S H O R T   C O M M U N I C A T I O N S

ROCZNIKI PSYCHOLOGICZNE / ANNALS OF PSYCHOLOGY
2013, XVI, 2, 355-365

MAGDALENA SZUBIELSKA
The John Paul II Catholic University of Lublin
Institute of Psychology
KATARZYNA JAROSZEK
BARTŁOMIEJ KILJANEK

THE COSTS OF CHANGING MODALITY IN VISUO-HAPTIC RECOGNITION OF SCENES

The experiment is aimed at investigating the factors that may modulate the costs of cross-modal visuo-haptic recognition of scenes. Participants learned a scene either visually or by touch (in the latter case they were blindfolded); then, following a delay, they identified the scene using the same or changed modality. The level of difficulty was adjusted by introducing two or three changes in the placement of scene elements at the recognition stage. It has been demonstrated that the costs of modality change, related to both decreased accuracy of recognition and extended time for making decision, occur only in a situation when a significant burden is imposed on working memory, i.e., with tactile learning of a scene and a high level of difficulty of the recognition task.

Keywords: cross-modal recognition, vision, touch.

THEORETICAL INTRODUCTION

Researchers investigating spatial images (e.g., Loomis et al., 2012) define these as representations either created as a result of visual, auditory, and tactile perception of three-dimensional space or originating from long-term memory. These representations function within working memory and contain information on the spatial properties (e.g., location or orientation) of an isolated stimulus or a set of stimuli. Research on the effect of intermodal priming (Easton, Srinivas,

Magdalena Szubielska – Institute of Psychology, The John Paul II Catholic University of Lublin, Al. Racławickie 14, 20-950 Lublin; e-mail: magdasz@kul.pl

& Greene, 1997; Easton, Srinivas, & Greene, 1997; Reales & Ballesteros, 1999) suggests that cognitive representations based on visual and haptic perception share the same information related to structural dimensions. Similar cortical areas, called visual areas, are involved in the visual and tactile recognition of objects (Amedi et al., 2001). During an experiment using fMRI, Lacey et al. (2010) demonstrated that activation caused by visual object imagery was more strongly correlated to activation during haptic perception of semantic objects in comparison with that of non-semantic objects (e.g., rubber duck vs. abstract shape). Furthermore, it is known that sighted individuals report using visualization strategy in the course of tactile learning of small non-semantic spaces (e.g., Szubielska, 2009). Therefore, a conclusion can be made about the visual nature of spatial imagery evoked by either vision or touch in the population of sighted individuals. Mental visualization of spatially complex tactile stimuli is connected with allocentric representation dominating in sighted individuals (see e.g. Pasqualotto & Proulx, 2012). However, there are haptic tasks, such as judgment of parallelity between bars, where sighted participants are found to use less complex egocentric representation based on body-centered signals and those encoded by movement. In the case of such tasks, a shift to allocentric representation may be facilitated by a delay between learning and test (Zuidhoek, Kappers, van der Lubbe, & Postma, 2003) and by allowing participants to view stimuli unrelated to the task (Newport, Rabb, & Jackson, 2002).

Despite the fact that both visual and haptic perception may lead to creating spatial images, it seems that the coding of spatial information in these two modalities occurs in different ways. This is confirmed by the costs of changing modality in examinations of cross-modal visuo-haptic recognition. Experiments focusing on remembering faces (Casey & Newell, 2007), isolated objects (Ernst, Lange, & Newell, 2007), and scenes (Newell et al., 2005) showed that a change of modality at the stage of recognition to one different from that used at the stage of learning resulted in poorer accuracy of recognition compared with a situation when the same modality was applied at both stages. Yet, these costs are not always observed.

Newell et al. (2001) did not report lower accuracy resulting from changed modality in a task involving the recognition of isolated objects, i.e., shapes built of LEGO bricks, when these were seen from one viewpoint. When at the stage of learning the same shapes were perceived from multiple viewpoints, a change in modality significantly decreased the accuracy of recognition (Ernst, Lange, & Newell, 2007). In a series of three experiments concerning the recognition of scenes, Newell et al. (2005) found, in two trials, an effect of changed modality involving an increased number of errors in recognition under cross-modal conditions. This effect was not fully confirmed in only one of these experiments, where the articulatory suppression variable was introduced. In that study, the cost of modality change was found when compared with the condition of visual learning and recognition. According to these authors, visual perception of space imposes less burden on working memory than tactile perception; therefore, when working memory was additionally overloaded, scene recognition was less accurate in haptic-to-haptic than in visual-to-visual task procedure.

While discussing their findings, Newell et al. (2005) suggest that by using less complex shapes than those applied in their experiments, or by increasing the time of learning a scene, it might be possible to neutralize the costs of cross-modal recognition. Notably, during the trials the authors presented scenes consisting of 7 flat animal figures selected randomly from the entire set of 15, and distributed on a round platform. That stage was followed by a 20-second interval, after which the participants were presented with the test platform where the positions of two figures on the scene were swapped. Critical analysis of the procedure adopted by these authors inspires a suggestion that in subsequent studies focusing on visuo-haptic recognition of scenes a few factors should be controlled. Firstly, the participants' verbal statements showed that some of them identified figures as representations of specific animals and named them while others treated the figures as non-semantic shapes. In fact, it is known that sighted individuals more effectively remember shapes which are haptically perceived once these have been identified and named (Pathak & Pring, 1989). Furthermore, it is easier to visualize semantic objects than non-semantic ones (cf. Lacey et al., 2010). Actually, one of the experiments conducted by Newell et al. (2005) included the articulatory suppression variable designed to control the option of naming the figures at the stage of learning the scene. Yet, the suppression task, involving repetition of the word "the," may have been too easy to completely prevent verbalization (this may be confirmed by the lack of main effect related to the suppression factor). Secondly, it is not obvious to what degree the participants made effort to maintain the spatial representation of the scene in their working memory during the 20-second interval between learning and recognition. This may have distorted the results of the study. Indeed, Pensky et al., (2008) showed an interaction between the type of measurement (using data from working memory; requiring a representation to be retrieved from long-term memory) and the type of modality (visual; haptic) at the stage of learning and recognizing spatial objects. Thirdly, it is uncertain to what degree participants may have been affected by the visual context of the room in which the trials were held (according to reports by Newport, Rabb, & Jackson, 2002, "noninformative vision" improves haptic spatial perception). During the haptic procedure the scene was located behind a curtain. It is unclear whether or not the participants in this situation closed their eyes. If so, this may have contributed to their focus on an egocentric reference frame and to decreasing their performance accuracy.

RESEARCH PROBLEMS AND HYPOTHESES

While comparing studies focusing on cross-modal recognition of stimuli, we can ask the following question: Is poorer recognition in cross-modal conditions related to the modality used in the course of learning and to task difficulty, both of which result in higher demand on working memory during the experimental task? In accordance with the assumed hypothesis, the working memory load modulates costs resulting from the change in the modality in visuo-haptic recognition of scenes: it is only with significant burden that the cognitive system bears the cost of modality change.

METHOD

The participants in the experiment were 60 university students (52 females, 8 males) aged 18-27 (*M* = 20.98; *SD* = 1.38), who were randomly assigned to one of four study groups distinguished on the basis of the modality of learning and recognition (visual; haptic). The participants had no visual impairments or corrected-to-normal vision; they had no haptic deficits, either.

The materials used in the experiment included a square LEGO Duplo baseplate, with a side length of 38.5 cm, and 3D figure-pieces of the following animals: horse, cow, sheep, little pig, cat, and chicken. These were selected from a group of 21 figures (including 10 “farm animals”: a horse, a foal, a cat, a hen, a dog, a rabbit, a sheep, a little pig, a cow, a calf) based on the results of pilot studies. Six blindfolded students participated in Pilot Study 1. They were asked to name animals represented by each of the 21 separately explored figures. A response was recognized as correct if the name referred to the basic level or the subordinate level of the concept (e.g., when naming a foal it was enough to say it was a horse). The sixteen figures that were successfully identified by a minimum 50% of the participants during the first pilot study were selected to be used in the remaining pilot studies (notably, the participants had considerable difficulty indentifying some animals – no participant was able to recognize the calf by touching the figure, and only 17% identified the bear, the tiger, and the zebra). Pilot Study 2 was conducted with twenty blindfolded students, none of whom had participated in Pilot Study 1. This time the name of an animal was given prior to tactile exploration of the figures. The participants were asked to recognize (name) the consecutive figures explored by touch. The six pieces that were selected as experimental material had been recognized by the highest percentage of those participating in Pilot Study 2, and they fit in a common category ("farm animals").

Trials performed by each participant individually were presented to them at random order. Each participant solved eight subtests (four per two different levels of task difficulty depending on the number of changing scene components), all of which consisted of two essential parts. During the first part, participants were asked to memorize the layout of the figures representing a horse, a cow, a sheep, a little pig, a cat and a chicken on the board (names of the animals were given to the participants before the experiment). Each scene presented to all participants during the learning stage contained the same layout of figures, previously selected at random and arranged with their fronts towards the viewer. The board was placed on the table at which the participants were seated. Half of the participants performed this trial by viewing the board (for 10 seconds), and the other half by touching the board, with their eyes covered (for 60 seconds – the time of exploration was the same as in the experiment by Newell et al., 2005; additionally, Pilot Study 2 showed that the duration was sufficient for exploring six figure-pieces). Then the board was removed, and there was a 60-second delay interval, during which the participants solved a symbol coding task (analogous to the coding test on Wechsler Intelligence Scale /WAIS-R/). The task was designed to "clean" working memory of the spatial information related to the model board. After the delay interval, the test board was shown and the participants recognized it using the same or different modality. The boards presented for recognition differed from the initial arrangement in such a way that the positions of two or three figures previously selected at random were swapped (four boards per each condition, respectively). The orientation of the figures was retained: the front of the removed animal was replaced with the front of the animal that took its place (see: Figure 1).

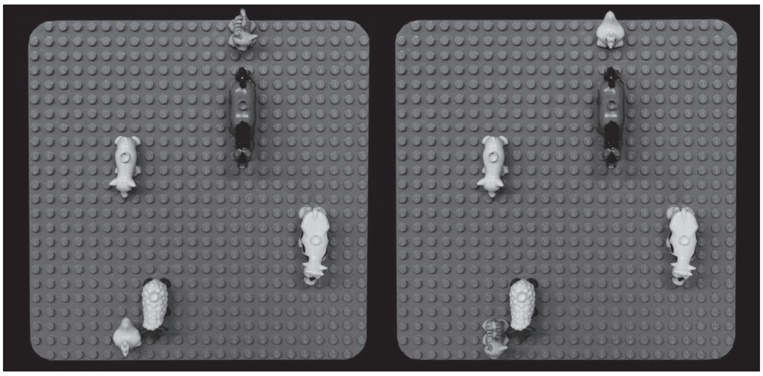

Figure 1. Sample experimental subtest: a change in two elements of the scene (hen and cat).

The participants were supposed to identify the animals whose positions changed on the board in comparison to the learned scene. Measurements included the accuracy and time of response (in seconds). Performance of a single trial was assessed as correct and scored 1 point when all animals that were swapped on the board were identified properly and at the same time no animals with unchanged position were named (i.e., when there was no error of false alarm). Any other responses were assessed as incorrect (0 score). Reaction time was recorded with a stopwatch operated by the experimenter. The stopwatch was switched on the moment the participant either removed the blindfold (visual conditions) or initiated tactile exploration of the board (haptic conditions); it was switched off when the participant informed the experimenter that he/she had finished answering.

RESULTS

The accuracy of scene recognition in four experimental groups was compared using ANOVA with repeated measurement for inter-object factors, i.e., the modality of learning (visual; haptic) and modality change at the stage of recognition (yes; no) as well as intra-object factor – the level of difficulty of the recognition task (change in two or three elements of the scene).

The analysis showed a main effect only for the intra-object factor ($F(1, 56) = 28.40$; $p < .001$; $\eta^2 = .34$) – more accurate responses were given when two figures were swapped ($M = 2.73$; $SD = 0.16$) than when the positions of three elements of the scene were changed ($M = 1.88$; $SD = 0.14$). The findings did not show main effects for the factors of modality of learning ($F(1, 56) = 1.94$;

p = .170) and modality change (F(1, 56) = 1.27; p = .265) or any significant interaction of factors.

In order to compare response time in four experimental groups, an analogous ANOVA was performed for the dependent variable of recognition time (expressed in seconds). Again, a main effect was found for the intra-object factor (F(1, 56) = 28.42; p < .001; η^2 = .34). A greater number of changes in the positions of figures on the board was connected with a longer time needed for identifying the swapped figures (M = 27.34; SD = 1.18) than was the case for a smaller number of changes (M = 22.60; SD = 1.15). No main effects were found for the factors of modality of learning (F(1, 56) = 2.60; p = .112) and modality change (F(1, 56) = .21; p = .650). Interaction was found for the variables of modality of learning and modality change (F(1, 56) = 86.95; p < .001; η^2 = .61), connected with the fact that recognition time was longer for haptic than for visual recognition (see: Figure 2). The findings also showed the interaction of both between-participant variables with the intra-object factor (F(1, 56) = 4.54; p = .037; η^2 = .08). After running simple effects it was found that a change in modality significantly impacted the time of scene recognition only when the scene was being recognized by viewing, and when three elements changed their position on the board (t(28) = -2.10; p = .045; Es = -2.01) – recognition time was longer when the test modality was different from the modality of learning (haptic-visual) than when it was unchanged (visual-visual) (see: Figure 2). Reaction time did not differ significantly when three elements were changed for the visual-haptic and haptic-haptic conditions (t(28) = -0.24; p = .816) or with the change of two elements in the following conditions: haptic-visual and visual-visual (t(28) = -1.83; p = .078) as well as visual-haptic and haptic-haptic (t(28) = -1.01; p = .320). No other interactions were significant.

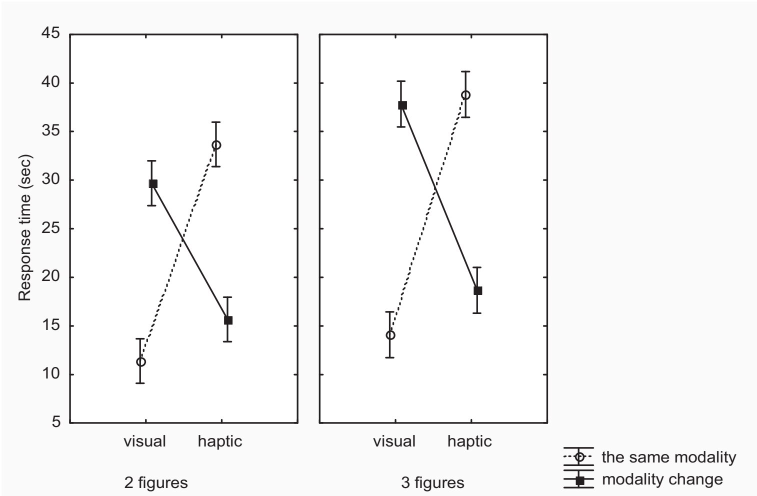

Figure 2. Reaction time depending on the modality of learning (visual; haptic), modality change at the stage of recognition (the same modality; modality change), and the level of difficulty of the recognition task (change in two or three figures of the scene). Error bars represent ± 1.0 standard error of mean.

In order to verify the research hypothesis, independent groups *t*-tests were conducted comparing the accuracy of scene recognition during cross-modal and intra-modal recognition procedures for the data distinguished by the modality of learning and the level of difficulty in the recognition task. A change in modality resulted in a decreased accuracy of scene recognition only when the participants learned the scene by touch and when three objects were swapped on the board (comparison of trial conditions: haptic-haptic and haptic-visual: $t(28) = -2.25$; $p = .032$; Es = 0.09). In the remaining cases, i.e. for the comparison of visual--visual and visual-haptic conditions with two or three changed objects ($t(28) = 0.33$; $p = .747$; $t(28) = 0.00$; $p = 1.00$, respectively), as well as haptic-haptic and haptic-visual conditions with two swapped figures ($t(28) = 0.83$; $p = .411$), the change in modality did not impact the accuracy of scene recogition (see: Figure 3).

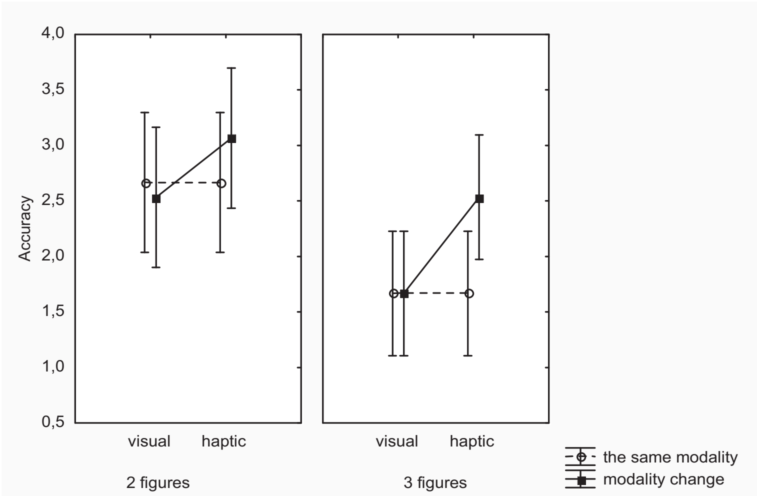

Figure 3. Accuracy in scene recognition depending on the modality of learning (visual; haptic), modality change at the stage of recognition (the same modality; modality change), and the level of difficulty of the recognition task (change in two or three figures of the scene). Error bars represent ± 1.0 standard error of mean.

DISCUSSION

It can be concluded that the research hypothesis was verified positively. It was demonstrated that working memory burden affects cross-modal spatial performance. The findings of the study show that a change in the modality of recognition caused both a decrease in the accuracy of responses and an increase in the time of taking decision only in the conditions of the greatest burden imposed on working memory which was connected with both tactile learning of the scene and a higher level of difficulty in the recognition task.

Therefore, these findings confirm the suggestion advanced by Newell et al. (2005) that the cost of change in modality is borne by the cognitive system only in the case of more difficult tasks (cf. Ernst et al., 2007; Newell et al., 2001). In our experiment a decrease in recognition accuracy related to modality change was found in those cases where during the learning stage the scene was explored by touch and where three pieces swapped position during the test. The procedure, which involved learning the scene by touch (with eyes covered), may have hindered the participants from using an allocentric reference frame (Newport, Rabb, & Jackson, 2002). Therefore, it may be hypothesized that, when learning a tactile scene, they were trying to visualize it (Szubielska, 2009), and at the same time they were partly encoding it in a form of egocentric representation. While identifying an altered scene by touch, they were able to use the same type of representation. On the other hand, while viewing a changed scene they had to compare the mental representation retrieved from long-term memory, possibly egocentric to some extent, with the image available to their perception. In the case of less complex tasks (change of two elements on the board) the discrepancy between the nature of mental and perceptual representations did not matter; on the other hand, it did hinder the recognition of elements that changed in a scene during a more complex task (change of three figures). Arguably, the difference in the nature of scene representations created in the course of haptic vs. visual learning, which constitutes an obstacle in comparing spatial information in more complex tasks, is also revealed by the longer time needed for the visual recognition of a scene with three swapped figures if the participant learned the scene by touch in comparison to the situation when visual modality was used for learning (a similar effect in tests investigating delayed recognition of isolated common objects was obtained by Easton, Greene, and Srinivas, 1997; but these authors also showed the effect of modality change on visual memorization and tactile recognition). By way of comparison, when exploring the scene visually, the participants created its mental image. While identifying the new scene, they could retrieve from memory the mental image in which, using an allocentric reference frame, they stored information on spatial relations between specific objects. Therefore, it was of little consequence to them whether they were comparing these relations to what they were viewing or to what they were touching when analyzing the changed scene.

There was an interactive impact of learning modality and modality change on the time required for solving a task. If the scene had been learned visually, more time was needed for recognition due to the change in modality, whereas in the case of tactile learning the change in modality resulted in shorter time required to provide an answer. The observed interaction resulted from the fact that, generally, the time of haptic recognition of a scene was longer than the time required for visual recognition, which intuitively seems obvious (it takes more time to identify an item which your hand encounters inside a pocket that to assess the same object once you retrieve it from the pocket) and is consistent with other studies (e.g., Reales & Ballesteros, 1999).

The adopted method of assessing the accuracy of response constituted a limitation of the experiment. It became evident during the tests that the participants made mistakes by omitting elements with changed location in the scene and pointing to pieces with unchanged position (unfortunately, this information was not recorded in answer sheets). In order to precisely determine the performance in the task, it would have been necessary to take into account the number of both accurate and erroneous responses (this assessment method was employed, for example, by: Easton, Greene, Srinivas, Easton, Srinivas, & Greene, 1997; Reales & Ballesteros, 1999). Another procedure adopted in our experiment that should be improved in future studies is the method of measuring response time. Given the varied number of elements to be indicated in various experimental conditions, rather than measure the time until participants decide they have named all elements changed in the scene, the time measurement should be conducted until the moment participants start providing answers.

REFERENCES

Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. *Nature Neuroscience, 4*(3), 324-330.

Casey, S. J., & Newell, F. N. (2007). Are representations of faces independent of encoding modality? *Neuropsychologia, 45*(3), 506-513.

Easton, R. D., Greene, A. J., & Srinivas, K. (1997). Transfer between vision and haptics: Memory for 2-D patterns and 3-D objects. *Psychonomic Bulletin & Review, 4*(3), 403-410.

Easton, R. D., Srinivas, K., & Greene, A. J. (1997). Do vision and haptics share common representations? Implicit and explicit memory within and between modalities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(1), 153-163.

Ernst, M. O., Lange, C., & Newell, F. N. (2007). Multisensory recognition of actively explored objects. *Canadian Journal of Experimental Psychology, 61*(3), 242-253.

Lacey, S., Flueckiger, P., Stilla, R., Lava, M., & Sathian, K. (2010). Object familiarity modulates the relationship between visual object imagery and haptic shape perception. *NeuroImage, 49*(3), 1977-1990.

Loomis, J. M., Klatzky, R. L., McHugh, B., & Giudice, N. A. (2012). Spatial working memory for locations specified by vision and audition: Testing the amodality hypothesis. *Attention, Perception, & Psychophysics, 74*(6), 1260-1267.

Newell, F. N., Ernst, M. O., Tjan B. S., & Bülthoff, H. H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological Science, 12*(1), 37-42.

Newell, F. N., Woods, A. T., Mernagh, M., & Bülthoff, H. H. (2005). Visual, haptic and crossmodal recognition of scenes. *Experimental Brain Research, 161*, 233-242.

Newport, R., Rabb, B., & Jackson, S. R. (2002). Noninformative vision improves haptic spatial perception. *Current Biology, 12*(19), 1661-1664.

Pasqualotto, A., & Proulx, M. J. (2012). The role of visual experience for the neural basis of spatial cognition. *Neuroscience and Biobehavioral Reviews, 36*(4), 1179-1187.

Pathak, K., & Pring, L. (1989). Tactual picture recognition in congenitally blind and sighted children. *Applied Cognitive Psychology, 3*(4), 337-350.

Pensky, A. E. C., Johnson, K. A., Haag, S., & Homa, D. (2008). Delayed memory for visual-haptic exploration of familiar objects. *Psychonomic Bulletin & Review, 15*(3), 574-580.

Reales, J. M., & Ballesteros, S. (1999). Implicit and explicit memory for visual and haptic objects: Cross-modal priming depends on structural descriptions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(3), 644-663.

Szubielska, M. (2009). *The role of visual experience on the strategies of blind and sighted individuals in mental imagery*. Poster presented on XI European Congress of Psychology. Oslo, Norway, July 7-10, 2009.

Zuidhoek, S., Kappers, A. M. L., van der Lubbe, R. H. J., & Postma, A. (2003). Delay improves performance on a haptic spatial matching task. *Experimental Brain Research, 149*(3), 320-330.