Integration of haptics and vision in human multisensory grasping

Ivan Camponogara* and Robert Volcic

Department of Psychology, New York University Abu Dhabi, Abu Dhabi, United Arab Emirates

Contact information:

Ivan Camponogara
Department of Psychology, New York University Abu Dhabi
PO Box 129188, Abu Dhabi, United Arab Emirates
E-mail: ivan.camponogara@nyu.edu

Abstract:

Grasping actions are directed not only toward objects we see but also toward objects we both see and touch (multisensory grasping). In this latter case, the integration of visual and haptic inputs improves movement performance compared to each sense alone. This performance advantage could be due to the integration of all the redundant positional and size cues or to the integration of only a subset of these cues. Here we selectively provided specific cues to tease apart how these different sensory sources contribute to visuo-haptic multisensory grasping. We demonstrate that the availability of the haptic positional cue together with the visual cues is sufficient to achieve the same grasping performance as when all cues are available. These findings provide strong evidence that the human sensorimotor system relies on non-visual sensory inputs and open new perspectives on their role in supporting vision during both development and adulthood.

Keywords: Multisensory integration, vision, haptics, proprioception, touch, grasping

Introduction

The planning and execution of a successful reach-to-grasp movement relies on the translation of visual inputs into motor commands that are involved in the control of the hand.[1–3] However, actions in everyday life are not only directed toward objects we see but also toward objects we feel with our hands. For instance, even without looking at it, we can easily reach and grasp the cap of a pen while holding the pen with the other hand. Thus, the afferent haptic inputs (proprioception and touch) from the hand in contact with the object are sufficient to estimate the object's position and its size and successfully guide the action of the opposite hand. However, movements toward haptically sensed objects are usually slower and show a wider grip aperture compared to movements directed toward visually detected objects.[4–6] Importantly, the simultaneous availability of vision and haptics leads to faster movements with narrower grip apertures compared to when objects are only visually, or haptically sensed.[4,7] Thus, haptics can effectively integrate with vision to plan and execute grasping actions, but how does each sense contribute to this multisensory-motor transformation is still poorly understood.

Even though vision and haptics provide redundant cues about the position and the size of an object, the two sensory systems acquire and process these cues in fundamentally different ways. In the visual domain, the estimated size of an object is tightly linked to its estimated position. In essence, determining an object's size requires the scaling of its retinal projections according to its distance from the observer.[8–12] Instead, in the haptic domain, the estimated size and position of the object are entirely independent. The object size is provided by the proprioceptive and tactile inputs from the digits enclosing the object,[13–16] whereas the object position is provided by the proprioceptive inputs from the muscles, joints, and skin of the flexed arm.[17] Hence, there are multiple ways in which visual and haptic cues could be integrated in multisensory grasping.

A full integration model posits that both visual and haptic cues about the object size and its position are all merged (Figure 1, left panel): the integration of visual and haptic position estimates promotes a better visual size estimate which is then integrated with the haptic size estimate. The enhanced performance in visuo-haptic multisensory grasping would thus stem from the availability of multisensory estimates of both position and size. An alternative model, the size integration model (Figure 1, middle panel), omits the haptic position cue and is thereby limited to the integration of haptic and visual size estimates only. The joint availability of visual and haptic size would be thus sufficient to improve performance in visuo-haptic grasping. This model resembles the models commonly used to describe visuo-haptic size

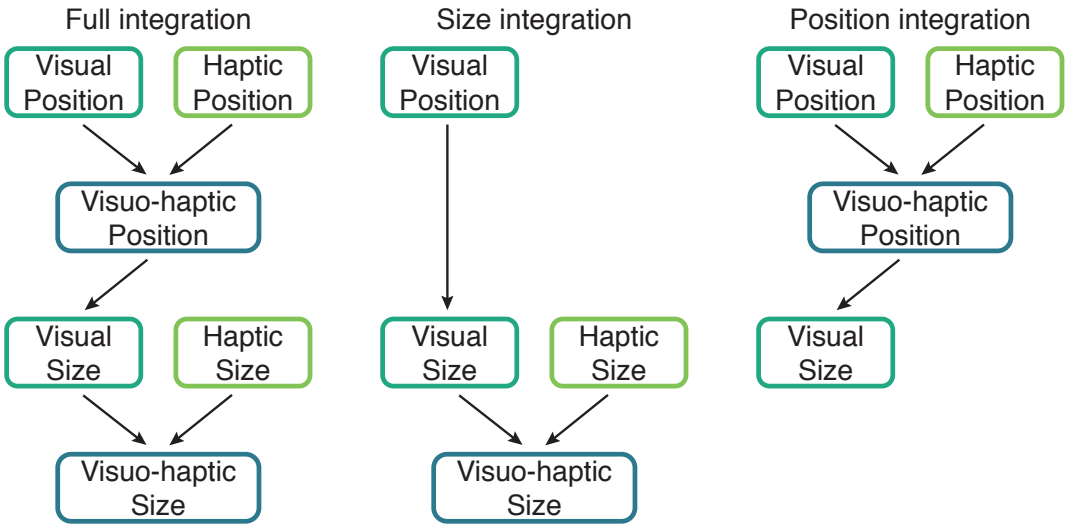

Figure 1: Integration of visual and haptic cues. Left panel, Full integration model: visual and haptic position estimates are merged into a visuo-haptic position estimate. The visual size estimate, which is based on the multisensory position estimate is then integrated with the haptic size estimate. Middle panel, Size integration model: visual object size is based on the visual position estimate only. The visual size estimate is then integrated with the haptic size estimate. Right panel, Position integration model: visual and haptic position estimates are integrated into a visuo-haptic position estimate which supports the visual object size estimate.

perception.[18–21] Another model, the position integration model (Figure 1, right panel), is instead focused on the role of haptics as an auxiliary position information cue.[22–25] This model is thus limited to the integration of only haptic and visual position estimates by disregarding the haptic size cue. In this case, the haptic position estimate would improve the visual position estimate which would lead to a better visual size estimate and, in turn, boost visuo-haptic multisensory grasping.

To determine the contribution of haptic and visual cues in multisensory grasping we have characterized grasping performance in two unisensory and two multisensory conditions (Figure 2a). In the haptic condition (H), vision was prevented, and grasping relied on the haptic size and position cues provided by the left hand holding the object. In the visual condition (V), grasping was based on the visually sensed location and size of the object. In the first multisensory condition, the visuo-haptic condition (VH), both visual and haptic cues about the object position and its size were available. In the second multisensory condition, the visuo-haptic position condition (VHP), participants held with the left hand a post which supported the object, instead of holding the object itself, while vision was fully available. Thus, haptics was informative only about the position of the object, but not about its size. In all these conditions, the

grasping action was performed by the right hand.

Based on previous research,[4,7] we expect the multisensory (VH) condition to exhibit faster grasping movements with smaller maximum grip apertures than the V and H unisensory conditions. The comparison between the two multisensory conditions (VH vs. VHP) is instead crucial to distinguish among the full integration, size integration and position integration models. Both the full integration and the size integration models rely, in part, on the availability of the haptic size cue to achieve the multisensory advantage. Thus, if the haptic size cue is absent, both models predict a worse grasping performance in VHP (in which no haptic size cue is provided) than in VH. Grasping movements should be as slow and with maximum grip apertures as large as in the V condition according to the size integration model and somewhere in-between the VH and the V conditions according to the full integration model. On the other hand, the position integration model depends only on the availability of the haptic position cue. Hence, if the haptic position cue is sufficient, this model predicts identical grasping performance in both VH and VHP.

Methods

Participants

Twenty students from the New York University Abu Dhabi took part in this study (10 male, age 19.4 ±0.9). All had normal or corrected-to-normal vision and no known history of neurological disorders. All of the participants were naïve to the purpose of the experiment and were provided with a subsistence allowance. The experiment was undertaken with the understanding and written informed consent of each participants and experimental procedures were approved by the Institutional Review Board of New York University Abu Dhabi.

Apparatus

In the H, V and VH conditions, the stimuli were five 3D-printed rectangular cuboids with a depth of 30, 40, 50, 60, 70 mm, all the same height (120 mm) and width (25 mm) (Figure 2b). In the VHP condition, the stimuli were five 60 mm high rectangular cuboids supported by a 60 mm high post (Figure 2c). While the upper part of these objects varied in depth as in the first set of objects, the depth of the post was kept constant (10 mm) and thus haptics was non-informative about the depth of the to-be-grasped object. In addition, we have manipulated the position of the objects by placing them at three different egocentric

distances in front of the participants: at 150 mm, 250 mm or 350 mm in the sagittal direction (depth) at table height. Two 5 mm high rubber bumps with a diameter of 9 mm were attached just in front of the participants, 300 mm to the left and to the right. These bumps were marking the start positions for the left and right hands (Figure 2b).

A pair of occlusion goggles (Red Scientific, Salt Lake City, UT, USA) controlled by a custom Matlab program was used to prevent vision of the workspace in the haptic condition and between trials. A pure tone of 1000 Hz and 100 ms length was used to signal the start of the trial, while another tone of 600 Hz with the same duration was used to signal its end. Index, thumb and wrist movements were acquired on-line at 200 Hz with sub-millimeter resolution by using an Optotrak Certus system (Northern Digital Inc., Waterloo, Ontario, Canada) controlled by the MOTOM toolbox.[26] The position of the tip of each digit was calculated during the calibration phase with respect to three infrared-emitting diodes attached on each distal phalanx.[27] An additional marker was attached on the wrist (styloid process of the radius).

Procedure

Participants sat comfortably in front of a table with their torso touching its edge. All the trials started with the participants' thumb and index digit of the right and left hand positioned on the respective start positions and the shutter goggles closed. Before each trial, one of the objects was positioned at one of the three positions. In the H condition, the experimenter signaled to the participants to hold the object with their left hand along its depth axis at its base (i.e., sense its size and position by means of tactile and proprioceptive inputs) while the shutter goggles remained closed. In the V condition, the goggles turned transparent to provide vision of the object, no haptic information was provided by the left hand. In the VH condition, the participants held the object with their left hand and then the goggles turned transparent. In the VHP condition, participants held the post on which the object was firmly placed with their left hand and then the goggles turned transparent. Unlike in VH, in which haptic inputs were informative about both the object size and its position, the haptic inputs in VHP were informative only about the object's position (Figure 2c).

After a variable period, the start tone was delivered and participants had to reach for and grasp the object along its depth axis. Movements were performed at a natural speed and no reaction time constrains were imposed. After 3 seconds the end sound was delivered, and, only in the H modality, the goggles were made transparent. Participants had to move their right and left hands back to the start positions and

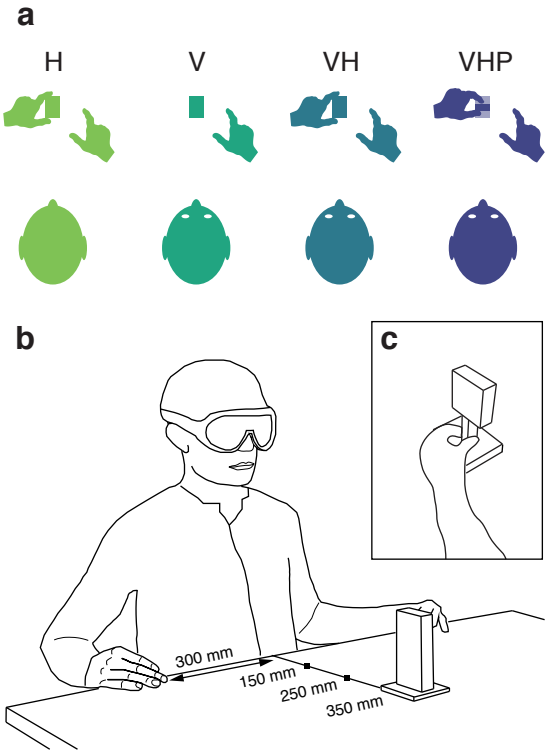

Figure 2: (a) Haptic (H), visual (V), visuo-haptic (VH), and visuo-haptic position (VHP) conditions. Grasping actions were always performed with the right hand. In H, VH and VHP the left hand was already holding the object or the post before the start of the grasping movement. (b) Experimental setup with the stimulus used in conditions H, V, and VH. (c) Left hand holding the post that supported the object in condition VHP.

then the goggles turned opaque. Another object and position were then selected and the next trial was ready to start.

The order of conditions was randomized across participants using a Latin square design, while size and position configurations were randomized within each condition. We ran five repetitions for each combination of object size and position, which led to a total of 300 trials per participant (75 for each condition). Before the experiment, a training session was performed in which ten trials were run in each condition to accustom the participants with the task.

Data analysis

Kinematic data were analyzed in R.[28] The raw data were smoothed and differentiated with a third-order Savitzky-Golay filter with a window size of 21 points. These filtered data were then used to compute velocities and accelerations in three-dimensional space for each digit and the wrist. Movement onset was

defined as the moment of the lowest, non-repeating wrist acceleration value prior to the continuously increasing wrist acceleration values,[4,29] while the end of the grasping movement was defined on the basis of the Multiple Sources of Information method.[30] We used the criteria that the grip aperture is close to the size of the object, that the grip aperture is decreasing, that the second derivative of the grip aperture is positive, and that the velocities of the wrist, thumb and index finger are low. Moreover, the probability of a moment being the end of the movement decreased over time to capture the first instance in which the above criteria were met.[4,29] Trials in which the end of the movement was not captured correctly or in which the missing marker samples could not be reconstructed using interpolation were discarded from further analysis. The exclusion of these trials (88 trials, 1.4% in total) left us with 5912 trials for the final analysis.

We focused our analyses on two dependent variables: the maximum grip aperture (MGA), defined as the maximum Euclidean distance between the thumb and the index finger, and, the Peak Velocity (PV) of the hand movement, defined as the highest wrist velocity along the movement. These two variables are indicative of the effect of object size and the effect of object position on the movement kinematics.[31]

We analyzed the data using Bayesian linear mixed-effects models, estimated using the brms package[32] which implements Bayesian multilevel models in R using the probabilistic programming language Stan.[33] The models used to fit the MGA and PV data included as fixed-effects (predictors) the categorical variable Condition (H, V, VH, VHP) in combination with the continuous variables Size and Position. The continuous variables Size and Position were centered before being entered in the models, thus, the estimates of the Condition parameters ($\beta_{Condition}$) correspond to the average performance of each Condition. The estimates of the parameters Size (β_{Size}) and Position ($\beta_{Position}$) correspond instead to the change in the dependent variables as a function of the object size and its position (i.e., distance from the participant). All models included independent random (group-level) effects for subjects. Models were fitted considering weakly informative prior distributions for each parameter to provide information about their plausible scale. We used Gaussian priors for the Condition fixed-effect predictor (MGA $\beta_{Condition}$: mean = 85 and sd = 40; PV $\beta_{Condition}$: mean = 950 and sd = 500). For the Size and Position fixed-effect predictors we used Cauchy prior distributions centered at 0 with a scale parameter of 2.5. For the group-level standard deviation parameters and sigmas we used Student t-distribution priors (MGA all sd parameters and sigma: $df = 3$, scale = 15; PV sd Condition and sigma: $df = 3$, scale = 166; PV sd Size and sd Position: $df = 3$, scale = 5). Finally, we set a prior over the correlation matrix that assumes that smaller correlations are slightly more likely than larger ones (LKJ prior set to 2).

For each model we ran four Markov chains simultaneously, each for 16,000 iterations (1,000 warm-up samples to tune the MCMC sampler) with the delta parameter set to 0.9 for a total of 60,000 post-warm-up samples. Chain convergence was assessed using the $\hat{R}$ statistic (all values equal to 1) and visual inspection of the chain traces. Additionally, predictive accuracy of the fitted models was estimated with leave-one-out cross-validation by using the Pareto Smoothed importance Sampling (PSIS). All Pareto k values were below 0.5.

The posterior distributions we have obtained represent the probabilities of the parameters conditional on the priors, model and data, and, they represent our belief that the "true" parameter lies within some interval with a given probability. We summarize these posterior distributions by computing the medians and the 95% Highest Density Intervals (HDI). The 95% HDI specifies the interval that includes with a 95% probability the true value of a specific parameter. To evaluate the differences between parameters of two conditions, we have simply subtracted the posterior distributions of $\beta_{Condition}$, β_{Size} and $\beta_{Position}$ weights between specific conditions (H-V, VH-H, VH-V, VH-VHP). The resulting distributions are denoted as the credible difference distributions (CDD) and are again summarized by computing the medians and the 95% HDIs.

For statistical inferences about model parameters (β_{Size} and $\beta_{Position}$) we assessed the overlap of the 95% HDI with zero. A 95% HDI that does not span zero indicates that the predictor has an effect on the dependent variable. For statistical inferences about the differences of the model parameters between conditions, we applied an analogous approach. A 95% HDI of the credible difference distribution that does not span zero is taken as evidence that the model parameters in the two conditions differ from each other.

Results

We report the main results in two separate sections. In the first section, we describe and compare the performance in the H, V and VH conditions to establish how unisensory and multisensory inputs affects grasping behavior. In the second section, we focus on the critical comparison between the VH and VHP conditions. This comparison allows us to establish whether the haptic size cue is crucial to optimize multisensory grasping movements and will reveal which model (full integration, size integration or position integration) is the most plausible.

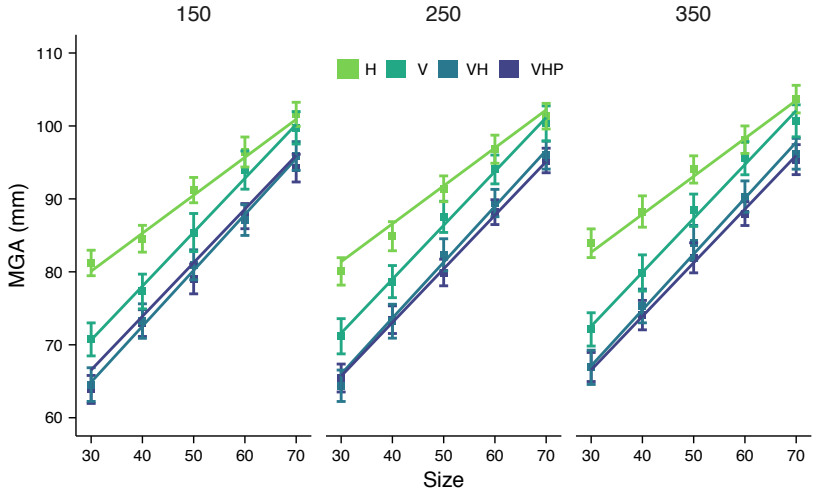

Figure 3: Average MGA as a function of size and position (separate panels) in the H, V, VH and VHP conditions. Error bars represent the standard error of the mean. Solid lines show the Bayesian mixed-effects model fits.

Grasping with haptic, visual and visuo-haptic inputs

The MGA was clearly affected not only by the size and position of the object, but, most importantly, also by the available sensory inputs (Figure 3). As can be also seen in Figure 4a, the average MGA was larger in H (91.8 mm, 95% HDI = 88.2, 95.0) than in V (86.3 mm , 95% HDI = 82.5, 90.3) which, in turn, was larger than in VH (81.2 mm, 95% HDI = 77.7, 84.7). The comparisons between conditions, based on the CDDs of the $\beta_{Condition}$ weights (Figure 4d), showed that the MGA in H was credibly larger than in both V (H−V = 5.4 mm, 95% HDI = 1.3, 9.4) and VH (H−VH = 10.5 mm, 95% HDI = 6.7, 14.5). Moreover, the MGA in V was credibly larger than in VH (V−VH = 5.1 mm, 95% HDI = 2.1, 8.0). The result that the MGA is smallest in VH replicates and extends previous findings,[4,7] and supports the idea that the simultaneous availability of visual and haptic inputs leads to a substantial multisensory advantage (approximately 5 mm and 10 mm smaller average MGA compared to V and H, respectively).

The MGA increased as a function of object size in all sensory conditions (Figure 3). However, the slope represented by β_{Size} weights was shallower in H (0.52, 95% HDI = 0.49, 0.54) than in V (0.73, 95% HDI = 0.71, 0.76) and VH (0.76, 95% HDI = 0.73, 0.80), as can be seen in Figure 4b. The comparisons between conditions, based on the CDDs of the β_{Size} weights (Figure 4e), showed that MGA increased credibly less as a function of object size in H than in both V (H−V = −0.21, 95% HDI = −0.25, −0.18) and VH (H−VH = −0.24, 95% HDI = −0.28, −0.20). Instead, the changes in MGA as a function of object size

were essentially identical in V and VH (V−VH = −0.02, 95% HDI = −0.06, 0.01). This suggests that the MGA in H is less sensitive to changes in object size and, thus, the MGA modulation in multisensory grasping is mainly based on the visual size cue.

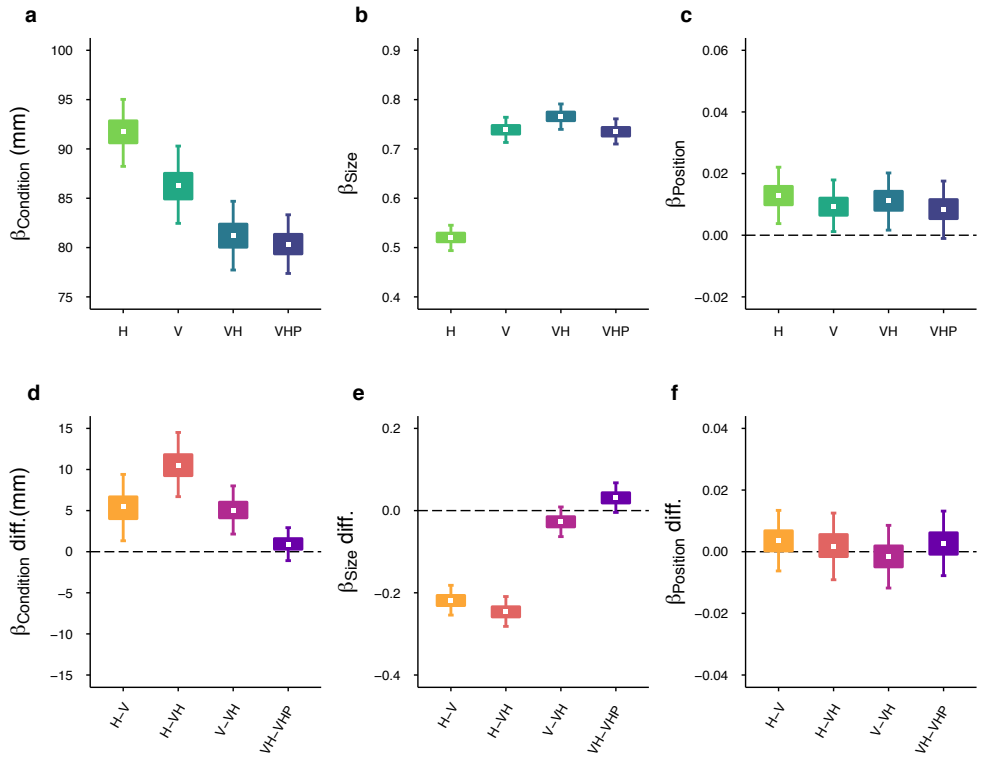

Figure 4: MGA results. Top row: Posterior beta weights of the Bayesian linear mixed-effects regression model for the predictors Condition (a), Size (b) and Position (c). Bottom row: CDDs between conditions for the predictors Condition (d), Size (e) and Position (f). White dots represent the median, the boxes represent the 50% HDIs, and the areas between whiskers represent the 95% HDIs of the distributions.

The MGA credibly increased also as a function of object position (Figure 4c), but only moderately (H = 0.012, 95% HDI = 0.003, 0.022; V = 0.009, 95% HDI = 0.001, 0.018, VH = 0.011, 95% HDI = 0.001, 0.02). Most importantly, the comparisons between conditions, based on the CDDs of the $\beta_{Position}$ weights (Figure 4f), showed that these MGA increases were indistinguishable between conditions (H−V = 0.003, 95% HDI = −0.005, 0.01; H−VH = 0.001, 95% HDI = −0.008, 0.01; V−VH = −0.001, 95% HDI = −0.01, 0.007).

The PV was mainly affected by the position of the object, but, notably, also by the available sensory inputs (Figure 5). The grasping movements in the unisensory conditions (H = 948 mm/s, 95% HDI = 906, 990; V = 936 mm/s, 95% HDI = 900, 973) reached a lower PV than in the multisensory condition (VH = 971 mm/s, 95% HDI = 948, 1018), as can be also seen in Figure 6a. The comparisons between

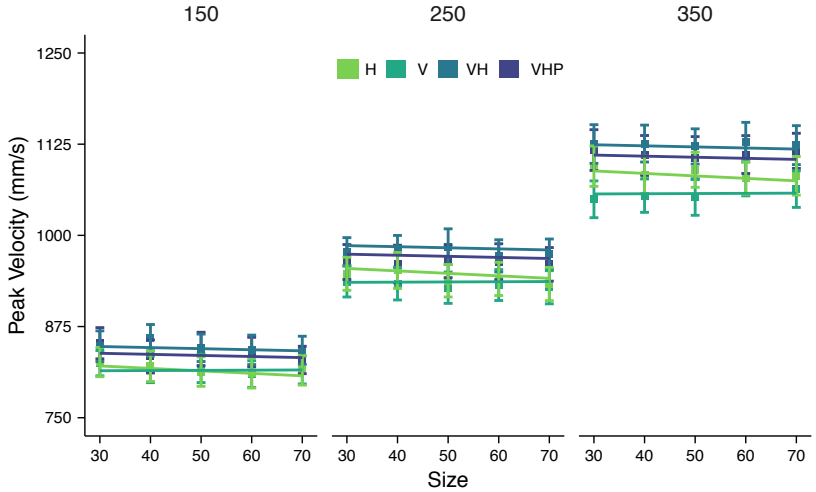

Figure 5: Average PV as a function of size and position (separate panels) in the H, V, VH and VHP conditions. Error bars represent the standard error of the mean. Solid lines show the Bayesian mixed-effects model fits.

conditions, based on the CDDs of the $\beta_{Condition}$ weights (Figure 6d), showed that the PV in V was credibly lower than in VH (V−VH = −47 mm/s, 95% HDI = −79, −16). No difference in terms of PV was found between H and V (H−V = 12 mm/s, 95% HDI = −33, 55). Although the CDD for the comparison between H and VH partially overlapped with zero (H−VH = −35 mm/s, 95% HDI = −73, 4), the bulk of the distribution was clearly negative. These results suggest that visual and haptic inputs are successfully integrated to speed-up multisensory grasping movements.

The PV was insensitive to changes in object size in V and VH conditions, as can be gathered from the slopes represented by β_{Size} weights (Figure 6b), which were relatively flat both in V (V = 0.02 s^{-1}, 95% HDI = −0.18, 0.25) and in VH (VH = −0.14 s^{-1}, 95% HDI = −0.37, 0.06). Instead, in H, the PV credibly decreased as a function of object size (H = −0.34 s^{-1}, 95% HDI = −0.56, −0.12). This decrease might be due to a compensatory strategy specific to haptically-guided grasping wherein participants minimize the risk of colliding with larger objects by decreasing movement velocity instead of increasing the grip aperture. This might be also taken as a symptom of worse sensitivity to size in H than in V. The comparisons between conditions, based on the CDDs of the β_{Size} weights (Figure 6e), showed that the slopes were credibly different only for the comparison between H and V (H−V = −0.36 s^{-1}, 95% HDI = −0.68, −0.06). There was no evidence of any other difference (H−VH = −0.19 s^{-1}, 95% HDI = −0.49, 0.12; V−VH = 0.17 s^{-1}, 95% HDI = −0.13, 0.48).

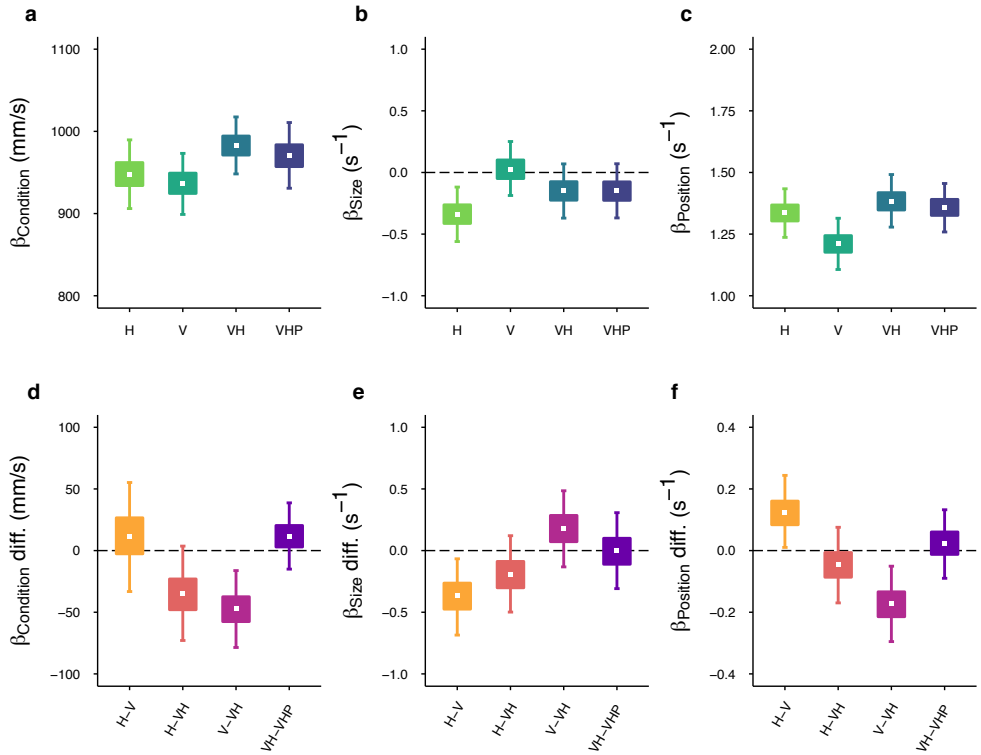

Figure 6: PV results. Top row: Posterior beta weights of the Bayesian linear mixed-effects regression model for the predictors Condition (a), Size (b) and Position (c). Bottom row: CDDs between conditions for the predictors Condition (d), Size (e) and Position (f). White dots represent the median, the boxes represent the 50% HDIs, and the areas between whiskers represent the 95% HDIs of the distributions.

The object position credibly influenced PV in all sensory conditions (Figure 6c). However, when haptic cues were available, either alone or in combination with vision, the increases in PV as a function of object position were greater than when movements were under visual guidance only (H = 1.33 s^{-1}, 95% HDI = 1.23, 1.43; V = 1.21 s^{-1}, 95% HDI = 1.10 s^{-1}, 1.31; VH = 1.38, 95% HDI = 1.27, 1.49). This result was confirmed by the comparisons between conditions, based on the CDDs of the $\beta_{Position}$ weights (Figure 6f). Whereas the PV increases in H and VH were credibly different from those found in V (H−V = 0.12 s^{-1}, 95% HDI = 0.01, 0.24; V−VH = −0.16 s^{-1}, 95% HDI = −0.29, −0.05), no difference was observed between H and VH (H−VH = −0.04 s^{-1}, 95% HDI = −0.16, 0.07). These findings provide a first piece of evidence that multisensory integration in grasping relies on the haptic estimates of egocentric object distance more than on the haptic estimates of object size. The comparison between the VH and VHP conditions will further clarify the specific role of the individual haptic cues.

Visuo-haptic grasping with only the haptic position cue

If the haptic size cue is crucial to achieve faster movements with smaller grip apertures in multisensory grasping, then the absence of this cue should deteriorate grasping performance in VHP (Figure 1, left and middle panels) as predicted by the full integration and the size integration models. If, on the other hand, only the haptic position cue is sufficient to boost performance in multisensory grasping, the position integration model predicts that the grip apertures in VHP should be as small as in VH and movements should be as fast (Figure 1, right panel). Indeed, we found that grasping performance in VH and VHP was essentially identical (Figure 3 and 5).

The average MGA in VHP (VHP = 80.3 mm, 95% HDI = 77.4, 83.3) was the same as in VH (Figure 4d, VH−VHP = 0.9 mm, 95% HDI = −1.1, 2.9) and the MGA in VHP increased as a function of object size at the same rate as in VH (VHP = 0.73, 95% HDI = 0.70, 0.76), even when the haptic cue about object size was lacking (Figure 4e, VH−VHP = 0.03, 95% HDI = −0.01, 0.06). With regards to object position, the MGA in VHP increased with distance (0.008, 95% HDI = −0.001, 0.001) as much as in VH (Figure 4f, VH−VHP = 0.002, 95% HDI = −0.006, 0.01). Similarly, the average PV in VHP (971 mm/s, 95% HDI = 931, 1011) was the same as the one we observed in VH (Figure 6d, VH−VHP = 12 mm/s, 95% HDI = −15, 39) and the PV in VHP (VHP = −0.14 s^{-1}, 95% HDI = -0.36, 0.07) was as insensitive to changes in object size as in VH (Figure 6e, VH−VHP = -0.0003 s^{-1}, 95% HDI = −0.30, 0.30). Finally, in terms of object position, the PV scaling to distance did not differ between VH and VHP (VHP = 1.35 s^{-1}, 95% HDI = 1.25, 1.45; VH−VHP = 0.02 s^{-1}, 95% HDI = −0.09, 0.13).

Taken together, these results suggest that the haptic position cue, and not haptic size, is required for optimized multisensory grasping movements. Thus, faster movements with smaller grip apertures are achieved by integrating visual and haptic position cues that are then jointly used to better scale visual size, consistent with the predictions of the position integration model.

Discussion

The study presented here was designed to investigate the role of visual and haptic cues in visuo-haptic multisensory grasping. The results show that the availability of both vision and haptics produced faster reach-to-grasp movements with considerably narrower grip apertures than in the unisensory conditions (vision-only or haptics-only) extending our previous findings[4,7] over a wide range of object sizes and object positions. Critically, when full vision was coupled with the haptic position cue only (VHP),

grasping movements were indistinguishable from those found in the full visuo-haptic condition (VH) in which also the haptic size cue was available.

Because both the full integration and the size integration models rely on the availability of the haptic size cue, its absence should have degraded grasping performance. Our results show that this was evidently not the case, and clearly support the position integration model. Therefore, we must conclude that the main contribution of the haptic modality is related to the estimation of the object position in space. The integration of this haptic position estimate with the available visual information it thus sufficient to enhance grasping performance to the same level as when also the haptic size cue is accessible. The use of haptics in supporting the visual estimation of the object position rather than its size might be related to the intrinsic positional function of the proprioceptive receptors, whose inputs provide a continuous flow of information about the extension of the limbs.[17] Our findings thus strengthen the idea that non-visual inputs from the hand holding the object actively support vision in estimating object properties.[22–25]

The fact that the haptic size cue does not play a role in multisensory grasping might, at first glance, seem surprising. After all, the purpose of multisensory processing is to augment the information contributed by the single modalities.[20] However, ignoring the haptic size cue could be seen as a virtue rather than a problem, because integrating this specific cue would be in many cases detrimental. Common objects have irregular shapes and, thus, the size perceived by the hand holding the object would rarely match the size of the part of the object that is the target of the grasping action (e.g., reaching with one hand for the cap of a bottle we hold in the other hand). Moreover, actions towards handheld objects do not necessarily end by grasping them (e.g., tapping on a smartphone). In these cases, the haptic position cue can still support vision in movement control, whereas the haptic size cue is essentially irrelevant. By relegating haptic size cues to a secondary role, the sensorimotor system could actually achieve greater robustness to variations in objects shape and action goals. This view does not, however, exclude the possibility that the haptic size cue is used when vision becomes less reliable, such as when objects are positioned in visual periphery.[34]

An additional point that is worth discussing concerns the role of multisensory grasping in learning the mapping between the visual and the motor systems. Infants start to develop the ability to reach and grasp objects at between four and six months of age before learning to perform grasping movements that resemble those of adults at the age of 10–12 months.[35,36] Importantly, their inability to produce

successful grasps is not due to an immature motor system,[37] to low visual acuity[38] or to undeveloped stereovision,[39,40] but rather to an unformed mapping between visual inputs and motor plans. This visuomotor mapping is thought to be achieved through an embodied process that requires infants to rely on proprioceptive and, more generally, haptic inputs from their reaching hand.[41–43] Our findings suggest that valuable haptic spatial cues provided by the hand holding the object could also assist the visuomotor learning process by promoting the development of precise visually controlled reach and grasp movements. In fact, anyone who has witnessed the motor development of a newborn must have observed that they can easily direct a grasping movement toward a handheld object long before they can grasp a distal object they can only see.

It is also important to note that the plasticity of the mapping between visual and motor systems does not end at the onset of adulthood. The visual feedback about the ongoing movement,[29,44–49] the terminal haptic feedback obtained by the grasping hand,[50–55] and the sensory prediction errors from past trials[?,?,12] have all been recognized to be important factors in calibrating visually guided grasping. We suggest that haptic spatial cues about the handheld object could presumably also play a relevant role in maintaining the correct visuomotor calibration over the lifespan or even aid the acquisition of visually guided skills in specific populations, such as children with dense bilateral congenital cataracts who recover vision years after birth.[?,?]

Even though this study was not designed to distinguish between different theories about grasping, it provides some important insights to be considered. First, according to the two-visual-systems hypothesis,[?] binocular vision provides estimates of distance and size that are accurate and reliable and is thus critical for grasping guidance. Indeed, switching from binocular to monocular vision has clear detrimental effects on grasping.[?,?,?] Thus, one might expect that providing a haptic position cue should compensate for the loss of visual distance cues in conditions of limited vision. This is, in fact, what has been found.[25] Our finding that the haptic position cue reduces the grip aperture and increases movement velocity even in conditions of full binocular vision instead suggests that vision-for-action is less optimal than it is usually claimed. Second, our finding that holding the pole on which the object is placed (i.e., only haptic position cue available) leads to identical grasping performance as when both haptic size and position cues are available is not fully compatible with the theories about grasping that propose that either a single digit[?,?,?,?,?] or both digits[?,?] are transported to specific positions on the object. In the VHP condition, the positions on the object toward which the fingers were moving did not coincide with the haptic positions felt by the fingers of the left hand. Integrating these visual and haptic position estimates

should have thus led to a worse grasping performance than when these estimates were congruent (VH vs. VHP), which was however not the case. It is rather more plausible that multisensory grasping movements are planned and executed based on a visuo-haptic estimate of the object's centroid which is also used to better scale visual size.

In summary, our study provides a new perspective on the role of haptics in sensorimotor control by highlighting how non-visual inputs influence vision during object manipulation. Whether it be reaching for the cap of a handheld bottle, or passing a glass from one hand to the other, humans rely on haptics to guide movements more than we realize. Often, haptics can even take a leading role when vision is driven away from the manipulated object (e.g., when looking at a glass while uncorking a bottle). Hence, haptics is a remarkably underestimated source of information we draw from to guide efficient reaching and grasping movements in everyday life.

Data Availability

The authors can make the data available upon request.

References

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

2. Castiello, U. The neuroscience of grasping. *Nature Rev. Neurosci.* **6**, 726–36 (2005).

3. Janssen, P. & Scherberger, H. Visual guidance in control of grasping. *Annu. Rev. Neurosci.* **38**, 69–86 (2015).

4. Camponogara, I. & Volcic, R. Grasping movements toward seen and handheld objects. *Sci. Reports* **9**, 3665 (2019).

5. Pettypiece, C. E., Culham, J. C. & Goodale, M. A. Differential effects of delay upon visually and haptically guided grasping and perceptual judgments. *Exp. Brain Res.* **195**, 473–479 (2009).

6. Pettypiece, C. E., Goodale, M. A. & Culham, J. C. Integration of haptic and visual size cues in perception and action revealed through cross-modal conflict. *Exp. Brain Res.* **201**, 863–873 (2010).

7. Camponogara, I. & Volcic, R. Grasping adjustments to haptic, visual, and visuo-haptic object perturbations are contingent on the sensory modality. *J. Neurophysiol.* **122**, 2614–2620 (2019).

8. Epstein, W., Park, J. & Casey, A. The current status of the size-distance hypothesis. *Psychol. Bull.* **58**, 491–514 (1961).

9. van Damme, W. J. M. & Brenner, E. The distance used for scaling disparities is the same as the one used for scaling retinal size. *Vis. Res.* **37**, 757–764 (1997).

10. Brenner, E. & van Damme, W. J. M. Perceived distance, shape and size. *Vis. Res.* **39**, 975–986 (1999).

11. Volcic, R., Fantoni, C., Caudek, C., Assad, J. A. & Domini, F. Visuomotor adaptation changes stereoscopic depth perception and tactile discrimination. *J. Neurosci.* **33**, 17081–17088 (2013).

12. Volcic, R. & Domini, F. The endless visuomotor calibration of reach-to-grasp actions. *Sci. Rep.* **8**, 14803 (2018).

13. Langfeld, H. S. The differential spatial limen for finger span. *J. Exp. Psychol.* **2**, 416 (1917).

14. Gaydos, H. F. Sensitivity in the judgment of size by finger-span. *The Am. J. Psychol.* **71**, 557–562 (1958).

15. Durlach, N. *et al.* Manual discrimination and identification of length by the finger-span method. *Atten. Percept. & Psychophys.* **46**, 29–38 (1989).

16. Berryman, L. J., Yau, J. M. & Hsiao, S. S. Representation of object size in the somatosensory system. *J. Neurophysiol.* **96**, 27–39 (2006).

17. Proske, U. & Gandevia, S. C. The proprioceptive senses: Their roles in signaling body shape, body position and movement, and muscle force. *Physiol. Rev.* **92**, 1651–1697 (2012).

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

19. Gepshtein, S. & Banks, M. S. Viewing geometry determines how vision and haptics combine in size perception. *Curr. Biol.* **13**, 483–488 (2003).

20. Ernst, M. O. & Bülthoff, H. H. Merging the senses into a robust percept. *Trends Cogn. Sci.* **8**, 162–169 (2004).

21. Gepshtein, S., Burge, J., Ernst, M. O. & Banks, M. S. The combination of vision and touch depends on spatial proximity. *J. Vis.* **5**, 7–7 (2005).

22. Carey, D. P. & Allan, K. A motor signal and "visual" size perception. *Exp. Brain Res.* **110**, 482–486 (1996).

23. Battaglia, P. W. *et al.* Within-and cross-modal distance information disambiguate visual size-change perception. *PLoS Comput. Biol.* **6**, e1000697 (2010).

24. Sperandio, I., Kaderali, S., Chouinard, P. A., Frey, J. & Goodale, M. A. Perceived size change induced by nonvisual signals in darkness: The relative contribution of vergence and proprioception. *J. Neurosci.* **33**, 16915–16923 (2013).

25. Chen, J., Sperandio, I. & Goodale, M. A. Proprioceptive distance cues restore perfect size constancy in grasping, but not perception, when vision is limited. *Curr. Biol.* **28**, 927–932 (2018).

26. Derzsi, Z. & Volcic, R. MOTOM toolbox: MOtion Tracking via Optotrak and Matlab. *J. Neurosci. Methods* **308**, 129–134 (2018).

27. Nicolini, C., Fantoni, C., Mancuso, G., Volcic, R. & Domini, F. A framework for the study of vision in active observers. In Rogowitz, B. E., Pappas, T. N. & de Ridder, H. (eds.) *Human Vision and Electronic Imaging XIX*, vol. 9014 of *Proc. SPIE*, 901414 (2014).

28. R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria (2019). URL https://www.R-project.org/.

29. Volcic, R. & Domini, F. On-line visual control of grasping movements. *Exp. Brain Res.* **234**, 2165–2177 (2016).

30. Schot, W. D., Brenner, E. & Smeets, J. B. J. Robust movement segmentation by combining multiple sources of information. *J. Neurosci. Methods* **187**, 147–155 (2010).

31. Jeannerod, M. Intersegmental coordination during reaching at natural visual objects. In Berthoz, A., Long, J. B. & Baddeley, A. D. (eds.) *Attention and Performance IX*, 153–169 (Erlbaum, Hillsdale, NJ, 1981).

32. Bürkner, P.-C. brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.* **80**, 1–28 (2017).

33. Carpenter, B. *et al.* Stan: A probabilistic programming language. *J. Stat. Softw.* **76**, 1–32 (2017).

34. Schlicht, E. J. & Schrater, P. R. Effects of visual uncertainty on grasping movements. *Exp. Brain Res.* **182**, 47–57 (2007).

35. Karl, J. M. & Whishaw, I. Q. Haptic grasping configurations in early infancy reveal different developmental profiles for visual guidance of the reach versus the grasp. *Exp. Brain Res.* 3301–3316 (2014).

36. Gonzalez, C. L. & Sacrey, L.-A. R. The development of the motor system. In Gibb, R. & Kolb, B. (eds.) *The Neurobiology of Brain and Behavioral Development Ch 9.*, 235 – 256 (Academic Press, 2018).

37. Wallace, P. S. & Whishaw, I. Q. Independent digit movements and precision grip patterns in 1–5-month-old human infants: Hand-babbling, including vacuous then self-directed hand and digit movements, precedes targeted reaching. *Neuropsychol.* **41**, 1912–1918 (2003).

38. Banks, M. S. & Salapatek, P. Acuity and contrast sensitivity in 1-, 2-, and 3-month-old human infants. *Investig. Ophthalmol. & Vis. Sci.* **17**, 361–365 (1978).

39. Braddick, O. *et al.* Cortical binocularity in infants. *Nature* **288**, 363–365 (1980).

40. Held, R., Birch, E. & Gwiazda, J. Stereoacuity of human infants. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5572–5574 (1980).

41. Corbetta, D., Thurman, S. L., Wiener, R. F., Guan, Y. & Williams, J. L. Mapping the feel of the arm with the sight of the object: On the embodied origins of infant reaching. *Front. Psychol.* **5** (2014).

42. Thomas, B. L., Karl, J. M. & Whishaw, I. Q. Independent development of the reach and the grasp in spontaneous self-touching by human infants in the first 6 months. *Front. Psychol.* **5** (2015).

43. Corbetta, D., Wiener, R. F., Thurman, S. L. & McMahon, E. The embodied origins of infant reaching: Implications for the emergence of eye-hand coordination. *Kinesiol. Rev.* **7**, 10–17 (2018).

44. Connolly, J. D. & Goodale, M. A. The role of visual feedback of hand position in the control of manual prehension. *Exp. Brain Res.* **125**, 281–286 (1999).

45. Schettino, L. F., Adamovich, S. V. & Poizner, H. Effects of object shape and visual feedback on hand configuration during grasping. *Exp. Brain Res.* **151**, 158–166 (2003).

46. Winges, S. A., Weber, D. J. & Santello, M. The role of vision on hand preshaping during reach to grasp. *Exp. Brain Res.* **152**, 489–498 (2003).

47. Schenk, T., Mair, B. & Zihl, J. The use of visual feedback and on-line target information in catching and grasping. *Exp. Brain Res.* **154**, 85–96 (2004).

48. Rand, M. K., Lemay, M., Squire, L. M., Shimansky, Y. P. & Stelmach, G. E. Role of vision in aperture closure control during reach-to-grasp movements. *Exp. Brain Res.* **181**, 447–460 (2007).

49. Bozzacchi, C., Brenner, E., Smeets, J. B., Volcic, R. & Domini, F. How removing visual information affects grasping movements. *Exp. Brain Res.* **236**, 985–995 (2018).

50. Bingham, G. P., Coats, R. & Mon-Williams, M. Natural prehension in trials without haptic feedback but only when calibration is allowed. *Neuropsychol.* **45**, 288–294 (2007).

51. Mon-Williams, M. & Bingham, G. P. Calibrating reach distance to visual targets. *J. Exp. Psychol. Hum. Percept. Perform.* **33**, 645–656 (2007).

52. Coats, R., Bingham, G. P. & Mon-Williams, M. Calibrating grasp size and reach distance: Interactions reveal integral organization of reaching-to-grasp movements. *Exp. Brain Res.* **189**, 211–220 (2008).

53. Weigelt, C. & Bock, O. Adaptation of grasping responses to distorted object size and orientation. *Exp. Brain Res.* **181**, 139–146 (2007).

54. Weigelt, C. & Bock, O. Adaptation of the precision grip orientation to a visual-haptic mismatch. *Exp. Brain Res.* **201**, 621–630 (2010).

55. Bozzacchi, C., Volcic, R. & Domini, F. Effect of visual and haptic feedback on grasping movements. *J. Neurophysiol.*

Author contributions statement

I.C. conducted the experiment and analyzed the data. I.C. and R.V. designed the experiment and wrote the manuscript.

Additional information

Competing interests The authors declare no competing interests.

Correspondence and requests for materials should be addressed to I.C.