Static Weight Perception Through Skin Stretch and Kinesthetic Information: Detection Thresholds, JNDs, and PSEs

Femke E. van Beek, *Member, IEEE*, Raymond J. King, Casey Brown, Massimiliano Di Luca, *Member, IEEE*, and Sean Keller, *Member, IEEE*

***Abstract*—We examined the contributions of kinesthetic and skin stretch cues to static weight perception. In three psychophysical experiments, several aspects of static weight perception were assessed by asking participants either to detect on which hand a weight was presented or to compare between two weight cues. Two closed-loop controlled haptic devices were used to present cutaneous and kinesthetic weights, in isolation and together, with a precision of 0.05 g. Our results show that combining skin stretch and kinesthetic information leads to better weight detection thresholds than presenting uni-sensory cues does. For supra-threshold stimuli, Weber fractions were 22-44%. Kinesthetic information was less reliable for lighter weights, while both sources of information were equally reliable for weights up to 300 g. Weight was perceived as equally heavy regardless of whether skin stretch and kinesthetic cues were presented together or alone. Data for lighter weights complied with an Optimal Integration model, while for heavier weights, measurements were closer to predictions from a Sensory Capture model. The presence of correlated noise might explain this discrepancy, since that would shift predictions from the Optimal Integration model towards our measurements. Our experiments provide device-independent perceptual measures, and can be used to inform, for instance, skin stretch device design.**

***Index Terms*—Psychophysics, weight perception, skin stretch, tactile, kinesthetic, device design guidelines.**

I. Introduction

Weight perception has been a topic of scientific enquiry since the origin of the field of psychophysics [1]. Numerous researchers have measured the precision (usually expressed as *JND* - just noticeable difference—or *threshold*) and accuracy (usually expressed as *PSE* - point of subjective equality - or *bias*) of static weight perception in various tasks, as can be seen in a review by Jones [2]. Various attempts have been made to disentangle the contributions of the two primary sources of information in haptic weight perception: kinesthetic and tactile [3]. Kinesthetic mechanoreceptors encode information on the state of muscles, tendons, and joints, while the four tactile mechanoreceptors respond to skin deformations. It has proven difficult, however, to analyze the contributions of the two sources of information in isolation. One of the limitations has been the absence of a device that allows for independent control of kinesthetic and tactile cues. In this experiment, we used a closed-loop controlled haptic device to render weight with a precision of 0.05 g. The device simulated the static weight of a virtual object held in a stationary pinch grasp by exerting force on the end-effector held between the fingers. This approach allowed us to greatly attenuate kinesthetic and tactile weight cues independently. For tactile information specifically, the major cue in such an interaction is most likely tangential deformation (i.e., shearing) of the skin on the fingertips [4]. Throughout this paper, we will refer to this type of information as *skin stretch*.

Research has shown that tactile cues play a big role in grip and load forces exerted during object interactions. When participants are asked to lift objects with locally anesthetized fingers, which attenuates tactile sensation, they struggle to maintain a proper background level of grip force [5], [6] and more often drop objects [6]. Chronically deafferent participants show incorrectly timed grip force profiles with a bigger safety margin than healthy controls do when moving objects point-to-point [7], [8]. The numerous recent attempts to simulate virtual object weight using skin stretch devices (e.g., [4], [9]–[11]) point to the importance of understanding how this cue specifically affects the perceptual aspect of the motor-control loop. Some of the literature on device design also present results of psychophysical experiments that assess the relation between the skin stretch and perceived weight [12], [13]. Only a few studies assessed the contribution of both skin stretch and kinesthetic information to providing the sensation of weight [13]–[15]. For example, Minamizawa *et al.* [13] combined a force feedback device strapped to the hand with devices attached to the fingertips. Their results, although obtained from only three participants, suggest that skin stretch information is more precise (i.e., results in lower JNDs) than

Manuscript received November 12, 2019; revised June 8, 2020; accepted July 5, 2020. Date of publication July 15, 2020; date of current version March 19, 2021. *(Massimiliano Di Luca and Sean Keller contributed equally to this work.) (Corresponding author: Femke E. van Beek.)*

Femke E. van Beek was with the Department of Physics and Astronomy, California State University, Los Angeles, CA 98052, USA. She is now with Facebook Reality Labs Redmond, WA 98052 USA (e-mail: femke.van.beek@oculus.com).

Raymond J. King, Casey Brown, and Sean Keller are with Facebook Reality Labs, Redmond, WA 98052 USA (e-mail: raymond.king@fb.com; casey.brown@oculus.com; seankeller@fb.com).

Massimiliano Di Luca was with Facebook Reality Labs, Redmond, WA 98052 USA. He is now with the Centre for Computational Neuroscience and Cognitive Robotics, School of Psychology University of Birmingham, Birmingham B15 2TT U.K. (e-mail: m.diluca@bham.ac.uk).

Digital Object Identifier 10.1109/TOH.2020.3009599

kinesthetic information for weights of 100 g and below, while kinesthetic information is more precise than skin stretch for weights of 300 g and above. The results obtained by Matsui *et al.* [14], using a combination of a force feedback device and a shear plate acting on a single finger, point in the same direction. These studies all induced skin stretch using a wearable device, which therefore necessarily causes parasitic reaction forces on the finger or hand. The force pattern as a whole is thus quite different from the forces that induce skin stretch when lifting a real object. Another limitation of these types of studies is that the magnitude of skin stretch stimulation is often either not reported in units of mass, or it is not continuously monitored. For example, in Webster *et al.* [16], the stimulation provided by the rotating ball display is expressed in revolutions per minute. The studies of Matsui *et al.* and Minamizawa *et al.* [13], [14] are examples of the latter problem, as the force producing the skin stretch cue was not being monitored during the experiments. This is problematic because the actual force exerted by the device, and thus the actual amount of skin stretch, is unknown and could even vary during the experiments (i.e., due to power losses during force transfer from motor to end-effector or due to device slip at the finger pad). Therefore, the results obtained in these papers [13], [14] cannot be generalized and can only be replicated using the specific devices described in the papers. Giachritsis *et al.* mitigated this problem by using real weights to assess the precision of both types of information [15]. To separate the cues, they used thimbles to eliminate tactile information and a hand rest to reduce kinesthetic contributions. Their results show that the combined percept seemed to be more precise than the single cues, suggesting that skin stretch and kinesthetic cues are integrated when assessing the weight of a real object. Their results are inconclusive about a difference between tactile and kinesthetic information, possibly due to the limited number of subjects (n was 5). Although their method has merit, the use of real weights introduces confounding factors, such as the influence of different lifting styles on the magnitude of inertial forces and the limits in stimulus range imposed by having to manufacture each stimulus. Therefore, we used an approach similar to Giachritsis *et al.* [15], but we used carefully-controlled virtual weights with a slow onset and offset, and asked participants to not move their hands during weight presentation. This allowed us to study the (static) perception of weight, without having to consider effects of inertia.

In the current experiment, we investigated how skin stretch and kinesthetic information contribute to the final percept of static weight, while carefully controlling the presentation of the physical weight cues that were provided through both types of information. This was done following the method of Giachritsis *et al.* [15], using a hand rest for the tactile condition and thimbles for the kinesthetic condition. Forces were presented using closed-loop controlled, externally grounded force feedback devices, with a precision of 0.05 g. This setup allowed us to assess whether the kinesthetic and skin stretch cues were integrated using a precise and accurate setup that did not cause parasitic forces. The maximum likelihood estimation (MLE) framework has been used to quantitatively explain the increase in precision when two sources of information are present by showing how the means and variances of the information can be pooled, i.e., statistically optimally integrated [17], [18]. We used this framework to make predictions on optimal performance in the combined cue condition. As an alternative model, we tested a Sensory Capture model, in which the most reliable modality is the only one that is represented in the multi-sensory percept, and information from the other modality is disregarded [19]. Finding an underlying model would allow making predictions about stimuli that were not tested in this experiment.

To systematically map the space of the contribution of skin stretch and kinesthetic cues to the perception of static weight, we investigated Detection Thresholds (DTs), Just Noticeable Differences (JNDs) and Points of Subjective Equality (PSEs) in three experiments. In all experiments, we used a tactile condition in which kinesthetic information was greatly attenuated, a kinesthetic condition in which tactile information was greatly attenuated, and a combined condition in which both types of information were present. The detection thresholds measured in Experiment 1 provide us with a guideline for the minimum amount of force that a skin stretch and/or a kinesthetic device needs to be able to exert to provide a detectable percept of static weight. In Experiment 2, we studied the supra-threshold precision of weight perception for the single and combined cues by measuring the JNDs. This experiment was performed to understand how many perceivably different levels of force are present in a weight range, so they describe how many noticeably different weights a device can present. For both detection thresholds and discrimination thresholds, optimal performance for combined cues was predicted from two candidate models (Optimal Integration and Sensory Capture), which allowed us to compare measurements and predictions. We have presented the results of Experiments 1 and 2 in prior work [20]. In Experiments 3a and 3b, we directly compared the perceived magnitude of the single and combined cues by measuring the PSEs between them. This experiment was done to understand if providing one of the cues alone can lead to the same perceived weight as providing congruent cues would. If single cues were enough for rendering virtual weight, that would simplify device design. Together, these three experiments provide an overview of the contribution of skin stretch and kinesthetic information to the perception of weight, and provide force-based guidelines for rendering weight of virtual objects through one or both types of information.

II. Material and Methods

A. Participants

In Experiment 1, 10 participants performed the study, 8 males and 2 females. Three additional participants completed this experiment, but due to technical issues their data sets were not recorded correctly and could not be used. The participants were 34 ± 5 years old (mean $\pm$ standard deviation), and all were right handed.

In Experiment 2, 19 participants performed the study, 6 males and 13 females. Two additional participants completed the

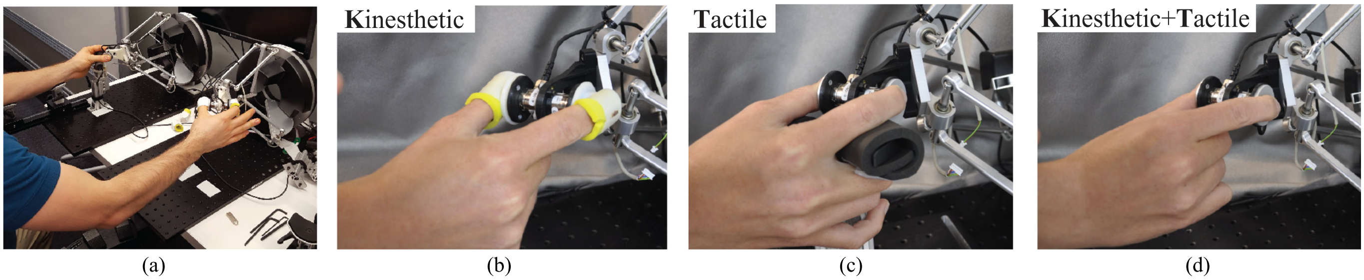

Fig. 1. Overview of the setup and the three experimental conditions. (a) Overview of the entire setup, showing the arm supports, hand rests, thimbles, and haptic devices. The participant is experiencing condition K with his right hand and condition T with his left hand. (b) Close-up of the kinesthetic condition (K), in which the participant's elbow is supported while he actively holds up his forearm and hand. The custom-fitted thimbles attenuate tactile information. (c) Close-up of the tactile condition (T), in which the participant's elbow and forearm are supported, and his thumb and index finger are resting on a padded finger rest, while holding the end-effector with his bare fingers. In this condition, kinesthetic information is attenuated. (d) Close-up of the combined cue condition (KT), in which the participant holds the end-effector with his bare fingers and only his elbow is supported.

experiment, but their performance never exceeded chance level, so their data was not used in the final analysis. The participants were 35 ± 10 years old, and 17 of them were right handed while 2 were left handed.

In Experiment 3a, 13 participants performed the study, 11 males and 2 females. The participants were 33 ± 7 years old, and 11 of them were right handed while 2 were left handed. In Experiment 3b, 10 participants performed the study, 7 males and 3 females. The participants were 32 ± 8 years old, and 7 of them were right handed while 3 were left handed.

All participants gave written informed consent prior to taking part and were naive to the purpose of the experiment. None of them had any history of neurological disorders, and all were compensated for their time. All experiments were approved by WIRB, and were carried out in accordance with the relevant guidelines and regulations.

B. *Experimental setup and protocol*

For all three experiments, we used a setup comprised of two 3-DoF haptic devices (Omega 3.0, ForceDimension, customized in a similar way as discussed in [21]), each equipped with two 6-axis force-torque sensors at the end-effector (Nano17, ATI) to allow closed-loop control of the rendered stimuli. The precision of force rendering was 0.05 g once the system had reached target force level. Participants held the custom end-effectors of the Omega in a pinch grasp between their thumbs and index fingers. The force-torque sensors were placed directly underneath the aluminum finger plates. Three conditions were used, see Fig. 1 for an overview of the setup and conditions. The first condition was kinesthetic (K): a uni-sensory condition in which participants wore the custom thimbles to ensure that the majority of the tactile weight information was removed. A variety of 3D printed thimbles with different thumb angles were used to ensure that each participant could comfortably hold the end effector. The thimbles were padded with participant-adjustable foam to ensure a tight fit on all finger sizes. In this way, the thimbles provided pressure around the fingers, which prevented the skin from stretching and thus removed most of the task-relevant tactile information. Participants rested their elbows on a support, while holding up their forearms and hands. The second condition was tactile (T): a uni-sensory condition in which participants rested their elbows and forearms on a support, while resting their thumbs and index fingers on a padded finger rest, such that most of the kinesthetic information was removed. The third condition was kinesthetic-tactile (KT): a multi-sensory condition in which participants held the device with bare fingers, while their arm posture was the same as in condition K. In Experiments 1 and 2, the conditions were always the same for both hands. In Experiment 3, one of the hands was always presented with condition KT, while the other hand was presented with a uni-sensory condition.

Throughout the experiment, participants wore headphones playing white noise to cancel any possible auditory cues. They were asked to provide their responses using foot pedals. In Experiment 1, they wore custom glasses that prevented them from seeing their hands. In Experiments 2 and 3, participants wore a head-mounted display (Rift, Oculus VR) for this purpose. Visual information, presented on a screen (Experiment 1) or the HMD (Experiments 2 and 3), was used to guide participants to the center of the workspace, such that they would always start a trial from the same position. To do this, the desired positions were represented with larger blue wire-frame spheres, while the actual hand positions were depicted with smaller solid white spheres, one for each hand. The participants were instructed to position the smaller spheres inside the larger spheres by moving their hands. As soon as they achieved this, visual feedback was removed and the trial started.

The participants' task was to hold the instrumented end-effectors of the two haptic devices as stationary as possible and to compare the sensation of weight between their two hands. A 2-alternative forced choice (2AFC) task was used, so participants had to choose which side was perceived to be heavier and indicate this using foot pedals. The stimuli comprised only downward forces, and no inertial effects were rendered in response to participants' movements. Each cue was initiated with a linear increase (3 s in Experiment 1, 2 s in Experiments 2 and 3). Once the first stimulus reached the stationary force level, the second stimulus was ramped up 0.5 ± 0.05 s (mean $\pm$ standard deviation) after. One second after the second force stimulus had reached its constant level, participants were prompted for a response by a sound. They

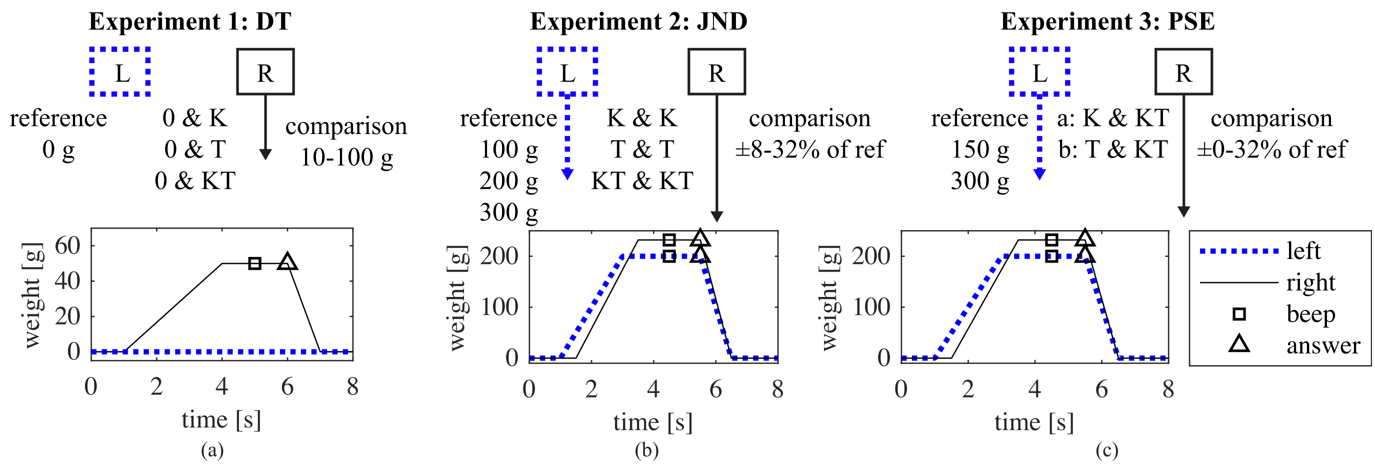

Fig. 2. Illustration of the experimental protocol. For illustration purposes, all subplots depict the left hand (L, dashed blue) receiving the reference force, and the right hand (R, solid black) receiving the comparison force. In the experiments, the side at which the reference force was presented was counterbalanced. The reference and comparison force ranges are indicated for all experiments. The letters in the middle of each force pair indicate the combination of the conditions that the hands received. The force profiles at the bottom illustrate the flow of the trial. One second after both the forces had reached their maximum force level, a beep indicated that participants could provide their answer. Upon answering, the force was ramped down in 1 s. a) Experiment 1, in which a 3 s force ramp was used and only one hand was presented with a force. b) Experiment 2, in which both hands received a force cue, which was ramped up in 2 s in a staggered fashion. Both hands always received cues of the same condition. c) Experiment 3, one hand always received a KT cue, while the other hand received a K cue (Experiment 3a) or a T cue (Experiment 3b). As in Experiment 2, the force cue was ramped up in 2 s in a staggered fashion.

were instructed to keep their hands as steady as possible and to base their perception on the stationary force level. Both forces remained constant until participants provided a response, after which they were ramped down (1 s). By staggering the ramp-up force while keeping the ramp time constant, participants could neither use the ramp time nor directly compare the ramp slope as an indication of the final force magnitude. In Experiment 1, a slower ramp was used to eliminate any additional cues about the presence of a stimulus, like potential hand movement, which would have confounded the results. For an illustration of the protocol, see Fig. 2.

In Experiment 1, we measured force detection thresholds for the three different cue types. The experiment consisted of three blocks and took about one hour in total. In each block the threshold of a single cue was measured, by presenting a force cue on one hand only, and asking participants to indicate which hand received a cue. The stimuli were weights of 10, 20, 30, 40, 50, 60, 80, or 100 g. The side at which the force was applied was pseudo-randomized. Each stimulus pair was repeated 12 times, resulting in 96 trials per condition.

In Experiment 2, we measured supra-threshold precision (JND) of weight perception for the three different cue types. The experiment consisted of nine blocks — three reference weights for three cue types — and took about three hours, divided over three one-hour sessions. In each block the perceptual precision of a single cue type was measuered. The same type of cue was presented to each hand, and participants were asked to indicate which hand received the heavier cue, at 3 reference weights: 100, 200, and 300 g.The comparison stimuli deviated from the reference weight by ± 8, 16, 24, or 32%. The side at which the reference weight was applied was pseudo-randomized. Each stimulus pair was repeated 12 times, resulting in 96 trials per reference weight.

In Experiments 3a and 3b, we measured the perceived magnitude of each of the uni-sensory cues (conditions K and T) with respect to the perceived magnitude of the multi-sensory cue (condition KT), i.e., the PSE. Both experiments were divided into four experimental blocks, and each experiment took about one hour. The PSE is defined as the weight at which the proportion of responses is 0.5. From this value the bias could be determined, which is the difference between the PSE and the reference stimulus, and it indicates the difference in perceived magnitude between the tested conditions. In Experiment 3a, participants were asked to compare weights provided through conditions KT and K. In Experiment 3b, they were asked to compare conditions KT and T. In both experiments, two different reference weights were used: 150 and 300 g. The comparison stimuli could be identical to the reference weight, or deviate from it by ± 8, 16, 24, or 32%. Each stimulus pair was repeated 16 times, presented counterbalanced between two arm configurations, resulting in 144 trials for each reference weight.

In all experiments, the order of the force cues was randomized and the order of reference and comparison stimulus was counterbalanced. The order of the blocked conditions was counterbalanced between participants. In Experiment 3, since two different types of stimuli were compared, each condition was divided into two experimental blocks, between which the configuration of the setup was changed to switch the cues between hands. Moreover, trials where the multi-sensory stimulus served as the reference and trials where it served as the comparison were pseudo-randomized. At the start of all experiments, 12 training trials were performed for familiarization with the procedure and the device. Participants did not receive feedback during the training trials.

C. Statistical Analyses and Models

For Experiments 1 and 2, we calculated the proportion with which the comparison stimulus was chosen as being the heavier stimulus, as a function of the weight of the comparison

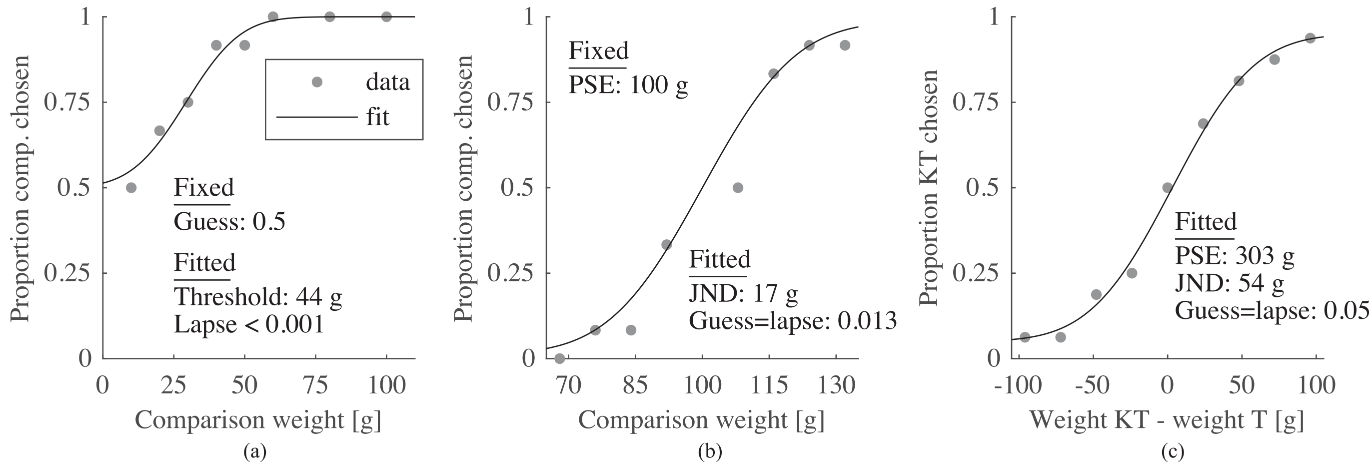

Fig. 3. Representative example data sets and fits for all experiments. (a) The detection threshold (Experiment 1) is the weight at which the proportion of responses was 0.84 in the fitted function. (b) The JND (Experiment 2) is the difference between the weight at which the proportion of responses was 0.84 in the fit and the reference weight. (c) The PSE is the weight at which the proportion of responses was 0.5 in the fitted function and the stimuli are thus perceived as perceptually equal. Note that in this subfigure, the data are represented slightly differently (the proportion that KT is picked, instead of the proportion that the comparison is picked, is plotted). This was needed to ensure that all data could be analyzed together.

stimulus. Representative examples can be seen in Figs. 3 a and b. For Experiment 3, the proportion of responses was calculated as the proportion with which the KT stimulus was chosen as being the heavier stimulus, expressed as a function of the weight of the uni-sensory stimulus (K in Experiment 3a, and T in Experiment 3b) subtracted from the multi-sensory stimulus (KT). An example of this type of data set can be found in Fig. 3 c. The slightly different approach in Experiment 3 was taken to make sure that all data could be converted into one psychometric curve, since half of the time the single stimulus served as the standard stimulus, while the other half of the time the combined stimulus had this role. In this convention, a positive (negative) bias indicates that the uni-sensory (multi-sensory) stimulus is perceived as being heavier when the physical stimuli are equally heavy.

For all experiments, a psychometric function (ψ) was fitted to the proportion of responses of each participant and condition by using the maximum likelihood procedure in the Palamedes toolbox [22] (see Fig. 3 for a typical example). A cumulative normal distribution was fitted as shown below, from which the precision (JND) and accuracy (PSE) could be determined:

$$\psi(x;\alpha,\beta,\gamma,\lambda) = \gamma + (1-\gamma-\lambda)\frac{\beta}{\sqrt{2\pi}} \times \int_{-\infty}^{x} exp\left(-\frac{\beta^2(t-\alpha)^2}{2}\right)dt. \quad (1)$$

In this equation, x is the presented weight, β is the reciprocal of the standard deviation, γ is the guess rate (i.e., the lower asymptote), λ corresponds to the lapse rate, and α corresponds to the PSE when λ and γ are equal. After obtaining fits for all these parameters, the inverse of the psychometric function was used to find the presented weight that corresponded to the proportion of interest (usually 0.5 for the PSE, and the difference between 0.84 and 0.5 for the JND).

In Experiments 1 and 3, the PSE was fitted, while in Experiment 2 it was constrained to be at the reference weight. In all experiments, the slope was fitted as a free parameter, and the lapse rate was fitted in the range [0,0.05]. The guess rate was set to 0.5 in Experiments 1 and 3, while it was constrained to be the same as the lapse rate in Experiment 2. In Experiment 1, we calculated the detection threshold by determining the weight at which the proportion of responses was 0.84. In Experiment 2, we used the difference between the weight at which the proportion of responses was 0.84 and the reference weight for assessing the JND. In Experiment 3, we calculated the PSE by determining the comparison stimulus weight at which the proportion of the KT stimulus being chosen was 0.5. To determine the goodness-of-fit of the psychometric curves, the model used to fit the data was compared to a "saturated" model in 1000 simulations. For a more detailed description of this procedure, see Kingdom and Prins [23]. A goodness-of-fit of less than 0.05 was considered to be an outlier, and was removed from the analysis.

For predicting multi-sensory performance using the Optimal Integration and Sensory Capture models, measured uni-sensory data were used. Note that no parameters were fitted, we are only predicting performance based on measurements. For the Optimal Integration model, predictions of detection thresholds were made using the Pythagorean Theorem [24], [25], which states:

$$d'_{KT} = \sqrt{d'^2_K + d'^2_T} \quad (2)$$

for which d' was calculated from the response proportions using the Palamedes toolbox. For predicting multi-sensory variances from uni-sensory measurements, we take the combined cue as a weighted sum $wK + (1-w)T$ as in Oruç *et al.* [26], in which the variance of the combined cue is:

$$\sigma^2_{KT} = w^2\sigma^2_K + (1-w)^2\sigma^2_T + 2w(1-w)\rho\sigma_K\sigma_T \quad (3)$$

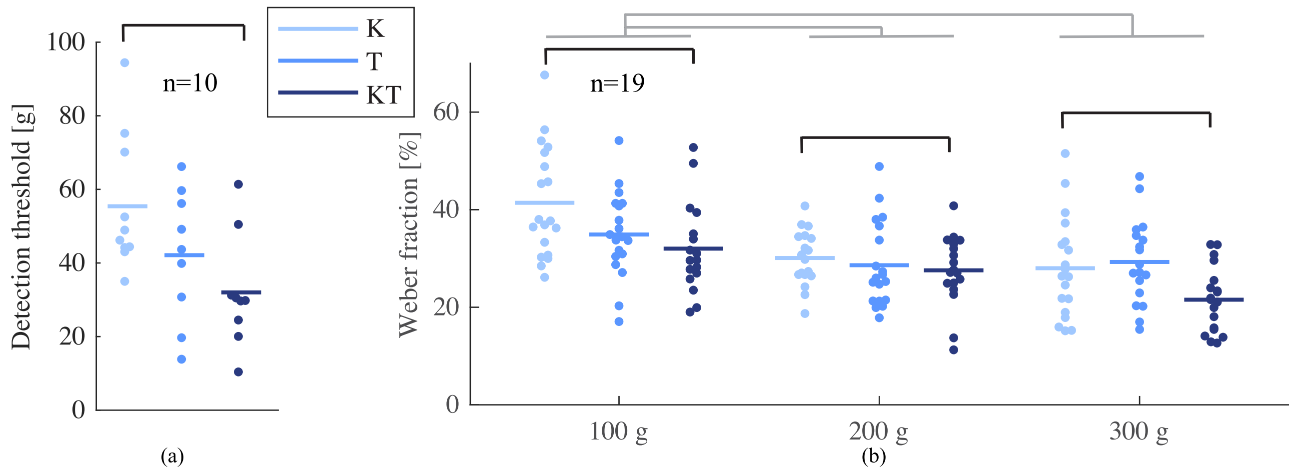

Fig. 4. Detection thresholds and JNDs for conditions K, T, and KT, with each marker representing one participant, and horizontal lines indicating mean values. Grey brackets illustrate that for both experiments, post hoc comparisons showed that conditions K and KT were significantly different. (a) Detection thresholds for Experiment 1, for which a significant main effect of 'condition' was found. (b) Weber fractions for Experiment 2, with significant main effects 'condition' and 'reference weight'. The gray brackets illustrate that post hoc tests revealed JNDs for a 100 g reference weight being significantly higher than those for 200 g and 300 g references. Note that the kinesthetic condition does not perform well for light weights.

where σ^2_{KT} is the variance of KT, σ^2_K and σ^2_T are the uni-sensory variances, w is the weight of K, and ρ is the correlation in the noise on the cues. The Optimal Integration model [17] (which is equivalent to the optimal Weighted Summation model [19]), constrains the cues to be uncorrelated, in which case the optimal weights are the reciprocals to the variance of the cues, which reduces equation 3 to:

$$\sigma^2_{KT} = \frac{\sigma^2_K\ \sigma^2_T}{\sigma^2_K + \sigma^2_T} \tag{4}$$

for which JND is $\sqrt(2)\sigma_{KT}$. Thus, using the Optimal Integration model, the multi-sensory condition is always predicted to have lower JNDs than either of the uni-sensory conditions.

For predicting DTs and JNDs using the Sensory Capture model, the following equations were used [19]:

$$\mathrm{DT_{KT}} = \min(\mathrm{DT_K}, \mathrm{DT_T}) \tag{5}$$

$$\mathrm{JND_{KT}} = \min(\mathrm{JND_K}, \mathrm{JND_T}) \tag{6}$$

The Sensory Capture model thus results in the multi-sensory condition having the same noise level as the best performing uni-sensory condition.

For statistically testing our hypotheses, we performed parametric and Bayesian repeated measures ANOVAs, t-test and Wilcoxon signed-rank tests. For ANOVAs, Greenhouse-Geisser correction was used when the sphericity criterion was not met, and Bonferroni-corrected post-hoc tests were performed for all significant main effects. All t-test results employed two-tailed probabilities. For ANOVAs and t-tests, an α level of 0.05 was used. For Bayesian statistics, BF_{01} were used, which represents the degree to which the data supports a hypothesis (i.e., the presence of a main effect) [27]. A $\mathrm{BF}_{01} < 0.067$ ($\mathrm{BF}_{01} > 3$) is considered strong evidence that the main effect is present (absent).

Movement data were analyzed by calculating the difference between the lowest and highest vertical position of the end effector on each trial. Force data were analyzed by calculating the average inward force exerted on the index and thumb force sensor by taking the magnitude of the inward component for each sensor and adding them, thus representing the average squeeze force exerted during the phase of the trial where the force had reached the plateau level. For both the movement and force data, this yielded two data points with their own presented weights per trial: one for the right and one for the left hand. To assess the effect of weight on movement and force data, linear regressions with intercept and slope were calculated for each participant, condition, and experiment.

III. Results

In Experiment 1, we measured detection thresholds for conditions K (mainly kinesthetic cues), T (mainly tactile cues), and KT (both cues present), as shown in Fig. 4 a. Two of the 30 psychometric curves were discarded for not meeting the goodness-of-fit criterion. The resulting thresholds were 55 ± 6 g for K, 42 ± 6 g for T, and 32 ± 5 g for KT (mean $\pm$ standard error). All statistics are shown in Table I. A one-way repeated measures ANOVA on the measured thresholds showed a significant effect of condition. Post hoc tests show that the KT condition differed significantly from the K condition, whereas the other conditions did not differ significantly from each other. The threshold for KT predicted from modeling was 31 ± 5 g for Optimal Integration and 43 ± 5 g for K for Sensory Capture. Two paired sample t-tests, with Bonferroni-corrected αs of 0.025, showed that KT measurements did not differ from predictions of the Optimal Integration model, while they did differ from the Sensory Capture ones, as shown in Fig. 5 a.

In Experiment 2, we measured discrimination thresholds for supra-threshold stimuli in conditions K, T, and KT, by asking participants to indicate which hand received the heavier cue.

TABLE I
STATISTICS OF DETECTION THRESHOLDS IN EXPERIMENT 1

	F/t	d.f.	p	η_p^2	BF$_{01}$
ANOVA+post hoc					
condition	F=13	2,14	<0.001*	0.65	0.021
KT - K	t=4.7	-	0.007*	-	-
KT - T	t=2.7	-	0.083	-	-
K - T	t=2.5	-	0.11	-	-
Paired sample t-test					
OI$_{KT}$ - meas$_{KT}$	t=-0.37	7	0.72	-	2.8
SC$_{KT}$ - meas$_{KT}$	t=3.1	8	0.015*	-	0.21

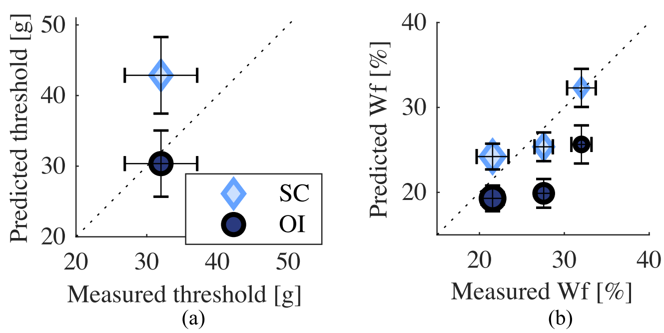

Fig. 5. Correlations between multisensory (KT) threshold measured in the experiments and predicted from the Sensory Capture (SC) and Optimal Integration (OI) models. Error bars indicate ± 1 standard error. (a) Correlations for detection thresholds in Experiment 1, in which the Optimal Integration model makes significantly better predictions than Sensory Capture does. (b) Correlations for JNDs for the 3 reference weights, with increasing marker size indicating increasing reference weights (small: 100 g, medium: 200 g, large: 300 g). Here, both models fail to predict multi-sensory thresholds correctly.

Both hands received the same type of cue, and a range of test stimuli were compared to reference stimuli of 100, 200, and 300 g. The JNDs for the three reference weights and conditions are shown in Fig. 4 b, with five of the 171 fits being discarded because of not meeting the goodness-of-fit criterion. All statistics are shown in Table II. A two-way repeated measures ANOVA on the JNDs with the within-subjects factors ‘condition’ and ‘reference weight’ showed a significant effect of both weight and condition, and the interaction effect was significant too. Post hoc testing of the ‘condition’ and ‘weight’ factors showed that conditions K and KT differ significantly, while the others do not. Reference weights of 100 g differ from 200 and 300 g, while 200 and 300 g do no differ significantly from each other. To test the predictions from the Optimal Integration and Sensory Capture models (see Fig. 5 b), two separate repeated measures ANOVAs were performed, in which the measured and predicted KT thresholds were compared, while using weight as the second ‘within-subject’ factor. The measured JNDs differed significantly from predictions from Optimal Integration, while they did not differ significantly from predictions from Sensory Capture.

In Experiments 3a and 3b, we measured if the perceived weight magnitudes differed between uni-sensory and multi-sensory weight cues, by presenting participants with a uni-sensory cue to one hand and a multi-sensory cue to the other hand. Participants were asked which hand received a heavier weight, and reference weights of 150 and 300 g were used. For Experiment 3a (KT to K), three of the 26 fits yielded a

TABLE II
STATISTICS OF JUST NOTICEABLE DIFFERENCES IN EXPERIMENT 2

	F/t	d.f.	p	η_p^2	BF$_{01}$
ANOVA+post hoc					
condition	F=4.5	2,26	0.021*	0.26	0.048
weight	F=13	2,26	<0.001*	0.51	<0.0067
condition x weight	F=2.7	4,52	0.040*	0.17	1.2(BF$_{12}$)
KT - K	t=3.2	-	0.008*	-	-
K - T	t=0.94	-	1.0	-	-
KT - T	t=2.2	-	0.099	-	-
100g - 200g	t=3.4	-	0.005*	-	-
100g - 300g	t=4.8	-	<0.001*	-	-
200g - 300g	t=2.0	-	0.15	-	-
ANOVA					
OI$_{KT}$ - meas$_{KT}$	F=27	1,13	<0.001*	0.67	<0.0067
SC$_{KT}$ - meas$_{KT}$	F=0.028	1,15	0.87	0.002	4.6

TABLE III
STATISTICS OF POINTS OF SUBJECTIVE EQUALITY IN EXPERIMENT 3

	t	d.f.	p	BF$_{01}$
1-sample t-test				
3a: 150 g - 0	-1.5	9	0.18	1.4
3a: 300 g - 0	-1.6	12	0.14	1.3
3b: 150 g - 0	-0.55	8	0.60	2.7
3b: 300 g - 0	0.045	9	0.97	3.2
Paired sample t-test				
3a: 150g - 300g	-0.026	9	0.98	3.2
3b: 150g - 300g	-0.008	8	0.99	3.1

goodness-of-fit that did not meet the criterion and were discarded; for Experiment 3b (KT to T), this was the case for one of the 20 fits. All statistics are shown in Table III. For Experiment 3a, we found non-significant biases of -7.5 $\pm$ 5.2 g for a 150 g reference and -12 $\pm$ 7.3 g for a 300 g reference, as shown in Fig. 6 a. For Experiment 3b, the biases were also non-significant and corresponded to -1.6 $\pm$ 2.9 g for a reference of 150 g and 0.31 $\pm$ 6.8 g for a reference of 300 g. A positive (negative) bias indicates that the uni-sensory (multi-sensory) stimulus is perceived as being heavier when the physical stimuli are equally heavy. Paired sample t-tests failed to show a difference between the biases for the different reference weights. For Experiment 3a, there was a modest but significant correlation between the two biases measured for each participant, as depicted in Fig. 6 b (R^2 = 0.50, p = 0.021). However, in Experiment 3b, no significant correlation between the two biases was found, as shown in Fig. 6 d (R^2 = 0.033, p = 0.64).

To investigate whether the perceptual results had an origin in the way participants used their hands, we looked at the movement and grip force data across experiments and conditions, as shown in Fig. 7. All statistics are shown in Table IV. Both the grip force data (Fig. 7 a) and the movement data (Fig. 7 b) show a gradual increase with presented weight, which is confirmed by the slopes of a regression of the force and movement data being significantly larger than 0 in a Wilcoxon signed-rank test. An ANOVA with fixed factors ‘experiment’ and ‘condition’ revealed that the grip force intercepts show an effect of condition, with condition K having a larger intercept than conditions T and KT. There was no significant effect of experiment on the intercepts. The grip force slopes show an effect of experiment, with experiment 1 having

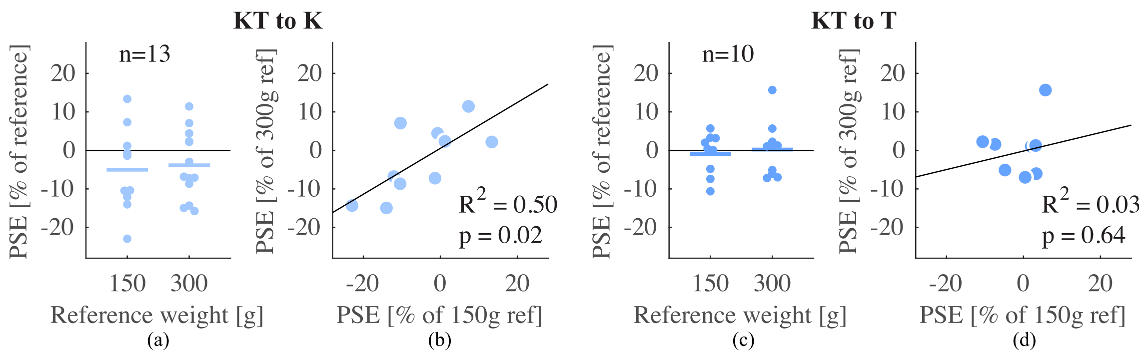

Fig. 6. Results for (a,b) Experiment 3a (showing KT to K) and (c,d) Experiment 3b (showing KT to T). (a,c) Biases measured in experiments, in which the horizontal lines indicate the mean biases, and each marker indicates one participant. A positive (negative) bias indicates that the uni-sensory (multi-sensory) stimulus is perceived as being heavier when the physical stimuli are equally heavy. None of the biases differ significantly from 0, thus giving no evidence that the weights were perceived differently between the different cues. (b,d) Correlations between the biases of the participants for the two different reference weights. In Experiment 3a, there is a significant correlation, meaning that participants with larger biases for 150 g also have larger biases for 300 g. However, in Experiment 3b, there was no significant correlation between the data sets.

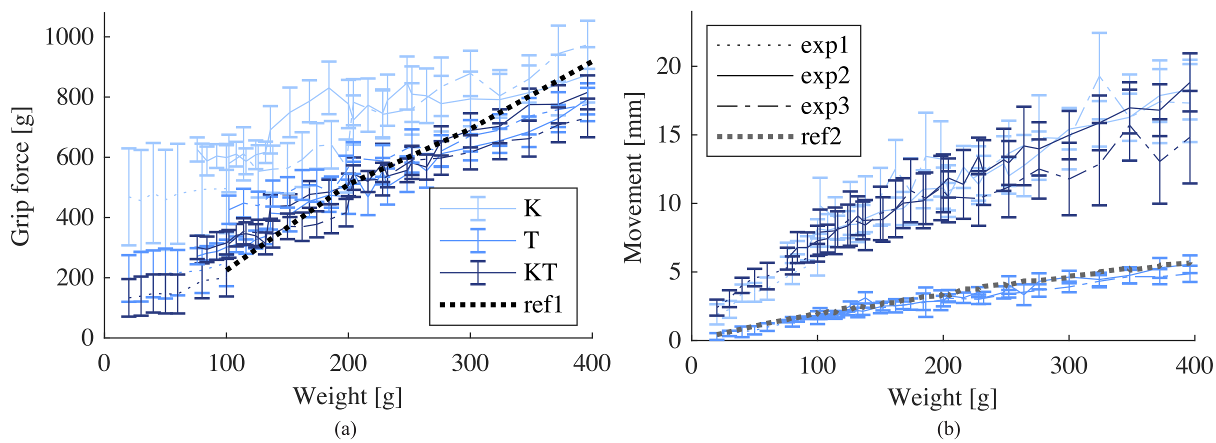

Fig. 7. Median force and position data across all experiments (line styles) and conditions (colors), with error bars indicating ±standard error. (a) Grip force data, representing the mean squeeze force between thumb and index finger per trial. In the detection threshold experiments, grip force did not seem to be modulated strongly by presented force, probably since many forces were imperceivable. For supra-threshold stimuli, grip force increased with presented weight. For smaller weights, participants exerted larger grip forces in the K condition than in the other 2 conditions, which was probably caused by wearing the thimbles. For comparison, data from Westling and Johansson are shown as ref1 in a dotted black line [5]. These data represent measurements from users statically holding silk-padded objects, which indicates that the grip force modulation in conditions T and KT resembles that observed for real objects with comparable material properties. (b) Movement data, representing the downward movement per trial. For the tactile condition, expected vertical movements based on finger impedance measurements from literature (dotted gray line, ref2) match the measured movements closely [29]. Thus, very little movement beyond that caused by skin stretch was present in the tactile condition.

smaller slopes than experiment 2 and 3. There was no significant effect of condition on the grip force slopes. The movement intercepts and slopes show significant effects of both experiment and condition. Experiment 1 has a smaller movement intercept and a larger movement slope than experiment 2 and 3 do. For both movement intercept and slope, condition T showed smaller values than condition K and KT did.

We compared measured force and movement to data obtained in the literature. It is well-known that grip and load forces are tightly coupled when manipulating objects [28]. The required grip force for holding a hard object depends on the friction of the object, which is determined by its material properties. The end effectors on our setup had a smooth aluminum surface, so to approximate this situation, we selected data from literature from users statically holding a silk-padded object [5]. The comparison data, plotted in Fig. 7 a, suggest that grip forces for conditions without thimbles are in the range expected from literature, while they are higher for the lighter weights in condition K. To compare movements measured in condition T to skin stretch data from literature, we used finger pad impedance data from Pataky *et al.* [29]. The authors show that finger pad stiffness depends on grip force, so we calculated expected stiffnesses and resulting predicted movements from measured grip forces, which align well with measured movements in the T condition (see Fig. 7 b).

IV. Discussion

We investigated the contribution of skin stretch and kinesthetic cues to the perception of static weight up to 300 g. Our experiments provide device-independent measures of detection thresholds, precision (JNDs), and accuracy (PSEs) for both types of information alone, and for their combination. Combining skin stretch and kinesthetic information led to better weight detection thresholds than presenting uni-sensory cues did. Weber fractions ranged from 22 to 44% for supra-threshold stimuli. Kinesthetic

TABLE IV
STATISTICS OF MOVEMENT AND FORCE REGRESSIONS

	W/F/t	d.f.	p	η_p^2	BF_{01}
Wilcoxon signed-rank					
force slope - 0	W=8497	129	<0.001*	-	<0.067
movement slope - 0	W=8513	129	<0.001*	-	<0.067
ANOVA+post hoc					
grip force intercept					
condition	F=32	2,121	<0.001*	0.35	<0.067
experiment	F=0.37	2,121	0.70	0.006	13
KT - K	t=7.6	-	<0.001*	-	-
K - T	t=6.0	-	<0.001*	-	-
KT - T	t=1.2	-	0.75	-	-
grip force slope					
condition	F=0.84	2,121	0.43	0.014	4.7
experiment	F=18	2,121	<0.001*	0.23	<0.067
Exp1 - Exp2	t=-5.8	-	<0.001*	-	-
Exp1 - Exp3	t=-4.5	-	<0.001*	-	-
Exp2 - Exp3	t=1.2	-	0.70	-	-
movement intercept					
condition	F=8.7	2,121	<0.001*	0.13	<0.067
experiment	F=6.9	2,121	0.001*	0.10	<0.067
KT - K	t=-0.76	-	1.0	-	-
K - T	t=3.1	-	0.006*	-	-
KT - T	t=4.0	-	<0.001*	-	-
Exp1 - Exp2	t=2.5	-	0.043*	-	-
Exp1 - Exp3	t=-3.7	-	<0.001*	-	-
Exp2 - Exp3	t=-1.6	-	0.36	-	-
movement slope					
condition	F=14	2,121	<0.001*	0.19	<0.067
experiment	F=7.6	2,121	<0.001*	0.11	0.16
KT - K	t=0.93	-	1.0	-	-
K - T	t=5.0	-	<0.001*	-	-
KT - T	t=4.3	-	<0.001*	-	-
Exp1 - Exp2	t=2.9	-	0.014*	-	-
Exp1 - Exp3	t=3.8	-	<0.001*	-	-
Exp2 - Exp3	t=1.3	-	0.70	-	-

information was generally less reliable for lighter weights, whereas for heavier weights up to 300 g the two cues were roughly equally reliable. Weight was perceived as not to change significantly regardless of whether skin stretch and kinesthetic cues were presented together or alone. These results can be used as guidelines for designing skin-stretch devices for presenting weight to users. In this section, we will assess the validity of our experimental setup, compare our measurements to results from literature and to the models we proposed to make an inference as to how these cues are being combined into a final percept of weight, and speculate on the underlying neural mechanisms.

A. Validity of Experimental Setup

To assess the validity of our experimental setup, we investigate the degree to which our setup was able to present tactile and kinesthetic weight cues, separately and in conjunction. We used a force onset ramp that was much slower than lifting an object in a natural setting. For instance, in Buckingham *et al.* [30] participants were asked to lift a 700 g weight in a natural fashion, and they took ~0.35 s on average to surpass the force needed for object lift-off, while our force ramps lasted 2-3 s. Moreover, participants were required to maintain a static posture in our experiment, while one of the most salient cues for weight perception is inertia, which is why the Exploratory Procedure for judging weight is moving an object up and down [31]. We restricted our experiment to static weight with a slow force increase for two reasons. Firstly, we wanted to be able to distinguish between the perception of static weight and of inertia, since they likely both influence the final percept of weight. Secondly, limiting kinesthetic stimulation is even harder in a dynamic task, since that would require a grounded finger rest that would move along with the participant's movements. The slow force increase was helpful to ensure participants kept their hands as static as possible in the kinesthetic condition.

The question now remains to which degree the static weight presentation might be unnatural, which could have affected the experience and thus the external validity of the results. Although we cannot be certain of the participants' subjective experience, we can look at signatures in the data that imply 'normal' behaviour. Such a signature is the tight coupling between lifting force and grip force, which is present in normal lifting of objects, and also in statically holding a weight [5], [28], [32]. Our grip force data, shown in Fig. 7 a, highlight a significant increase with presented weight. Comparing our T and KT data with grip forces from the literature for holding a silk-padded object (since our end-effectors had very smooth aluminum surfaces) shows a comparable range of grip forces [5]. Interestingly, grip forces intercepts were significantly higher in the K condition than in T and KT, but there was no difference between the conditions in grip force slopes. This is consistent with grip force data from the literature on locally anesthetized or chronically deafferent participants, for whom tactile sensations were absent [5], [6], [8]. In those groups, an increase in the background level of grip force, or 'safety margin,' is typical, which is consistent with the higher intercept in condition K. Moreover, deafferent participants are still able to adjust their grip forces to slow changes in load force, albeit with this higher safety margin [7], [8], which is consistent with our slopes not differing significantly between the conditions. We believe that the slopes being smaller for Experiment 1 is simply an artifact of the often even imperceptible forces in that experiment, which makes grip force adjustments over the baseline force less necessary. Taken together, these signatures all point to participants showing motor behavior consistent with naturally holding objects.

Another factor in the setup that could be considered is the increased grip width in the K condition due to wearing thimbles (see Fig. 1). As the thimbles only added 3 mm of grip width on each side of the end effector, the increase of grip width was ~12% compared to conditions T and KT (50 mm). Flanagan *et al.* [33] investigated the effect of grip width on weight perception of real objects, finding that changes of up to 130% did not influence weight sensitivity, while the maximum induced bias was a 3%. Therefore, we believe that our small differences in grip width cannot have affected our results.

Finally, we can assess how well we separated kinesthetic from tactile cues. For the kinesthetic condition, the custom thimbles with participant-specific padding were tight-fitting, so the skin was unable to move and pressure was exerted around the finger constantly, which was unrelated to the presented force. For the tactile condition, we can compare our movement data to measurements from literature. Fig. 7 b shows a good agreement between measured movements in our T condition and movements predicted from literature [29]. These data from literature represent impedance measurements when constraining the finger up to the Proximal Interphalangeal (PIP) joint, so very little movement beyond skin stretch was present when stretching the fingertip tangentially. This suggests that the most important cue in our T condition was indeed skin stretch.

B. Comparison to the Literature

In both Experiment 1 and 2, we see that multi-sensory information is outperforming at least one of the uni-sensory types of information, which is in line with results from literature [13]–[15]. However, the thresholds that we found are worse than those reported for static weight perception in literature. In Experiment 1, the detection threshold in our KT condition (which is closest to 'natural' weight presentation) was 32 g, while literature reports thresholds as low as 10 g [34], [35]. In Experiment 2, our Weber fractions for the KT condition were 20-30%, while literature reports a range between ~9-13% for unconstrained lifting of real objects, which is ~1.5 times worse for static perception of real objects placed in the hands [2]. Giachritsis *et al.* [15] report JNDs of 8% for active lifting of real objects with combined cues, and 10-12% for uni-sensory weight perception. We also see improvements in JNDs for multi-sensory cues compared to uni-sensory cues, but the range of our JNDs is quite different. This difference is probably due to the absence of inertial cues in our experiment, which resembles a very slow placement of a real object on a stationary hand. Thus, our result shows that even in a static situation, the force ramp caused by the inertia of placing a weight on a user's hand is an important cue for weight perception. Therefore, our results can be used to disentangle the contributions of static weight and inertia to weight perception.

Two previous studies have looked at the contribution of skin stretch and kinesthetic cues to passive weight perception [13], [14]. The work by Minamizawa *et al.* [13] is most comparable to our weight ranges. They suggest that tactile information is more precise than kinesthetic information for smaller weights, while for weights of 300 g and heavier, tactile sensitivity is greatly reduced and kinesthetic information becomes the more reliable source. We observe similar trends in our data, with the condition T outperforming the K for small weights while reaching similar performance for larger weights, but the patterns is not as pronounced as in Minamizawa's study. In Experiment 1, detection sensitivity for K is significantly worse than for KT, while T and KT do not differ significantly. The same pattern of performance is found for supra-threshold discrimination in Experiment 2, where K performs worse than KT, which is most obvious for 100 g. For 300 g weights, T and K seem equally reliable, so their relative reliabilities change from those at 100 g weights, but we do not observe the massive deterioration of uni-sensory tactile information that Minamizawa reported. The authors attribute the deterioration to saturation of the tactile stimulus, meaning that they believed they approached the limit of skin stretch sensation for the finger pad. However, their skin stretch device did not deliver force-controlled stimuli, so we cannot tell if their tactile 300 g cue reflected a physical 300 g weight cue. Additionally, the finger pad is unlikely to approach its stretch limit at 300 g, as work on the shear properties of the finger pad shows increasing displacements with increasing force up to 5 mm at 5 N (510 g) [29], which agrees with the movements in our T condition being ~4 mm for 400 g. Thus, the tactile sensation of the finger pad not being fully saturated at 300 g is in agreement with the material properties of the finger pad, and the results in Minamizawa *et al.* could be due to device limitations. Our setup did not allow us to use reference weights surpassing 300 g, both because holding a much larger force with the thimbles is challenging in the K condition, and because pilot testing suggested that torques on the finger at the edge of the hand rest start playing a considerable role in the T condition for weights over 500 g. To avoid hitting those limits, we choose our range to be far below that weight. This slight discrepancy between literature and our results actually indicates the importance of device-independent perceptual measures, such as the ones reported here. Thus, the novelty of our results is that they provide device guidelines in terms of force, which makes them useable across a broader range of skin stretch devices.

C. Model Comparison and Neural Mechanisms

When comparing our data to an Optimal Integration and a Sensory Capture model, the detection thresholds in Experiment 1 can be accounted for by an Optimal Integration model, whereas results of JNDs in Experiment 2 are more in line with predictions from the Sensory Capture model. The results of Experiment 3 cannot resolve this paradox, since both models predict the perceived magnitude of uni-sensory and multi-sensory cues being the same, which is what we indeed find. This is not a trivial observation, since some types of integration would have resulted in finding biases in Experiment 3. An example of this is described by Bruno *et al.* [36], who presented visual depth stimuli in which the presence of cues on relative size, height in the projection plane, occlusion, and motion parallax was manipulated. They found that these cues were summed rather than averaged in the final depth percept. Similarly, Di Luca *et al.* [37] asked participants to judge geometric properties of object shape from different mixtures of 3D cues, which they found to be a monotonically increasing function of Signal-to-Noise-Ratio. This resulted in increasing biases with increasing numbers of cues, while statistical optimal integration of signal estimates would predict averaging out of biases when adding more cues. Our PSEs do not show a summation effect, so they rule out these alternative integration models. However, they do not help us resolve the apparent paradox between Optimal Integration for light weights in Experiment 1 and Sensory Capture for the heavier weights in Experiment 2, and it seems unlikely that participants change their integration strategy when the weight range changes. Our results cannot conclusively resolve this paradox, but an alternative hypothesis is that our sensory inputs were corrupted by correlated noise, which reduces the benefits of Optimal Integration [26]. When we revisit equation 3 we can see that a correlation between cues will increase the variance of the multi-sensory cue for any $0 < w < 1$. Thus, even when cues are optimally integrated, multi-sensory JNDs will be higher in the presence of correlation (see Oruç *et al.* [26] for more details). As tactile and kinesthetic information is processed in separate physiological pathways [3], assuming uncorrelated noise seems valid for noise introduced on the physiological level. However, tactile and kinesthetic information was provided by the same device, so noise introduced at the rendering level would be correlated between the cues. As increasing the rendered force leads to increasing instability in haptic systems [38], it is likely that the rendering noise starts playing a bigger role for larger weights. Thus, we propose that Optimal Integration of skin stretch and kinesthetic information was present in all our experiments, but the benefit of the integration was reduced for supra-threshold stimuli.

D. Implications for Device Design Guidelines

Our results provide force-based design guidelines for skin stretch devices and kinesthetic devices aiming to deliver a percept of weight. The DTs found in Experiment 1 can be used to determine the minimum amount of force that a skin stretch and/or a kinesthetic device needs to be able to exert to provide a detectable percept of static weight. It needs to be noted that the detection thresholds are absolute minimum levels at which users can barely perceived the stimuli, so their main use would be to exclude devices that will not be able to meet this threshold. The JNDs in Experiment 2 can be used to understand how many perceivably different levels of force can be provided by the device, so they describe how many noticeably different weights a device can present. The PSEs in Experiment 3 show that to present an intended multi-sensory (i.e., 'natural') weight cue, the same force can simply be used for uni-sensory presentation through a skin stretch or kinesthetic device.

Interestingly, our results show that adding kinesthetic information to a skin stretch cue helps to lower detection thresholds and JNDs (Experiments 1 and 2), but does not change the perceived weight of that cue (Experiment 3). It should be noted that our experiment only manipulated the presence or absence of a cue, which showed that weight is perceived as equally heavy regardless of whether skin stretch and kinesthetic cues are presented together or alone. It is still unknown what would happen if the physical magnitude of the two cues were different, for instance, if a weight cue were composed of 1 N of kinesthetic weight and 2 N of skin stretch weight. For such a cue, integration would result in the perceived magnitude of the stimuli being between the physical magnitudes of the presented stimuli, but experiments are needed to verify this. Nonetheless, the hypothesis underlines the importance of designing tactile devices such that the provided skin stretch can be expressed in useful physical terms like force, in order to make sure that the perceptual output is as intended.

V. Conclusion

Our experiments provide force-based measures of detection thresholds, precision (JNDs), and accuracy (PSEs), for static weight perceived through skin stretch and kinesthetic information alone, and for their combination. In line with findings from literature, combining skin stretch and kinesthetic information leads to better weight detection thresholds than presenting uni-sensory cues does. For supra-threshold stimuli, Weber fractions for static weight perception ranged from 22 to 44%. Kinesthetic information was less reliable for lighter weights, whereas for heavier weights up to 300 g the two cues were roughly equally reliable. Weight was perceived as not to change significantly regardless of whether skin stretch and kinesthetic cues were presented together or alone. Discrimination performance was close to the predictions of Optimal Integration only for lighter weights, while for heavier weights, measurements were closer to predictions from a Sensory Capture model. The difference might be accounted for by the presence of correlated noise across the two cues with heavier weights, which would affect model predictions such that all our data could be explained through an Optimal Integration model. This knowledge can be used to inform, for instance, skin stretch device design.

Acknowledgment

The authors would like to thank Allison Okamura and Jake Suchoski for scientific discussions, Andrew Doxon for software support and Jonathan Browder for alternative views on the analysis.

References

[1] E. Weber, *De Pulsu, Resorptione, Audito et Tactu*. London, U.K.: Academic, 1834.
[2] L. A. Jones, "Perception of force and weight: Theory and research," *Psychological Bull.*, vol. 100, no. 1, pp. 29–42, 1986.
[3] S. J. Lederman and R. L. Klatzky, "Haptic perception: A tutorial," *Attention, Perception Psychophysics*, vol. 71, no. 7, pp. 1439–1459, 2009.
[4] Z. F. Quek, S. B. Schorr, I. Nisky, W. R. Provancher, and A. M. Okamura, "Sensory substitution and augmentation using 3-degree-of-freedom skin deformation feedback," *IEEE Trans. Haptics*, vol. 8, no. 2, pp. 209–221, Apr.–Jun. 2015.
[5] G. Westling and R. S. Johansson, "Factors influencing the force control during precision grip," *Exp. Brain Res.*, vol. 53, no. 2, pp. 277–284, 1984.
[6] A.-S. Augurelle, A. M. Smith, T. Lejeune, and J.-L. Thonnard, "Importance of cutaneous feedback in maintaining a secure grip during manipulation of hand-held objects," *J. Neurophysiol.*, vol. 89, no. 2, pp. 665–671, 2003.
[7] D. A. Nowak *et al.*, "The effects of digital anaesthesia on predictive grip force adjustments during vertical movements of a grasped object," *Eur. J. Neurosci.*, vol. 14, no. 4, pp. 756–762, 2001.
[8] D. A. Nowak, S. Glasauer, and J. Hermsdörfer, "How predictive is grip force control in the complete absence of somatosensory feedback?" *Brain*, vol. 127, no. 1, pp. 182–192, 2004.
[9] Z. F. Quek, S. Schorr, I. Nisky, A. Okamura, and W. Provancher, "Augmentation of stiffness perception with a 1-degree-of-freedom skin stretch device," *IEEE Trans. Human-Mach. Syst.*, vol. 44, no. 6, pp. 731–742, Dec. 2014.
[10] B. T. Gleeson, S. K. Horschel, and W. R. Provancher, "Design of a fingertip-mounted tactile display with tangential skin displacement feedback," *IEEE Trans. Haptics*, vol. 3, no. 4, pp. 297–301, Oct.–Dec. 2010.
[11] S. B. Schorr, Z. F. Quek, I. Nisky, W. R. Provancher, and A. M. Okamura, "Tactor-induced skin stretch as a sensory substitution method in teleoperated palpation," *IEEE Trans. Human-Mach. Syst.*, vol. 45, no. 6, pp. 714–726, 2015.
[12] Y. Kurita, S. Yonezawa, A. Ikeda, and T. Ogasawara, "Weight and friction display device by controlling the slip condition of a fingertip," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Sep. 2011, pp. 2127–2132.
[13] K. Minamizawa, D. Prattichizzo, and S. Tachi, "Simplified design of haptic display by extending one-point kinesthetic feedback to multipoint tactile feedback," in *Proc. IEEE Haptics Symp.*, 2010, pp. 257–260.

[14] K. Matsui, S. Okamoto, and Y. Yamada, "Relative contribution ratios of skin and proprioceptive sensations in perception of force applied to fingertip," *IEEE Trans. Haptics*, vol. 7, no. 1, pp. 78–85, Jan.–Mar. 2014.

[15] C. Giachritsis, R. Wright, and A. Wing, "The contribution of proprioceptive and cutaneous cues in weight perception: Early evidence for maximum-likelihood integration," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes Artificial Intell. Lecture Notes Bioinformatics)*, vol. 6191 LNCS, pp. 11–16, 2010.

[16] R. J. Webster, T. E. Murphy, L. N. Verner, and A. M. Okamura, "A novel two-dimensional tactile slip display: Design, kinematics and perceptual experiments," *ACM Trans. Appl. Percept.*, vol. 2, no. 2, pp. 150–165, 2005.

[17] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, no. 6870, pp. 429–433, 2002.

[18] R. J. Van Beers, A. C. Sittig, and J. J. Denier van der Gon, "Integration of proprioceptive and visual position-information: An experimentally supported model," *J. Neurophysiol.*, vol. 81, no. 3, pp. 1355–1364, 1999.

[19] M. Kuschel, M. Di Luca, M. Buss, and R. Klatzky, "Combination and integration in the perception of visual-haptic compliance information," *IEEE Trans. Haptics*, vol. 3, no. 4, pp. 234–244, Oct.–Dec. 2010.

[20] F. E. van Beek, R. J. King, C. Brown, and M. Di Luca, "The contributions of skin stretch and kinesthetic information to static weight perception," in *Proc. IEEE World Haptics Conf.*, Jul. 2019, pp. 235–240.

[21] N. Gurari and G. Baud-Bovy, "Customization, control, and characterization of a commercial haptic device for high-fidelity rendering of weak forces," *J. Neurosci. Methods*, vol. 235, pp. 169–180, 2014.

[22] N. Prins and F. Kingdom, Palamedes: Matlab routines for analyzing psychophysical data. 2009, [Online]. Available: http://www.palamedestoolbox.org

[23] F. A. A. Kingdom and N. Prins, *Psychophysics: A Practical Introduction*. Amsterdam, The Netherlands: Elsevier, 2010.

[24] D. Green and J. Swets, *Signal Detection Theory and Psychophysics* vol. 1. New York, NY, USA: Wiley.

[25] M. Pannunzi, A. Pérez-Bellido, A. Pereda-Baños, J. López-Moliner, G. Deco, and S. Soto-Faraco, "Deconstructing multisensory enhancement in detection," *J. Neurophysiol.*, vol. 113, no. 6, pp. 1800–1818, 2015.

[26] İ. Oruç, L. T. Maloney, and M. S. Landy, "Weighted linear cue combination with possibly correlated error," *Vision Res.*, vol. 43, no. 23, pp. 2451–2468, 2003.

[27] Z. Dienes and N. Mclatchie, "Four reasons to prefer bayesian analyses over significance testing," *Psychonomic Bull. Rev.*, vol. 25, no. 1, pp. 207–218, Feb. 2018.

[28] J. Flanagan, J. Tresilian, and A. M. Wing, "Coupling of grip force and load force during arm movements with grasped objects," *Neuroscience Lett.*, vol. 152, no. 1, pp. 53–56, 1993.

[29] T. C. Pataky, M. L. Latash, and V. M. Zatsiorsky, "Viscoelastic response of the finger pad to incremental tangential displacements," *J. Biomechanics*, vol. 38, no. 7, pp. 1441–1449, 2005.

[30] G. Buckingham, J. S. Cant, and M. A. Goodale, "Living in a material world: How visual cues to material properties affect the way that we lift objects and perceive their weight," *J. Neurophysiol.*, vol. 102, no. 6, pp. 3111–3118, 2009.

[31] S. Lederman and R. Klatzky, "Hand movements: A window into haptic object recognition," *Cogn. Psychol.*, vol. 19, no. 3, pp. 342–368, 1987.

[32] J. R. Flanagan and A. M. Wing, "Modulation of grip force with load force during point-to-point arm movements," *Exp. Brain Res.*, vol. 95, no. 1, pp. 131–143, 1993.

[33] J. R. Flanagan and C. A. Bandomir, "Coming to grips with weight perception: Effects of grasp configuration on perceived heaviness," *Perception Psychophysics*, vol. 62, no. 6, pp. 1204–1219, 2000.

[34] G. Baud-Bovy and E. Gatti, "Hand-held object force direction identification thresholds at rest and during movement," in *Haptics: Generating and Perceiving Tangible Sensations*, Lecture Notes in Computer Science, vol. 6192, A. M. L. Kappers, J. B. F. Van Erp, W. M. Bergmann Tiest, and F. C. T. Van der Helm, Eds. Berlin, Germany: Springer, 2010, pp. 231–236.

[35] F. E. van Beek, W. M. Bergmann Tiest, A. M. L. Kappers, and G. Baud-Bovy, "Integrating force and position: Testing model predictions," *Exp. Brain Res.*, vol. 234, pp. 3367–3379, 2016.

[36] N. Bruno and J. E. Cutting, "Minimodularity and the perception of layout," *J. Exp. Psychol., Gen.*, vol. 117, no. 2, pp. 161–170, 1988.

[37] M. D. Luca, F. Domini, and C. Caudek, "Inconsistency of perceived 3 d shape," *Vis. Res.*, vol. 50, no. 16, pp. 1519–1531, 2010.

[38] B. Hannaford and A. M. Okamura, *Haptics*. Berlin, Heidelberg: Springer, 2008, pp. 719–739.

Femke van Beek (Member, IEEE) received the B.Sc. and M.Sc. degree in sensory biology from Wageningen University. She received the PhD degree in haptic perception from Vrije Universiteit Amsterdam under the supervision of Astrid Kappers. In 2016, she joined Oculus as a Postdoctoral Researcher. Currently, she is a Visiting Scientist with Facebook Reality Labs. Her research interests include haptics, multi-sensory perception, and psychophysics.

Raymond King received the B.S. and M.S. degrees in engineering science and mechanics from Virginia Tech, Blacksburg, VA in 2009 and 2010, respectively, and the Ph.D degree in mechanical engineering from the University of Utah, Salt Lake City, UT in 2014. He currently works as Research Manager with Facebook Reality Labs in Redmond, Washington.

Casey Brown is currently a Technical Program Manager currently working at Facebook Reality Labs. He received a B.Sc degree in psychology from the University of Washington (Seattle) in 2013. Casey has supported usability research at Microsoft for Xbox and other Windows devices. His research interests include user experience and perception science.

Massimiliano Di Luca (Member, IEEE) achieved a masters degree in psychology from the University of Trieste (Italy) in 2000 and the Ph.D. degree in cognitive sciences from Brown University (USA), in 2006. He worked as a scientist at the Max Planck Institute for Biological Cybernetics (Germany). In 2011, he joined the University of Birmingham (UK) where he is currently a Senior Lecturer of Computational Neuroscience. In 2016, he spent a sabbatical at Oculus (USA) and a secondment leave at Facebook Reality Labs as research scientist. Dr Di Luca's research interest is to understand the computations that are involved in perception. Using psychophysical methods and computational modelling, he investigates how the human brain combines multisensory information for perception and action.

Sean Keller (Member, IEEE) received the B.S. and M.Eng. degrees from Cornell University, Ithaca, NY, USA, the M.S. degree from the University of Illinois at Urbana-Champaign, Urbana, IL, USA, and the Ph.D. degree from the California Institute of Technology, Pasadena, CA, USA. He is currently a Director of Research with Facebook Reality Labs (FRL). His current research interests include haptics, human-computer interaction, AR/VR input devices, soft goods, sensor systems, and modeling of perceptual and cognitive systems.