimulus *size*, within-participant factor) * 6 (*order* of tasks standard ME, full-range ME, adjustment, between-participant factor) mixed ANOVA was conducted for ME to test whether (a) participants scaled with stimulus size and (b) whether it mattered in which order the tasks were conducted. Such an ANOVA cannot be calculated for the adjustment task, since there is variability in the predictor variable; hence, we only modelled the response function. Greenhouse-Geisser correction (Greenhouse & Geisser, 1959) was applied for all results involving within factors with more than two levels. We report the corresponding correction factor ε_{gg} in these cases. For ME, we found an effect of *size* ($F(7, 84) = 241.5$, $p_{gg} < .001$, $\varepsilon_{gg} = .33$) and no effect of *order* ($F(5, 12) = 1.7$, $p = .213$), as well as no statistically significant interaction ($F(35, 84) = 2.1$, $p_{gg} = .053$, $\varepsilon_{gg} = .33$).

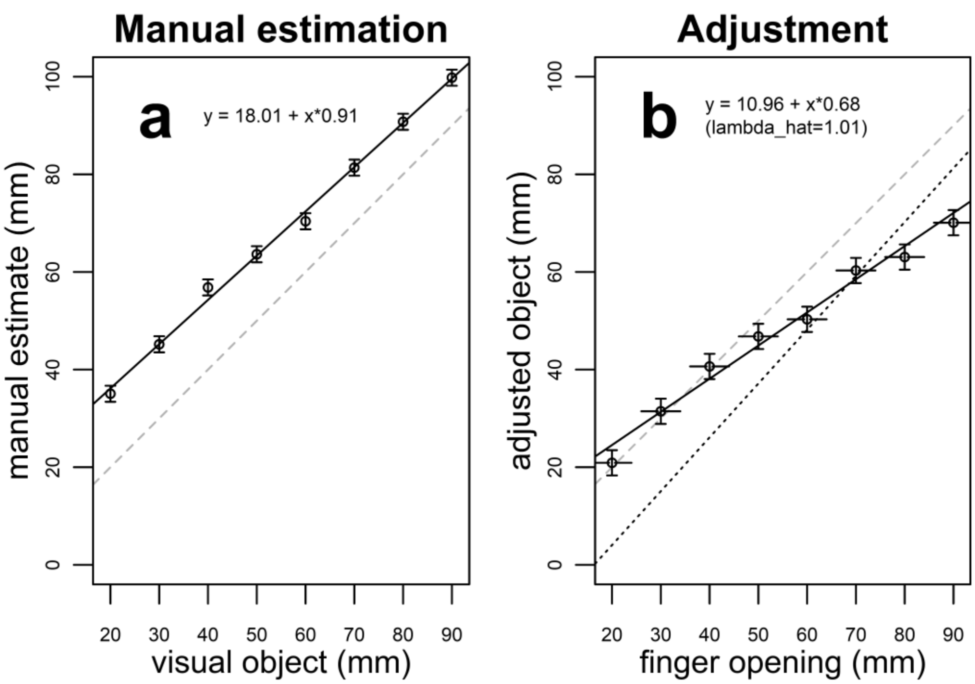

Figure 2: Mean responses by stimulus size in (**a**) ME and (**b**) visual adjustment task. Dashed grey line indicates unity, solid black line indicates the best linear fit, for which mean coefficients are given. Dotted black line in panel **b** indicates inverted response function from the ME task. Vertical error bars indicate within-subject SEM for the pooled differences between levels of factor size (Franz & Loftus, 2012; Loftus & Masson, 1994), horizontal bars represent the ± 4mm range of apertures that would have triggered the 'correct aperture' tone during the adjustment. Slope in the adjustment task was corrected for an estimated correction factor $\hat{\lambda}$ (see below and Frost & Thompson, 2000; Searle, Casella, & McCulloch, 1992).

To investigate the response functions more closely, we fit the mean responses (for each participant) to three different models: A linear function $y = a + bx$, a simple power model $y = a * x^b$ (as proposed by Stevens, 1957) and a power function with an additive constant $y = a * x^b + c$ (see R. Teghtsoonian, 1973). In both tasks, the more complex power model failed to predict the data better than the simpler model (as indicated by the difference in Akaike's information criterion, ΔAIC - Akaike, 1974; Burnham & Anderson, 2004), as the responses were already fitted very well with two free parameters (especially on aggregate, see figure 2). The AIC slightly preferred the linear model in ME and the power function in the adjustment task (ΔAIC of

-0.5 and -1.9, respectively). For the sake of simplicity and comparability, we chose the linear models for further analysis. The mean responses by size, as well as best-fitting linear models are shown in figure 2. As can be seen, participants scaled slightly shallower than unity in ME, with a mean slope of 0.91 ± SEM of 0.06. Results were similar for the adjustment task, with a mean slope of 0.68 ± 0.08. For the latter regression, we had to account for the noise of the predictor (the finger aperture) introduced by the fact that the 'go' tone was present when the aperture was within a region, not when it assumed a precise value. Thus, we corrected for dilution bias by a factor of $\lambda = 1 + \frac{\sigma_w^2}{\sigma_b^2}$, with σ_b^2 being the underlying, long-term variability of the predictor and σ_w^2 being the random error in its measurement (Frost & Thompson, 2000; Searle et al., 1992). To estimate the random error, we assumed a uniform distribution of apertures within the possible range of ±4 mm. This gave us a correction factor of 1.01. Standard errors of the slope were estimated using 1,000 bootstrap samples (Efron & Tibshirani, 1993). Importantly, these results show that ME and the inverse adjustment task do not display inverse response functions (contrary to what would follow from the principles named by Stevens, 1959), see the dotted black line in figure 2b. That is, participants' responses scaled with a slope of less than 1 in ME, so that an inverse response function would have to show slopes larger than 1. However, slopes in the adjustment task (which was designed to employ the input modality of ME as the response and the response modality of ME as the input) were also significantly smaller than 1 ($t(17) = -4.0$, $p < .001$). In addition, the slopes were also not inversely related on an individual level (product-moment correlation of -.28, $t(16) = -1.2$, $p = .266$; see figure 3).

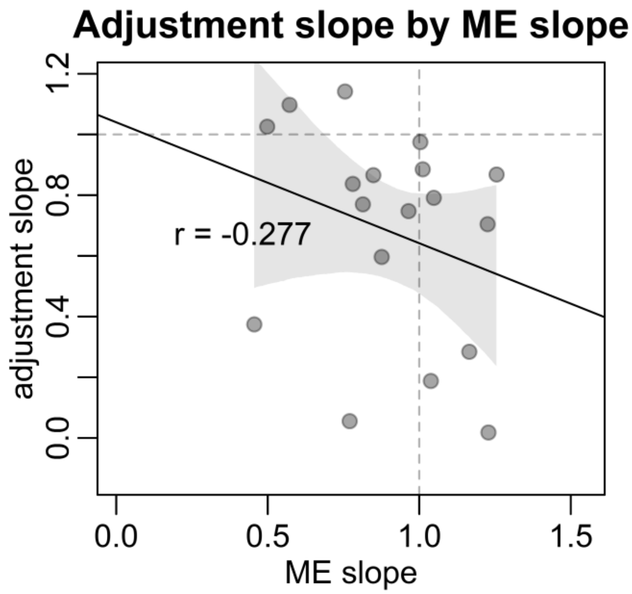

Figure 3: Slopes in the visual adjustment plotted by slopes in ME. Each point represents one participant: The x-coordinate indicates that participant's response slope in ME (in the linear model), y-coordinate that same participant's (linear) slope in visual adjustment. We see a slight negative correlation. Grey area depicts 95% confidence interval based on 10,000 bootstrap samples (Efron & Tibshirani, 1993).

We also included a full-range task to confirm that responses tended to tail off at a certain size. As can be seen in figure 4, this was the case for all participants. While this was to be expected, we note that there is virtually no sign of nonlinearity to be seen in the range employed in our other tasks.

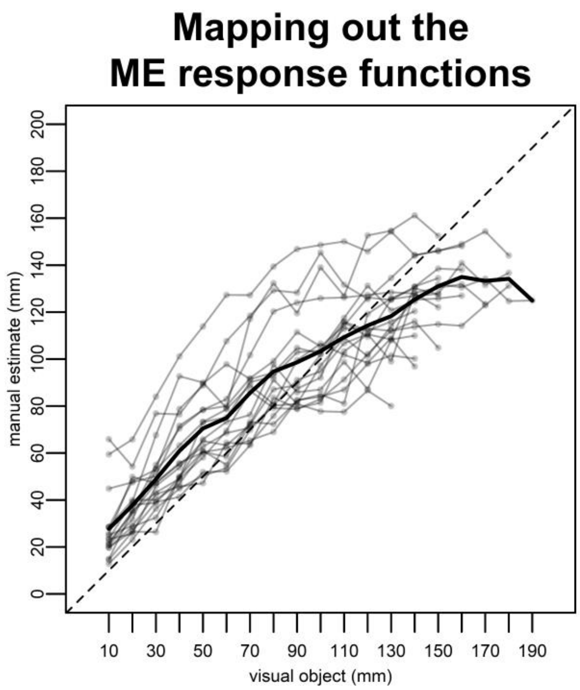

Figure 4: Individual responses by size in the 'full range' task of experiment 1, where participants manually estimated objects from 10 mm up to an object size almost equal to their maximal finger span (i.e., rounded down to the nearest multiple of 10). Dashed line indicates unity, solid black line indicates the mean response.

To investigate scalar variability of responses in our data and contribute to the recent debate about whether ME follows the predictions of Weber's Law (e.g., Bruno et al., 2016; Heath & Manzone, 2017), we calculated the SD and inter-quartile range (IQR) for responses to each stimulus size. Means of the SDs computed for each participant and each size separately can be seen in figure 5. Repeated-measures ANOVA on SDs with the factor 'size' (8 levels) revealed a significant main effect in ME ($F(7, 119) = 3.5$, $p_{gg} = .005$, $\varepsilon_{gg} = .78$), indicating that variability did indeed differ with stimulus size. Fitting a spline with a single knot at 40 mm to the ME data (40 mm being a proposed starting point for where ME's variability may start to plateau by Bruno et al., 2016) gave us a mean slope of 0.18 ($t(17) = 3.8$, $p = .001$) for the first component and -0.01 ($t(17) = -0.5$, $p = .636$) for the second component, consistent with the notion that scalar variability was only present at small stimulus sizes (figure 5a). Since it has been proposed that mechanical constraints might affect the variability in relatively large apertures during grasping (Löwenkamp et al., 2015; Utz, Hesse, Aschenneller, & Schenk, 2015) as well as ME (Bruno et al., 2016), we also show the skew of responses by stimulus size (figure 5e). If motor constraints biased the variability depending on size (i.e., if larger responses were rarer due to being more effortful for the participant), we would predict fewer responses above the mean for larger objects, leading to a negatively skewed (or left-skewed) distribution of responses. While there is a relationship between skewness and object size relative to hand span (figure 5f), this relationship is quite weak, and not apparent when looking just at object size.

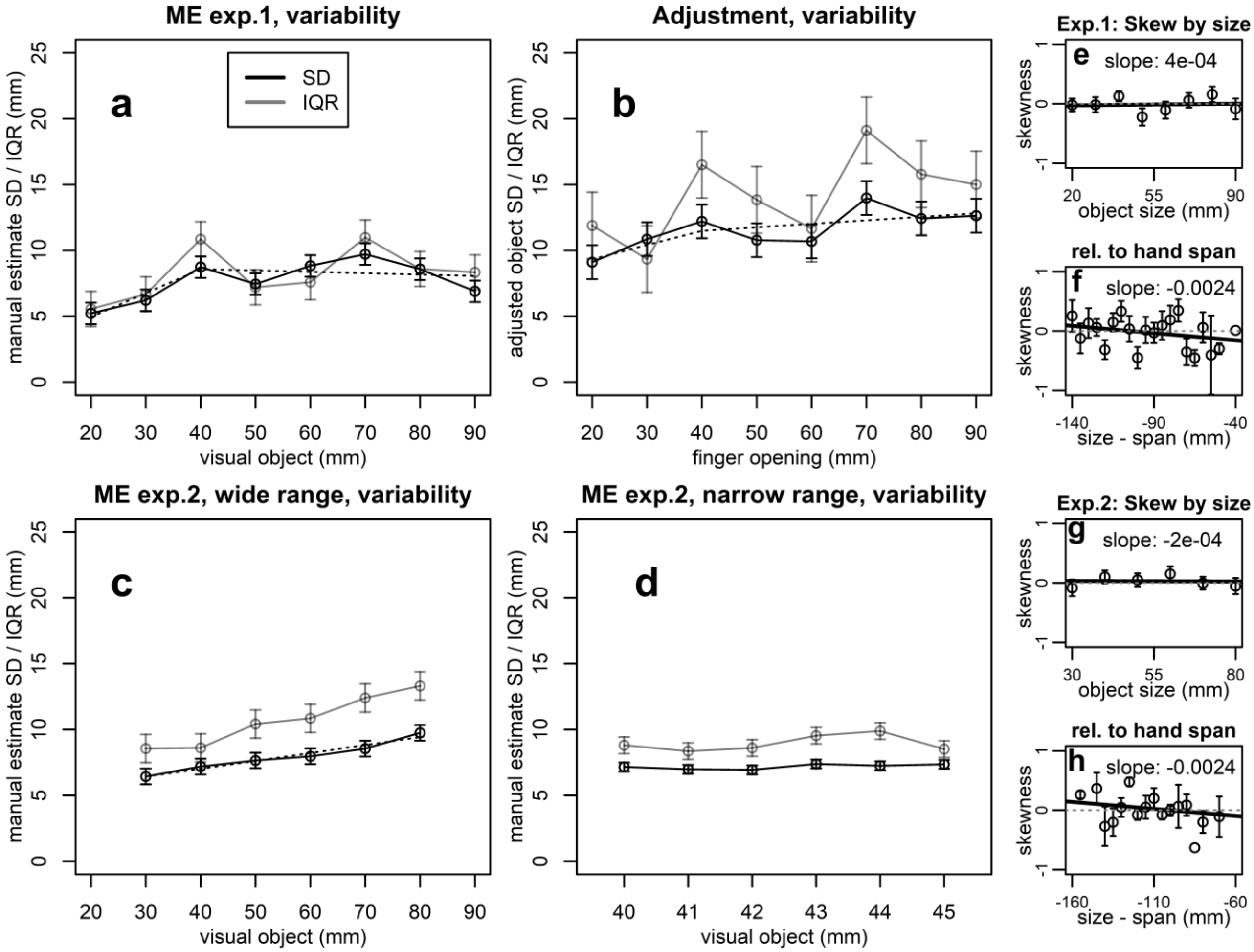

Figure 5: Variability measures in the four tasks, top row: Experiment 1, bottom row: Experiment 2. Plotted are SDs (black) and IQRs (grey) by stimulus size in experiment 1, (**a**) ME and (**b**) visual adjustment, as well as experiment 2, in the (**c**) wide range and (**d**) narrow range. Error bars indicate pooled within-subject SEMs for within-subject differences between levels of the factor *size* (Franz & Loftus, 2012; Loftus & Masson, 1994). Dotted line shows spline regression with a knot at 40 mm. Skewness by object size in the ME task of experiment 1 (**e**) and the large-range task of experiment 2 (**g**). Skewness plotted by *object size relative to maximal hand span* in panels (**f**) and (**h**), exploring the possibility of mechanical constraints driving skew. Error bars in panels e-h show between-subject SEMs.

3. Experiment 2: ME and stimulus range

To investigate the degree of range-dependence of the properties of ME, we conducted a second experiment in which we manipulated the set of stimuli presented, with each participant completing one session of an ME task with a large range of stimuli and one session with a small range of stimuli. Chiefly, we expected the response slope to be larger in the small-range condition than in the large-range condition, as is typically found in cross-modal matching (R. Teghtsoonian, 1973) and also in the ME literature (e.g., compare Franz, 2003 and Bruno et al., 2016). In a within-subject design with two blocks conducted on consecutive days, each participant performed an ME task with a narrow range of object sizes as well as an ME task with a wide range of object sizes, allowing us to test this notion.

3.1 Participants

A total of N=14 participants (all right-handed, 9 women, mean age 23.4 years, age range 19 to 39) took part in experiment 2. Recruitment, ethics, and data protection issues were treated the same way as in experiment 1. Participants received 8€ in compensation. None of the participants from experiment 1 participated in this experiment.

3.2 Stimuli and apparatus

Participants completed a similar ME task to the one administered in experiment 1. Only the height of the stimuli differed: Each participant completed a 'narrow range' session, in which the stimuli were rectangles of 40, 41, 42, 43, 44, 45 mm height, as well as a 'wide range' session with objects of 30, 40, 50, 60, 70, 80 mm height.

3.3 Procedure

Each participant completed two sessions (narrow range, wide range) of open-loop ME. The two sessions were always completed on consecutive days to avoid learning effects (as found in classic rating tasks where participants may develop and retain a scale for their responses based on trials at the beginning of the experiment, Haubensak, 1992). Order was counterbalanced between participants. Each participant was given 5 practice trials in each session, and completed 20 repetitions for each size, for a total of 120 experimental trials per session. Other than the stimulus sizes and the number of repetitions, the task was identical to the ME task in experiment 1. Overall, each session lasted between 20 and 30 minutes.

3.4 Results and discussion

A total of 43 trials were missing due to technical difficulties due to missing frames or were filtered as outliers, primarily in one participant (where too many frames were missing in 29 trials; despite this, no cell had fewer than 15 trials usable for evaluation). In the large range, 19 trials were repeated during the experiment due to missing frames, and 5 trials in the small range. A further 12 trials had to be removed in other participants' data due to a too high proportion of frames being invalid (see section 2.4). The same criteria as in experiment 1 were used to remove outliers, which concerned two trials, leaving us with a total of 1667 trials in the wide range and 1650 trials in the narrow range.

Mean trial durations were 2878 ms in the narrow range and 3104 ms in the wide range. For both the wide and the narrow range, we first conducted a mixed 2 (*order* of blocks; between-subject factor) * 6 (stimulus *size*) ANOVA akin to the one conducted for experiment 1 to ascertain that participants scaled their estimates with the height of the stimulus. In the wide range, we found the expected main effect of *size* ($F(5, 60) = 136.9$, $p_{gg} < .001$, $\varepsilon_{gg} = .28$), with no main effect of *order* ($F(1, 12) = 1.1$, $p = .324$) and no interaction ($F(5, 60) = 2.2$, $p_{gg} = .150$, $\varepsilon_{gg} = .28$). For the narrow range, the pattern was the same: A main effect of *size* ($F(5, 60) = 17.4$, $p_{gg} < .001$, $\varepsilon_{gg} = .24$), but not of *order* ($F(1, 12) = 1.1$, $p = .308$) and no interaction ($F(5, 60) = 1.0$, $p_{gg} = .361$, $\varepsilon_{gg} = .24$).

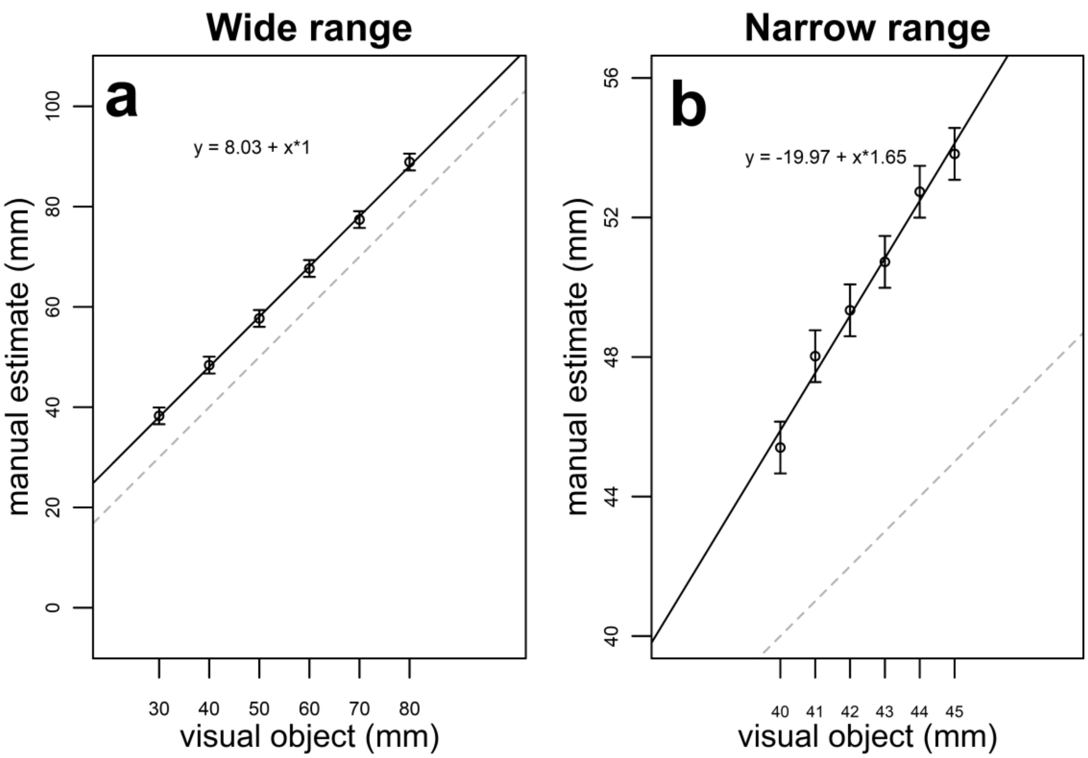

Figure 6: Mean ME responses in experiment 2 by stimulus size in (**a**) the wide range (30…80 mm) and (**b**) the narrow range (40…45 mm). Dashed grey line indicates unity, solid black line indicates the best linear fit, for which mean coefficients are given. Error bars indicate within-subject SEMs (Franz & Loftus, 2012; Loftus & Masson, 1994).

To investigate scaling more in-depth, we fit linear and power-function models to the data as was done in experiment 1. In both blocks, the linear model provided the best fit, although in each case the difference in fit with the power function was minimal, with ΔAIC = -0.1 in both cases[1]. We also found that participants' estimates were more responsive in the narrow range (y = -19.97 + 1.65*x) than in the wide range (y = 8.03 + 1.00*x), as was expected. This can be seen in figure 6. Comparing the slopes for the linear model revealed no significant difference, however ($t(13) = 2.0$, $p = .063$). As was also expected, the slopes for the wide range varied less than the slopes for the narrow range (SEMs of 0.08 and 0.26, respectively). This can also be seen in figure 7: Whereas slopes in the wide range fell between 0.54 and 1.47, slopes computed from the narrow range showed some participants barely responding to sizes differences (minimum slope: -0.12) with others strongly exaggerating the same differences (maximum slope: 3.95).

Interestingly, the slopes of these functions were strongly related across participants ($r = .8$; $t(12) = 4.7$, $p < .001$), by a mean function of $slope_{large} = 0.72 + 0.17 * slope_{small}$ (with SEMs of 0.03 and 0.07 for that function's slope and intercept, respectively; estimated via 10,000 bootstrap samples, Efron & Tibshirani, 1993), indicating that the larger slopes, as well as the larger variance in the narrow range, were not just random noise, but likely a systematic amplification of an existing inter-individual tendency to scale more or less strongly in ME (figure 7). This

[1] Two things should be noted for the fits in the narrow range: First, while these models provided a nearly perfect fit on aggregate (figure 6b), they did not provide a good fit on an individual basis. The reason for this is primarily the huge inter-individual variability of both the mean responses and the slopes (figure 7). Second, for the three-parameter power function $y = a * x^b + c$, the fitting algorithm did not converge due to the small spacing of values on the x-axis.

systematic variation may be further indication that the higher slope in the narrow range is not a statistical fluke, but depends on a mechanism similar to results found in classic psychophysics (Poulton, 1967; M. Teghtsoonian & Teghtsoonian, 1971; R. Teghtsoonian, 1973).

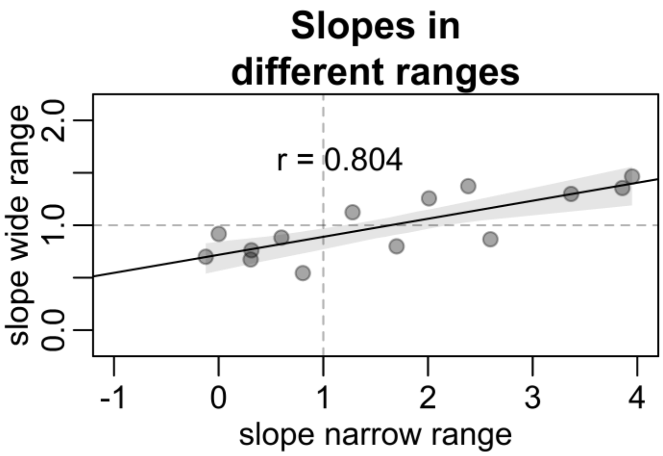

Figure 7: Slopes in the narrow range plotted by slopes in the wide range (both according to the linear model). Analogous to figure 3, each point represents one participant. Here, we see a strong correlation between slopes in the two ranges. Grey area depicts 95% confidence interval based on 10,000 bootstrap samples (Efron & Tibshirani, 1993).

Once again, we tested for variability scaling with object size (plotted in figure 5c, d) by conducting repeated-measures ANOVAs with mean SDs as the dependent variable and *stimulus size* as the only factor. We conducted this analysis only for the wide range, as the narrow range would be poorly suited to investigate this effect. The main effect of *size* was significant ($F(5, 60) = 3.7$, $p_{gg} = .034$, $\varepsilon_{gg} = .44$). A spline regression with a knot at 40 mm revealed no difference between slopes before (mean slope = 0.06, t-test against 0: $t(13) = 1.0$, $p = .339$) and after the knot (mean: 0.06, $t(13) = 2.3$, $p = .038$), although the latter was significantly different from 0.

As we did in experiment 1, we explored the possibility of mechanical constraints for larger objects by plotting the skewness of responses. Once again, we found little indication that these might have affected the results (see figure 5c, g, h), as skewness differed almost not at all by stimulus size and only slightly by *size-hand span*. In combination with the fact that the variability of responses scaled linearly with stimulus size, this is not consistent with an account of biomechanical constraints in manual estimates of large stimuli.

4. General discussion

Our experiments sought to provide a more detailed look at the properties of manual estimates to visual stimuli. Comparing the response functions obtained from two different matching procedures of visual size and finger opening, and from two stimulus ranges, we found the scaling of the response with stimulus size to be highly dependent on the task at hand. An active finger opening was matched to a different visual size than was matched to an induced opening, and narrow and wide ranges differed substantially with regards to the responsiveness of ME. Under all conditions, however, manual estimates scaled linearly with size.

Prior to running experiment 1, we fully expected the response functions of ME on the one hand and of matching visual size to a hand opening on the other hand to be inversely related. The fact that this was clearly not the case is not what one would expect following the simple model of cross-modal matching following Stevens (1959), but consistent with recent findings on matching positions, where Kuling and colleagues (Kuling, van der Graaff, Brenner, & Smeets, 2017) found systematic mismatches between proprioception and vision. Similar to our results, Kuling et al. (2017) reported different (and not mutually predictive) systematic errors in a task where participants moved either an unseen finger to a seen position, or a seen target to the position of an unseen finger. These response biases occur even in the absence of time constraints (Kuling, Brenner, & Smeets, 2013), which is also in line with our findings. Kuling et al. (2017) takes this as evidence that the transformation of visual to proprioceptive information biases the encoding of position in a different way than the inverse transformation from proprioception to vision. Such a mechanism could also explain our results. Another possible explanation is that the action of deliberately creating a finger opening contains information that is not available when the aperture is induced (van Kemenade, Arikan, Kircher, & Straube, 2016). This would involve feedforward mechanisms (Wolpert et al., 1995) using efference copies of motor commands to predict hand postures. Such an explanation would be consistent with the fact that we found a steeper slope (and closer to unity) when participants had this information, indicative of a stronger signal for the size of the aperture. It is also consistent with the much higher variability in adjustment responses, see figure 5, which can only be due to the difference in how the aperture is felt, as the visual information should be equally precise in both tasks – and in any case, it is known that visual perception is typically much less noisy than haptic or proprioceptive information (see e.g. Ernst & Banks, 2002).

Note that the noise inherent in this signal is not to be confused with noise inherent in our design. Using an auditory beep as indication of the correct aperture (as opposed to e.g. an object inserted between the fingers, Stevens & Stone, 1959) would necessarily introduce some noise since we had to specify a correct "region" of apertures. However, since the beep was either present or not, there was no uncertainty for the participant as to whether the aperture was correct or not. In addition, the added noise could only have been very small relative to the observed noise of the manual estimates. Assuming a uniform distribution for apertures within the specified region of 8 mm (4 in each direction), the added variance would have been 5.33 mm², which – if added to the variance inherent in ME – would result in an increase in SD of under 10% even for the smallest object, which is much smaller than the observed increase in visual adjustment (figure 5). Thus, like its effect on the slope (see section 2.4), the impact of stimulus-variability is small here. The issue of hand-opening judgements depending on how the opening was induced need not be a problem for standard ME tasks. In such tasks, the aperture will always be formed actively and in a very similar way in any experiment. It does, however, reiterate the question of whether ME is a purely perceptual task, as well as how strongly its properties depend on the exact design of the task. This relates to our second main finding.

The second major finding is that the responsiveness of ME is highly range-dependent (figure 6). This is not surprising, but it has consequences, as the property of responsiveness has been much-discussed (e.g., Franz, 2003; Haffenden et al., 2001; Kopiske, Bruno, Hesse, Schenk, & Franz, 2017; Kopiske, Bruno, et al., 2016; Whitwell & Goodale, 2017), and is of particular importance when comparing ME to other output measures. The good news is that our findings support the use of linear slope-correction as a viable way to account for mean ME slopes different from 1, since the mean responses reliably showed linear scaling with stimulus size. It is also important to note that even different response functions in different ranges seem to be indicative of similar processes, seeing that they are highly correlated between subjects (figure 7). The specific nature of these processes is still unknown (and has been for a long time, see e.g. Poulton, 1967; R. Teghtsoonian, 1973), especially given that not all hand openings are created (and sensed!) equal. An unfortunate implication of the demonstrated range-dependence of slopes is that raw effects across studies cannot be usefully compared. Rather, it is necessary to estimate slopes for each study individually, which is a noisy affair especially for small stimulus ranges (e.g., see figure 7).

As we mentioned earlier, ME tasks can be employed in several different ways, which should be kept in mind when considering the generalizability of our results. In particular, we decided to use an open-loop task without vision of the hand, whereas others have used closed-loop ME tasks with full vision of the hand and stimulus (e.g., Dewar & Carey, 2006; Kopiske, Bruno, et al., 2016). These are known to have different response properties compared to open-loop ME (Kopiske, Bruno, et al., 2016), such as shallower response slopes even when employing small ranges of stimuli. Thus, it is not clear whether the range dependence found in our experiment would also be present under closed-loop conditions. Similarly, presenting participants with virtual 2D stimuli throughout the entire trial is slightly different than the limited-duration presentations used in many previous studies. These studies often used real objects and removed vision not just of the hand but also of the stimulus during ME (e.g., Bruno et al., 2016; Franz, 2003; Haffenden et al., 2001; Heath & Manzone, 2017; Kopiske, Bruno, et al., 2016). It has been shown that matching 2D and 3D stimuli produces an excellent match, both in size comparison tasks and in ME (Franz, 2003; Kopiske, Bruno, et al., 2016), but it is plausible that keeping the stimulus visible throughout each trial might lead to slightly different behaviour than estimation based on memory. However, our main goal was to study ME as a measure of visual perception. To the best of our knowledge, none of the investigations into its responsiveness or variability rely on memory mechanisms for their interpretation.

Finally, we want to address the recent discussion about scalar variability in ME. It has been argued that ME shows scalar variability and that this can tell us something about the mode of visual processing in grasping, which does not show scalar variability (Davarpanah Jazi & Heath, 2014; Ganel et al., 2008). However, this finding has been disputed, as others have reported that ME's variability increases with size only for relatively small objects (Bruno et al., 2016), or only for 'functionally graspable' objects (Heath & Manzone, 2017). Mechanical constraints (Utz et al., 2015) or motor actions using finger positions rather than magnitude for movement planning (Smeets & Brenner, 2008) have been put forward as possible explanations (although it is an open question whether these accounts, originally put forward to explain grasping behaviour, apply to different motor actions like ME; see Schenk et al., 2017). While our experiments were not designed to test these hypotheses, some make quite explicit predictions for a design like ours, which we compared to our actual data. Specifically, we see perhaps some indication of a variability plateau in experiment 1, but nothing of the sort in experiment 2. Note that despite our use of fewer trials in experiment 1 than in previous experiments (5 repetitions, compared to 20 for Bruno et al., 2016; Heath & Manzone, 2017), this should not result in a problem of power for the spline analysis, at least if we assume effects of the magnitude found in the literature: An effect of size on SD that is of the magnitude as reported by Bruno et al. (2016) would have been easily detectable (the reported *t* value for the initial slope would work out to a Cohen's $d > 2$; our sample would have given us 90% power to detect an effect in the range of $d=0.8$, Cohen, 1988; computed using G*Power, Faul, Erdfelder, Lang, & Buchner, 2007). Of course, initial estimates for effects are known to often be somewhat inflated (Button et al., 2013), which might be the case here. If indeed there are effects of mechanical constraints but those are more subtle, then our design might well have been insufficient to detect them, which would also explain the somewhat inconsistent results from our two experiments. In our view, this would be a plausible possibility, which also relates to our next point: We see no evidence of any effect of hand size on variability or skewness of responses for the largest stimuli used in the ME blocks (figure 5f, h), although as can be seen in figure 4, the corresponding apertures for these stimuli were rarely close to the mechanical limits. This kind of effect would be predicted based on certain biomechanical constraints as well as the notion of functional graspability. Similarly, an absence of scalar variability has been found recently for bimanual grasping, where such considerations should not apply (Ganel, Namdar, & Mirsky, 2017). Here, the relatively large uncertainty and small differences in effect make this a case where more data might be needed to provide a strong test of these hypotheses (seeing that other studies used similar sample sizes to ours).

5. Conclusion

In our study, we explored the properties of manually estimating object size by conceptualizing it as the matching of visual size to felt finger opening. We find that properties such as responsiveness differ markedly depending on which size is used as the standard and which size is matched to it, in a way inconsistent with a simple cross-modal matching. We also find ME's responsiveness to be strongly dependent on the range of stimuli presented. Scaling was clearly linear, within the bounds set by hand span. Results are mixed with respect to whether ME's variability scales with stimulus size.

Acknowledgements

This work did not receive external funding. The authors have no competing interest in its publication. We thank Evan Cesanek for helpful advice and for proofreading the manuscript.

References

Akaike, H. (1974). A new look at the statistical model identification. In *IEEE Transactions on Automatic Control AC-19* (pp. 716–723).

Boyd, I. A. (1980). The isolated mammalian muscle spindle. *Trends in Neurosciences*, *3*, 258–265. http://doi.org/10.1016/0166-2236(80)90096-X

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

Burnham, K. P., & Anderson, R. P. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, *33*(2), 261–304. http://doi.org/10.1177/0049124104268644

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*(5), 365–376. http://doi.org/10.1038/nrn3475

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Psychology Press.

Davarpanah Jazi, S., & Heath, M. (2014). Weber’s law in tactile grasping and manual estimation: Feedback-dependent evidence for functionally distinct processing streams. *Brain and*

Cognition, *86*, 32–41.

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

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.

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

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.

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

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

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.

Frost, C., & Thompson, S. G. (2000). Correcting for regression dilution bias: Comparison of methods for a single predictor variable. *Journal of the Royal Statistical Society. Series A*, *163*, 173–189.

Gallivan, J. P., Logan, L., Wolpert, D. M., & Flanagan, J. R. (2016). Parallel specification of competing sensorimotor control policies for alternative action options. *Nature Neuroscience*, *19*(2), 320–326. http://doi.org/10.1038/nn.4214

Ganel, T., Chajut, E., & Algom, D. (2008). Visual coding for action violates fundamental psychophysical principles. *Current Biology*, *18*(14), R599–R601.

Ganel, T., Namdar, & Mirsky, A. (2017). Bimanual grasping does not adhere to Weber's law. *Scientific Reports*, *7*, 6467. http://doi.org/10.1038/s41598-017-06799-4

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–112.

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.

Heath, M., & Manzone, J. (2017). Manual estimations of functionally graspable target objects adhere to Weber's law. *Experimental Brain Research*. http://doi.org/10.1007/s00221-017-4913-8

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://doi.org/10.1016/j.cortex.2016.03.020

Kopiske, K. K., Bruno, N., Hesse, C., Schenk, T., & Franz, V. H. (2017). Do visual illusions affect grasping? Considerable progress in a scientific debate. A reply to Whitwell & Goodale, 2016. *Cortex*, *88*, 210–215. http://doi.org/10.1016/j.cortex.2016.10.012

Kopiske, K. K., Gornik, A., & Franz, V. H. (2016). Manual estimation: Feedback affects bias but not precision. *Journal of Vision*, *16*(12), 450. http://doi.org/10.1167/16.12.45

Kuling, I. A., Brenner, E., & Smeets, J. B. J. (2013). Proprioception is robust under external forces. *PLoS ONE*, *8*(9), e74236. http://doi.org/10.1371/journal.pone.0074236

Kuling, I. A., van der Graaff, M. C. W., Brenner, E., & Smeets, J. B. J. (2017). Matching locations is not just matching sensory representations. *Experimental Brain Research*, *235*, 533–545. http://doi.org/10.1007/s00221-016-4815-1

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

Manzone, J., Davarpanah Jazi, S., Whitwell, R. L., & Heath, M. (2017). Biomechanical constraints do not influence pantomime-grasping adherence to Weber's law: A reply to Utz et al. (2015). *Vision Research*, *130*, 31–35. http://doi.org/10.1016/j.visres.2016.09.018

Poulton, E. C. (1967). Population norms of top sensory magnitudes and S. S. Stevens' exponents. *Perception & Psychophysics*, *2*(7), 312–316. http://doi.org/10.3758/BF03211049

R Core Team. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org

Schenk, T., Utz, K. S., & Hesse, C. (2017). Violations of Weber's law tell us more about methodological challenges in sensorimotor research than about the neural correlates of visual behaviour. *Vision Research*, *140*, 140–143. http://doi.org/10.1016/j.visres.2017.05.017

Searle, S. R., Casella, G., & McCulloch, C. E. (1992). *Variance components*. New York, USA: Wiley.

Shadmehr, R. (2017). Distinct neural circuits for control of movement vs. holding still. *Journal of Neurophysiology*, *117*(4), 1431–1460. http://doi.org/10.1152/jn.00840.2016

Smeets, J. B. J., & Brenner, E. (1999). A new view on grasping. *Motor Control*, *3*(3), 237–271.

Smeets, J. B. J., & Brenner, E. (2008). Grasping Weber's law. *Current Biology*, *18*(23), R1089–R1090.

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.

Stevens, S. S., & Stone, G. (1959). Finger span: Ratio scale, category scale, and JND scale. *Journal of Experimental Psychology*, *57*(2), 91–95. http://doi.org/10.1037/h0048829

Teghtsoonian, M., & Teghtsoonian, R. (1971). How repeatable are Stevens's power law exponents for individual subjects? *Perception & Psychophysics*, *10*(3), 147–149. http://doi.org/10.3758/BF03205774

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

Teghtsoonian, R. (1973). Range effects in psychophysical Scaling and a revision of Stevens' law. *The American Journal of Psychology*, *86*(1), 3–27.

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 Kemenade, B. M., Arikan, B. E., Kircher, T., & Straube, B. (2016). Predicting the sensory consequences of one's own action: First evidence for multisensory facilitation. *Attention, Perception & Psychophysics, 78,* 2515–2526. http://doi.org/10.3758/s13414-016-1189-1

Whitwell, R. L., & Goodale, M. A. (2017). Real and illusory issues in the illusion debate (Why two things are sometimes better than one): Commentary on Kopiske et al. (2016). *Cortex*, *88*, 205–209. http://doi.org/10.1016/j.cortex.2016.06.019

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science*, *269*, 1880–1882. http://doi.org/10.1126/science.7569931