



A visual–haptic Necker cube reveals temporal constraints on intersensory merging during perceptual exploration

Nicola Bruno^{a,*}, Alessandra Jacomuzzi^a, Marco Bertamini^b, Georg Meyer^b

^a *Dipartimento di Psicologia and BRAIN Center for Neuroscience, Università di Trieste, via S. Anastasio 12, 34134 Trieste, Italy*

^b *School of Psychology, University of Liverpool, Liverpool, UK*

Received 25 August 2005; received in revised form 18 January 2006; accepted 30 January 2006

Abstract

When viewing a three-dimensional Necker cube with one eye, participants can experience illusory reversals even while they feel the cube with their hands. This surprising property of the visual–haptic Necker cube affords a unique opportunity to investigate temporal constraints on interactions between vision and touch during extended observation of a three-dimensional object. Our observers reported reversals while they viewed the cube and, at the same time, they either held it with two-finger grips, felt it with while their hands remained stationary, or actively explored it by moving one hand. Consistent with a multisensory approach to three-dimensional form perception, touch had a clear effect on both the number and the duration of illusory percepts. Additionally, when observers alternated between stationary and moving periods during exploration, transitions from stationary to moving-hand haptics played a crucial role in inhibiting illusory reversals. A temporal analysis of the probability of first reversals occurring after different types of motor transition revealed a “vetoing window” initiating approximately 2 s after the transition and lasting at least another 1–2 s. Implications for multisensory processes during exploration are discussed.

© 2006 Published by Elsevier Ltd.

Keywords: Vision; Touch; Haptics; Perceptual exploration; Intersensory conflict; Intersensory merging; Intersensory vetoing

1. Introduction

Perceiving the three-dimensional structure of an object often involves merging vision and haptics over extended periods of exploration. An interesting feature of this process is that as exploration progresses, new information may require changing how the two sensory signals are treated. Suppose you were looking at a simple object, say a cup, while you also explore it with one hand. If the hand is feeling the front of the cup, eye and hand inform about the same properties (such as local curvature, size and so on). In this case merging the two sensory signals would be justified, and presumably advantageous. But if, instead, the hand touches in the back of the cup, haptics may detect properties that are not available to vision. For instance, a change in surface curvature at the junction with the cup's handle, or a differently shaped cup nearby. That the two signals should be merged is now less obvious. In many such cases, in fact, the cor-

rect decision would be that the two signals are not to be merged at all. What process makes this kind of decisions in the human perceptual system?

Technically, the problem of handling intersensory discrepancies that arise during bimodal exploration may be solved by different strategies (for a recent review, see Ernst & Bulthoff, 2004). For instance, the system may merge the two signals by performing a weighted sum of the bimodal signals (bimodal integration). It is generally believed that such integration tends to occur for signals at similar spatial and temporal positions (see Stein & Meredith, 1993) and that the weights entered in the computation are based on the relative reliability of the two sensory channels (see Ernst & Banks, 2002). As an alternative, different bimodal signals may be handled by a more complex operation whereby complementary aspects of bimodal information are coordinated (bimodal combination). For instance, the perception of three-dimensional shape may combine information about the back of an object, which is typically acquired by touch, with information about its front, which is readily available to vision (Newell, Ernst, Tjan, & Bühlhoff, 2001). Finally, discrepant signals may be dealt with using internally represented

* Corresponding author.

E-mail address: nick@psico.univ.trieste.it (N. Bruno).

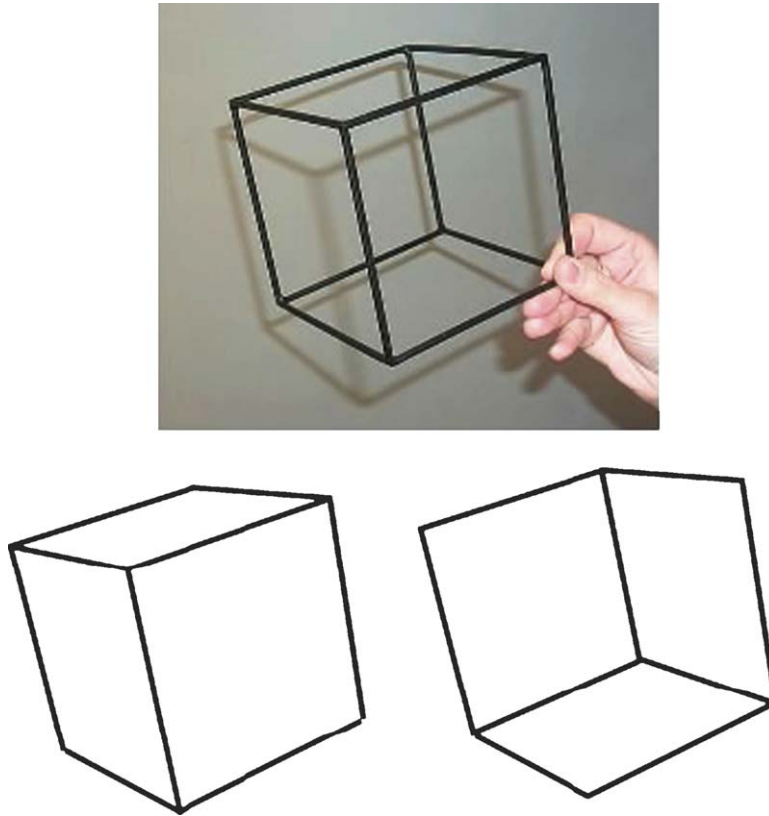


Fig. 1. Top: photograph of a 3D model of a Necker cube that can be held in the hands. Bottom: drawings of its two alternative interpretations.

knowledge that one sensory channel is more trustworthy under certain conditions (i.e. the “modality appropriateness” hypothesis of Welch & Warren, 1986, Chapter 25; see also Jacobs, 2002). Such a priori bias in favor of one channel may cause the other channel to be discarded (i.e. the “visual capture” observed by Rock & Victor, 1964).

It is currently unclear whether the human perceptual system uses all of these strategies to process discrepant signals. There is strong evidence that merging consistent sensory signals is often modelled very well by an integration approach (Alais & Burr, 2004; Ernst & Banks, 2002; van Beers, Sittig, & van der Gon, 1999). This scheme may be extended to deal with inconsistent signals in several ways. For instance, the system may monitor changes in the quality of sensory signals as conditions change during exploration. This could then result in intersensory reweighting (Gepshtein & Banks, 2003) or recalibration (Ernst, Banks, & Bühlhoff, 2000). These processes would effectively give greater importance to the most reliable of the discrepant signals. Note that reweighting that assigns a near-zero weight to one of the channels is equivalent to discarding it. Note also, however, that reweighting or recalibration may just as well be performed on the basis of a priori biases. For instance, in many situations, the system may be biased to use haptic information as the standard for recalibrating visual inputs (Atkins, Fiser, & Jacobs, 2001). This process is reminiscent of earlier theories in philosophy (Berkeley, 1709) and cognitive psychology (Piaget, 1937). Given alternative mechanisms for reweighting and recalibration, we need further information about bimodal processes

during exploration before we can distinguish between candidates.

In this paper, we intend to investigate this issue by studying sensory discrepancies in a visual–haptic Necker cube. The Necker cube is a well-known reversible figure (Necker, 1832). Less well known is that reversals occur, under monocular viewing, also with actual 3D models of the cube (Fig. 1) and even when such 3D models are explored haptically (Shopland & Gregory, 1964). This is striking, when one considers the perceived 3D alternatives. One of these is of course a cube, the veridical shape, which matches the shape felt by the hands. The other, however, is a truncated pyramid pointing in the opposite direction relative to that felt by the hands. When experiencing this second percept, one somehow has the impression that the cube loses its rigidity, or that one’s wrists are bent at impossible angles, consistent with the visually reversed shape instead of the haptically felt one. Odd as they are, these experiences seem to be due to some kind of bimodal process. Evidence for this conclusion is provided by changes in the frequency of reversals as well as durations of perceived alternatives. For instance, reversal frequency decreases when seeing and touching the cube, relative to when one sees it but cannot touch it (Shopland & Gregory, 1964). The average duration of bimodally consistent percepts is larger than that of inconsistent percepts (Ando & Ashida, 2003).

The above effects suggest that the visual–haptic Necker cube is an excellent model to investigate bimodal processes during extended periods of exploration. For instance, although natural

objects do not ordinarily reverse in depth, spontaneous reversals in the cube provide an interesting opportunity to assess adaptive processes that take place when previously consistent bimodal signals begin to conflict. As we have argued at the beginning of the paper, such conflict can take place when the hand explores locations that are not immediately visible, such as those in the back of objects. In addition, by tracking reversals as they are experienced under different haptic conditions, one could obtain information about bimodal processes occurring when the quality of information provided by separate sensory channels changes over time.

To address these questions, we present two coordinated studies. In the first, we varied the quality of information for 3D shape provided by haptics by changing conditions across separate sessions. The aim was to replicate known haptic effects on Necker cube reversals and percept durations (Ando & Ashida, 2003; Shopland & Gregory, 1964) and to confirm that these effects are indeed due to haptics (and not to confounded visual changes). In the second study, we varied tactile information within sessions by asking observers to alternate between periods of active and passive touch. This second study was aimed at obtaining information about the temporal dynamics of bimodal interactions during the exploration of the cube. In both studies, the perceived three-dimensional form was assessed by asking observers to verbally report reversals as they experienced them.

2. Methods

2.1. Participants

A total of 17 participants were included in the studies. Six (including the first two authors) served in the intermodal conditions of the first study as well as in the second study. An additional six participants took part in two control unimodal conditions of the first study. Finally, another five participants (including the third and fourth author) served only in the second study. All participants were either faculty members or graduate students in the Trieste or Liverpool departments and all gave their informed consent prior to their inclusion in the studies. With the exception of the authors, all other participants were fully naïve to the purpose of the studies. All were right-handed and either had normal vision or wore prescription lenses as was appropriate for them.

2.2. Materials and stimuli

The visual and haptic stimulus consisted of a wire-frame cube (side = 12.5 cm) made of thin iron bars (diameter = 4 mm). The frame was spray-painted with matte black colour. To minimize brightness differences due to directional illumination, a translucent semicircular screen was constructed and used as background during experimental sessions. A standard commercial video camera was used to record the participant's hands during the session as well as their vocal productions when reporting reversals (main conditions of study 1 and study 2). The scene camera mounted on the helmet of an ASL 5000 eye movement recording system was used to record views of the cube and of hands holding it from the viewpoint of the participant (control conditions of study 1). Camera output was fed on a PowerBook G4 Macintosh computer where the recordings were stored as multimedia files (.mov). Participants wore modified goggles with an opaque screen occluding the left eye. The goggles were constructed in order to permit wearing prescription glasses underneath, if needed. Finally, the three drawings in Fig. 2(a) were used in the main conditions of study 1 and in study 2 to show participants how they were required to hold the cube in different sessions and to insure that all had approximately the same monocular view of the cube.

2.3. Procedure and experimental conditions

The studies were performed in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki, as well with the guidelines for research involving human participants provided by the Universities of Trieste and Liverpool.

Participation in the experiments was preceded by training sessions. These served the purpose of demonstrating Necker cube reversals, of insuring that participants could identify them when they occurred, and of standardizing the verbal responses that were recorded. Once the training session was concluded, participants began their experimental sessions. In the three intermodal conditions of the first study, these consisted of five 1-min runs for each of the three experimental conditions described below, in random order. In the two control unimodal conditions of the first study, they consisted of four 2-min runs for each of two of the three main conditions. In the second study, finally, they consisted of four 2-min runs consisting of alternations between hand-stationary and hand-moving periods. Participants were allowed rest periods between runs, if they requested them. At the end, naïve participants were debriefed regarding the aims of the study.

In the first study, the training session began by showing participants the 3D Necker cube. Once participants noticed that they could invert the cube under monocular viewing, we drew their attention on how the alternating percepts corresponded to different positions in depth as well as to different 3D shapes. After this, we told them that in the study they were going to hold the cube in their hands. To illustrate the specific manners of holding the cube, we showed participants drawings (see Fig. 2(a)) that reproduced the monocular views they were required to hold. At this point, we explained participants that they were to report inversions of the 3D cube during prolonged viewing. We instructed them to say the word "inverted" as soon as the cube turned into a truncated trapezoid and to say "normal" as soon as the trapezoid turned into a cube.

To insure that they had understood the instructions and to familiarize them with the task, we first asked participants to try holding the cube with two-finger grips. This constituted the baseline condition, which is illustrated in the top panel of Fig. 2(a). After they reported reversals over a period of about 1 min, we requested them to cup their hands over two vertices of the cube as in the hand-stationary condition, which is illustrated in the middle panel of Fig. 2(a). After they experienced reversals in this new condition, we requested them to start moving the right hand as shown by the arrows in the bottom panel of Fig. 2(a), that is, to continuously explore the three sides that converged at the top right vertex of the cube. They were requested to remain on these three sides, however, and to avoid touching other vertices by bringing their hand to the front or to the back of the cube. This last mode of touching the cube defined the hand-moving condition. Training ended as soon as observers realized that they could experience reversals even while moving the hand.

In the two unimodal control conditions of study 1, training was performed exactly as in the three intermodal conditions. However, participants did not hold the cube in their hands, but simply looked at the cube and at the hands of an actor. The cube and the hands were presented in a video taken from a viewpoint that mimicked what the participants would have seen, if they had been holding the cube (see, again, the views in Fig. 2(a)). Thus, the videos used in this control condition reproduced the visual stimuli provided by two of the three intermodal conditions: the baseline condition and the hand-moving condition.

In the second study, the training session was the same as for the first study, plus an additional part at the end. This additional part served to familiarize participants with the task of the second study, which involved alternating between hand-stationary and hand-moving periods. Participants were told that the experimenter was to give them verbal instructions as to when to start or stop the hand movement at pseudorandom times. Participants reported reversals, as in the first part, over a period of about 1 min.

2.4. Data recording and analysis

Video files were inspected on a frame-by-frame basis using QuickTime™ Player Version 7.0.1 on a PowerBook G4 Macintosh computer. In both studies, vocal productions reporting Necker cube reversals were identified and their timing within the session was recorded in a spreadsheet for further analysis. In the second study, the .mov files were further inspected to identify transitions

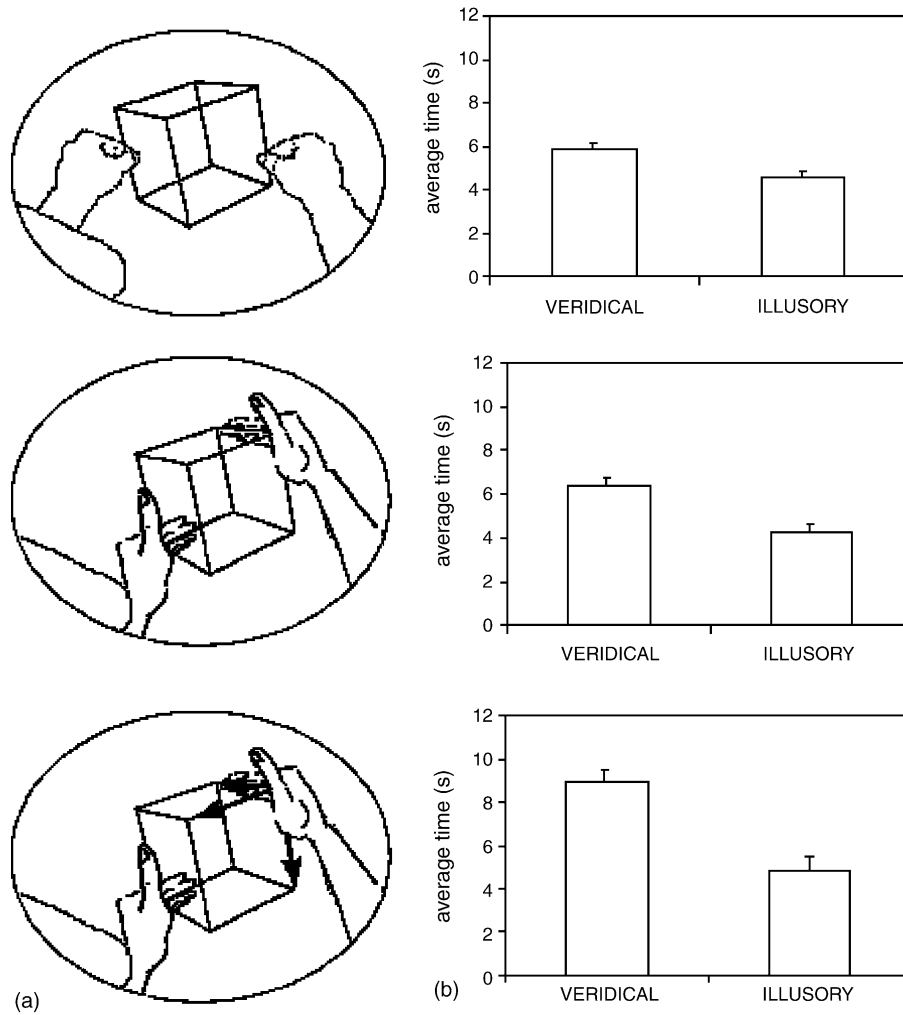


Fig. 2. First study, (a) schematic representations of the participant's right visual field in the experimental conditions; top, baseline condition; middle, hand-stationary condition; bottom, hand-moving condition; (b) average percept durations in the three experimental conditions. The drawings were actually used during the experiment to instruct participants on how to hold the cube. Error bars are 1 S.E.M.

from active to passive touch, or from passive to active touch. The timing of these transitions was also entered in the spreadsheet. To estimate probabilities, all timings were binned by rounding down to the nearest second. The type and timing of reported reversals, as entered in the spreadsheet, were also analyzed to estimate the number of reversals and the duration of periods whereby participants experienced a 3D cube (the "veridical" percept) or its reversed counterpart, a truncated pyramid (the "illusory" percept). Parametric analyses of these data were performed using Data Desk[®] Version 6.2.

3. Results and discussions

3.1. First study, intermodal conditions

In the baseline intermodal condition, the reversal rate was 68.4 min^{-1} . This remained almost unchanged in the hand-stationary condition, where it reduced only to 67 min^{-1} . Conversely, there was a substantial reduction in the hand-moving condition, where the reversal rate was 51 min^{-1} . The total durations of the veridical percept, computed over an average 1 min session, were 34.3, 37.4 and 40.6 s for the baseline, hand-stationary and hand-moving conditions, respectively. The bar charts in Fig. 2(b) present average durations for the two alter-

natives (veridical and illusory) and the corresponding standard errors in the three experimental conditions.

Before parametric analysis, the duration data were subjected to a transformation to correct for a marked asymmetry in the shape of their distribution (skewness = 3.9). Such asymmetry is typical of percept durations in reversible figures, which are well approximated by gamma distributions (Borsellino, De Marco, Allazetta, Rinesi, & Bartolini, 1972; see also Mamassian & Goutcher, 2005). To reduce violations of the normality assumption and improve the analysis (see, for instance, Mosteller & Tukey, 1977; Snedecor & Cochran, 1980, Chapter 15), we subjected the durations to a Box-Cox transformation (see Box & Cox, 1964) varying the transformation exponent until we found the value that minimized the observed skewness. This turned out to be equal to -0.057 (skewness = -0.003). Accordingly, we used the transformed durations rather than the original data to perform a 2 (percept type, veridical, or illusory) \times 3 (experimental condition, baseline, hand-stationary, or hand-moving) repeated-measures analysis of variance.

The analysis of variance yielded a significant main effect of percept type, $F(1, 5) = 7.7$, $p < 0.04$. This finding suggests

that, in all conditions, the duration of a veridical percept was on average longer (by about 3 s) than the duration of an illusory percept. The cause of this “veridical bias” may be traced back to several factors, such as the preference for regular 3D shapes (i.e. a cube) over less regular alternatives (a truncated pyramid), slight differences between the retinal size of the near and the far bars, or subtle cues about the true 3D shape of the object that may be provided by sensing the distribution of weight on the object, which would be slightly different for a cube and a truncated pyramid.

The analysis of variance also yielded a significant main effect of experimental condition, $F(2, 10) = 3.4, p < 0.05$. Post hoc pairwise comparisons using Tukey’s LSD measure demonstrated that the average duration of the alternative percepts was longer in the hand-moving condition (about 10 s) than in the other two conditions (about 6 s), both p ’s < 0.001 , whereas percept durations in the hand-stationary and the baseline condition were not statistically distinguishable, $p = 0.65$. Inspecting Fig. 2(b) suggests that this effect was due to an increase of the duration of the veridical percept, relative to its duration in the baseline and hand-stationary conditions, while the durations of the illusory percept remained essentially unchanged across conditions. Although the two-way interaction between percept type and experimental condition technically failed to reach significance, $F(2, 8) = 2.75, p = 0.06$, post hoc contrasts of the interaction simple effects supported this interpretation. The veridical percept lasted longer, on average, in the hand-moving than in the other two conditions, $p < 0.001$ or smaller. Conversely, the illusory percept did not differ statistically across the three conditions, $p > 0.12$ or bigger.

3.2. First study, unimodal control conditions

Given that participants could see their hands touching the cube, effects observed in the three intermodal conditions of the first study may be due to changes in the visual stimulus. We considered this possibility unlikely because Ando and Ashida (2003) reported similar effects while using a virtual reality system that prevented their participants from seeing their hands. However, to completely rule out this possibility, we run two unimodal control conditions using another six observers. These control conditions closely corresponded to the baseline and the hand-moving intermodal conditions. However, participants did not hold a cube but watched videos of an actor holding the cube or moving the hand on it. Given that these videos were taken from the viewpoint of someone holding the cube, they faithfully reproduced the visual stimulus one would have seen when holding the cube, but of course they provided no haptic information whatsoever.

The pattern of results in these two unimodal controls was markedly different from that of the corresponding intermodal conditions. First, the frequency of reversals was almost exactly the same: 63.6 and 63.4 m^{-1} , in the baseline and the hand-moving conditions, respectively. Second, the total durations of the veridical percept per average 1 min session were also the same: 36.9 and 36.6 s, in the same order. Third, and most important, the difference between the average durations of the alternative percepts did not change between the baseline (6.7 s

Table 1

Second study, total frequencies of veridical to illusory ($v \rightarrow i$) and of illusory to veridical ($i \rightarrow v$) reversals when the hand was stationary and when it moved

	Hand	
	Stationary	Moving
($v \rightarrow i$)	186	118
($i \rightarrow v$)	142	138

versus 5.4 s) and the hand-moving condition (6.8 s versus 5.7 s). Accordingly, an analysis of variance on transformed duration data (see description of transformation in the previous section) did not reveal statistically significant effects, although the bias in favor of the veridical percept came close to significance, $F(1, 5) = 4.5, p < 0.087$.

3.3. Second study

In the second study, we first computed the total number of each type of reversal, that is, reversals from the veridical to the illusory percept ($v \rightarrow i$) or from the illusory to the veridical percept ($i \rightarrow v$) and separated those occurring when the hand was stationary from those occurring when the hand moved. Note that in the case of this second study, it makes little sense to compute percept durations as any given percept could be experienced partly during hand-stationary and partly during hand-moving periods. Summing across all eleven participants, we observed about 600 reversals. Specifically, there were 304 ($v \rightarrow i$) reversals, 186 occurring when the hand was stationary and 118 when it moved, and 280 ($i \rightarrow v$) reversals, 142 occurring during stationary periods and 138 during moving periods (see Table 1). To test the association between reversal type and haptic condition in these data, we assumed reversal independence (as supported by Zhou et al., 2004) and computed $\chi^2(1) = 6.49, p < 0.02$.

Next, we computed first reversals occurring after transitions that increased or decreased the quality of haptic information, that is, transitions from hand-stationary to hand-moving ($s \rightarrow m$), or from moving to stationary ($m \rightarrow s$). There were about 300 such reversals, indicating that other reversals could take place after the first, and before a new motor transition occurred. Specifically, there were 171 ($v \rightarrow i$) first reversals, 67 occurring after ($s \rightarrow m$) transitions and 104 after ($m \rightarrow s$) transitions, and 118 ($i \rightarrow v$) reversals, 72 after ($s \rightarrow m$) transitions and 46 after ($m \rightarrow s$) transitions (see Table 2). To test the association between reversal type and touch transition in this data, we computed $\chi^2(1) = 13.3, p < 0.0003$.

Table 2

Second study, frequencies of veridical to illusory ($v \rightarrow i$) and of illusory to veridical ($i \rightarrow v$) reversals occurring first after transitions from stationary to moving ($s \rightarrow m$) and from moving to stationary ($m \rightarrow s$)

	Transition	
	($s \rightarrow m$)	($m \rightarrow s$)
($v \rightarrow i$)	67	104
($i \rightarrow v$)	72	46

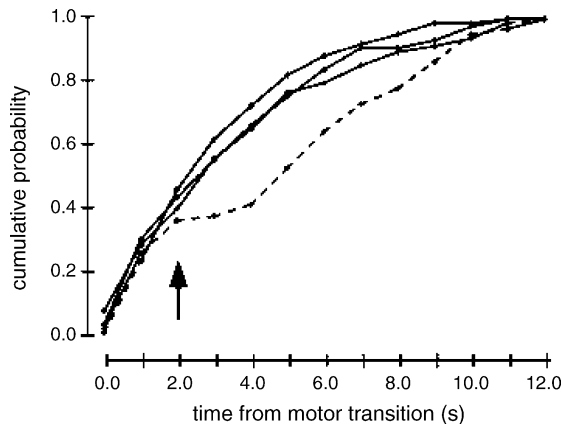


Fig. 3. Second study, cumulative probabilities of experiencing a reversal from veridical to illusory ($v \rightarrow i$) or from illusory to veridical ($i \rightarrow v$), after a transition from hand-stationary to hand-moving ($s \rightarrow m$) or from moving to stationary ($m \rightarrow s$). Continuous lines, curves for ($i \rightarrow v$)/($s \rightarrow m$), ($i \rightarrow v$)/($m \rightarrow s$) and ($v \rightarrow i$)/($m \rightarrow s$). These curves were well fit by almost identical cumulative gamma functions. Dashed line, curve for ($v \rightarrow i$)/($s \rightarrow m$). This curve was fit more poorly by a markedly different gamma function. Note that after about 2 s from the onset of hand movement, during a “vetoing window” of at least another 2 s it was essentially impossible to experience veridical to illusory reversals.

Finally, to test how changes in haptic quality affected the probability of a given reversal over time, we plotted the cumulative probability of each type of reversal given each type of transition, as a function of the temporal delay from the motor transition itself (see Fig. 3). Consider, for instance, the probability of ($v \rightarrow i$)/($s \rightarrow m$) within 1 s from the transition. This is estimated by the frequency of ($v \rightarrow i$)/($s \rightarrow m$) divided by the total of ($v \rightarrow i$)/($s \rightarrow m$) events. At each 1 s interval, the cumulative probability is then given by the sum of probabilities up to that interval, divided by the total. As can be seen from the figure, this plot revealed that the probability of all reversals tended to decrease with time following a smooth negatively accelerated curve, except for the ($v \rightarrow i$) reversals following a ($s \rightarrow m$) transition. In this case, the cumulative probability curve had a markedly different shape. More precisely, the plot demonstrated that after having reached a value of ≈ 0.4 at the 2 s bin (following a trend comparable to the other three curves), the cumulative probability curve for ($v \rightarrow i$)/($s \rightarrow m$) stopped growing, and remained fixed at approximately 0.4 for another 2 s. To evaluate the differences between this latter curve and the other two, we fitted cumulative gamma functions to the observed cumulative probabilities. As expected, three of the four curves showed excellent fits, $0.0022 < \text{RMSE} < 0.0136$, except for the ($v \rightarrow i$)/($s \rightarrow m$), $\text{RMSE} = 0.0638$. To insure that this difference applied equally to the naïve observers and to the four authors, we also replotted the data separately for the two groups. These plots were very similar, with the curve for ($v \rightarrow i$)/($s \rightarrow m$) similarly halting at the 2 s bin. Additional analyses confirmed that gamma functions fitted the data equally well, in the case of the first three curves, or equally badly, for that of the ($v \rightarrow i$)/($s \rightarrow m$) curve, in both groups. Specifically, we found that for the first three curves $0.0096 < \text{RMSE} < 0.0280$ and $0.0059 < \text{RMSE} < 0.0272$, whereas for the ($v \rightarrow i$)/($s \rightarrow m$) curve $\text{RMSE} = 0.0794$ and 0.0592 in the author and naïve groups,

respectively. The similar performance profiles of the four authors and the seven naïve participants is not surprising, given that the specific shape of the curves in these graphs was not expected, and only came to our attention after the analysis.

Finally, we can exclude that these effects are due to a non-specific, task-irrelevant motor activity. If the reversals were inhibited by simply moving the hand, independent of haptic shape information, then after a transition from stationary to moving we should have observed vetoing of ($v \rightarrow i$) reversals, as we did, but also of ($i \rightarrow v$) reversals, which we did not. In fact, the cumulative probability curve for this latter case was identical to the curves involving transitions from moving to stationary. Thus, these results suggest that changes that increased haptic quality prevented participants from experiencing illusory reversals, but only within a specific “vetoing window” that occurs in our data after about 2 s from the motor transition and lasts for an additional 1–2 s. A natural interpretation for this pattern is that the delay before the onset of the inhibitory period reflects the time required for haptic information to build up and enter the intersensory merging process, as one would expect if the discrepancy was handled by a process that registers the increase in quality of the haptic signal, and acts accordingly.

4. Conclusions

When participants explored the visual–haptic Necker cube employed in the present studies, they obtained information that could be either consistent or inconsistent across the two modalities. Specifically, when the visual signal supported a truncated pyramid (the illusory percept), the tactile signal conflicted with this interpretation in supporting a cube. Conversely, when the visual signal also supported a cube, the two signals agreed in supporting the same three-dimensional interpretation. In addition to changing visual information, during exploration our observers also received haptic information that varied in quality. For instance, when a stationary period was followed by the initiation of hand movement, our participants experienced an increase in the quality of the haptic information about three-dimensional form. When motion ceased, they experienced a decrease in haptic quality.

Our results suggest that the system was sensitive to these changes when handling discrepant information during exploration. Haptic information obtained by moving the hand on the cube tended to make veridical percepts more durable (as already observed by Ando & Ashida, 2003) and consequently reversals somewhat less frequent (as already observed by Shopland & Gregory, 1964). These effects appear to depend on intersensory vetoing of the illusory interpretation, occurring about 2 s after increases in haptic quality due to the onset of hand motion. In addition, such intersensory vetoing was not ever-lasting, but appeared to completely prevent normal to illusory reversals only for about 1–2 s. Thus, our findings suggest that, at least in the present conditions, intersensory discrepancies were dealt with by monitoring fluctuations in sensory signal quality over specific temporal windows, and by accordingly adjusting their relative importance in the merging process.

A number of papers have proposed temporal integration windows in processes of multisensory integration (Anastasio, Patton, & Belkacem-Boussaid, 2000; Colonius & Diederich, 2004), possibly involving cortical modulation of multisensory responses in the superior colliculus (Jiang & Stein, 2003). These proposals are often contrasted to approaches based on computing the statistics of population of neurons (Deneve, Latham, & Pouget, 1999; Pouget, Dayan, & Zemel, 2003), which would afford faster estimates of signal reliabilities for intersensory reweighting (Ernst & Banks, 2002). However, computing reliabilities in such a fashion is not trivial, and may be problematic if haptic changes involve large modifications in the neuronal populations involved. For instance, a PET study by Fink et al. (1999) suggests that intersensory conflicts involving comparisons between motor intentions and sensory information may be handled by different cortical structures than those consisting of mere conflicts between simple sensory information. When such differences are big, resorting to measurements of fluctuations within temporal windows may be an adaptive, even if slower, strategy.

A last, interesting feature of our results is that the observed vetoing window was fairly narrow, completely preventing illusory percepts only for about 1–2 s. It is possible that the system is not continuously measuring the quality of the sensory signal, but is instead sensitive to changes in this quality. Once these changes are registered, adaptation occurs that effectively eliminates the vetoing effect. Note that in a less constrained haptic task, one would presumably assume more varied hand positions and perform larger range motions, such that the quality of the information would vary in a more continuous fashion. Continuous change would prevent adaptation and therefore produce longer lasting vetoing effects. As an alternative, if the visual and tactile signals are processed fully and kept separate (Hillis, Ernst, Banks, & Landy, 2002), it is possible that participants stopped attending to the haptic signal some time after the transition, and switched to a “vision-only” attentional mode. Monitoring the pattern of eye fixations within and after the vetoing window may provide information about such attentional switches. Experiments aimed at measuring fixations during manual exploration of the visual–haptic Necker cube are currently under way in our laboratories and will be the subject of future reports.

References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262.
- Anastasio, T. J., Patton, P. J., & Belkacem-Boussaid, K. (2000). Using Bayes' rule of model multisensory enhancement in the superior colliculus. *Neural Computation*, *12*(11), 65–1187.
- Ando, H., & Ashida, H. (2003). Touch can influence visual depth reversal of the Necker cube. *Perception*, *32*, 97.
- Atkins, J. E., Fiser, J., & Jacobs, R. A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Research*, *41*, 449–461.
- Berkeley, G. (1709). *An essay towards a new theory of vision*. London: J.M. Dent.
- Borsellino, A., De Marco, A., Allazetta, A., Rinesi, S., & Bartolini, B. (1972). Reversal time distribution in the perception of ambiguous stimuli. *Kybernetik*, *10*, 139–144.
- Box, G. E., & Cox, D. R. (1964). An analysis of transformation. *Journal of the Royal Statistical Society B*, *26*, 211–243.
- Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: A time-window-of-integration model. *Journal of Cognitive Neuroscience*, *16*, 1000–1009.
- Deneve, S., Latham, P. E., & Pouget, A. (1999). Reading population codes: A neural implementation of ideal observers. *Nature Neuroscience*, *2*, 740–745.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–432.
- Ernst, M. O., Banks, M. S., & Bühlhoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience*, *3*, 69–73.
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*, 162–169.
- Fink, G. R., Marshall, J. C., Halligan, P. W., Frith, C. D., Driver, J., Frackowiak, R. S. J., et al. (1999). The neural consequences of conflict between intention and the senses. *Brain*, *122*, 497–512.
- Gepshtein, S., & Banks, M. S. (2003). Viewing geometry determines how vision and haptics combine in size perception. *Current Biology*, *13*, 483–488.
- Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science*, *298*(5598), 1627–1630.
- Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in Cognitive Sciences*, *6*, 345–348.
- Jiang, W., & Stein, B. E. (2003). Cortex controls multisensory depression in superior colliculus. *Journal of Neurophysiology*, *90*, 2123–2135.
- Mamassian, P., & Goutcher, R. (2005). Temporal dynamics in bistable perception. *Journal of Vision*, *5*, 361–375. doi:10.1167/5.4.7. <http://journalofvision.org/5/4/7/>.
- Mosteller, F., & Tukey, J. (1977). *Data analysis and regression*. Reading, MA: Addison–Wesley.
- Necker, L. A. (1832). Observations on some remarkable phenomena seen in Switzerland; and an optical phenomenon which occurs on viewing of a crystal or geometric solid. *Philosophy Magazine*, *3*, 329–337.
- Newell, F. N., Ernst, M. O., Tjan, B., & Bühlhoff, H. H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological Science*, *12*, 37–42.
- Piaget, J. (1937). *La construction du réel chez l'enfant*. Neuchâtel: Delachoux et Niestle.
- Pouget, A., Dayan, P., & Zemel, R. (2003). Computation and inference with population codes. *Annual Reviews of Neuroscience*, *1*, 381–410.
- Rock, I., & Victor, J. (1964). Vision and touch: An experimentally created conflict between the two senses. *Science*, *143*, 594–596.
- Shopland, J. C., & Gregory, R. L. (1964). The effect of touch on a visually three-dimensional figure. *Quarterly Journal of Experimental Psychology*, *16*, 66–70.
- Snedecor, G. W., & Cochran, W. G. (1980). *Statistical methods* (7th ed.). Ames, IA: The Iowa State University Press.
- Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. Boston, MA: MIT Press.
- van Beers, R. J., Sittig, A. C., & van der Gon, D. (1999). Integration of proprioceptive and visual position-information: An experimentally supported model. *Journal of Neurophysiology*, *81*, 1355–1364.
- Welch, R. B., & Warren, D. H. (1986). Intersensory interactions. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance*. New York: Wiley.
- Zhou, Y. H., et al. (2004). Perceptual dominance time distributions in multistable visual perception. *Biological Cybernetics*, *90*, 256–263.