20/04/2024 06:14

(Article begins on next page)

The Multilevel Modality-Switch Effect: What Happens When We See the Bees

Buzzing and Hear the Diamonds Glistening

Elisa Scerrati, Luisa Lugli, Roberto Nicoletti, Anna Maria Borghi

University of Bologna

Author Note

Elisa Scerrati, Department of Philosophy and Communication, University of Bologna,

Italy; Luisa Lugli, Department of Philosophy and Communication, University of Bologna,

Italy; Roberto Nicoletti, Department of Philosophy and Communication, University of

Bologna, Italy; Anna Maria Borghi, Department of Psychology, University of Bologna, Italy,

Institute of Cognitive Sciences and Technologies, Italian National Research Council, Rome,

Italy.

Correspondence concerning this article should be addressed to Elisa Scerrati,

Dipartimento di Filosofia e Comunicazione, Università di Bologna, Via A. Gardino, 23,

40122 Bologna, Italia Tel.: 051-2092218-19 E-mail: elisa.scerrati@unibo.it

Word count: 3898

Abstract

Previous studies demonstrated that the sequential verification of different sensory modality properties for concepts (e.g., *BLENDER-loud*; *BANANA-yellow*) brings about a processing cost, known as the modality-switch effect. We report an experiment designed to assess the influence of the mode of presentation (i.e., visual, aural) of stimuli on the modality-switch effect in a property verification and lexical decision priming paradigm. Participants were required to perform a property verification or a lexical decision task on a target sentence (e.g., "a *BEE buzzes*", "a *DIAMOND glistens*") presented either visually or aurally after having been presented with a prime sentence (e.g., "the *LIGHT is flickering*", "the SOUND *is echoing*") that could either share both, one or none of the target's mode of presentation and content modality. Results show that the mode of presentation of stimuli affects the conceptual modality switch effect. Furthermore, the depth of processing required by the task modulates the complex interplay of perceptual and semantic information. We conclude that the MSE is a task-related, multilevel effect which can occur on two different levels of information processing (i.e., perceptual and semantic).

The Multilevel Modality-Switch Effect: What Happens When We See the Bees

Buzzing and Hear the Diamonds Glistening

In people's everyday life, the majority of experiences involve multiple sensory modalities. We are thus required to be able to switch across different sensory modalities in different situations. A classic example involves the musicians in an orchestra: they are required to be able to quickly process visually presented auditory contents (i.e., sheet music along with the conductor's gestures) in order to perform. However, this only happens through years of studying. Indeed, recent research has shown that people experience a cognitive cost in shifting attention between different sensory modalities. Interestingly, such cognitive cost occurs both when switching between events presented in different modalities (Spence, Nicholls & Driver, 2001) as well as when switching between sentences having different-modality contents (Pecher, Zeelenberg & Barsalou, 2003). For example, switching from the sentence "*BLENDER is loud*" to the sentence "*BANANA is yellow*" incurs a processing cost much like switching from an auditory tone to a light flash. This phenomenon is known as Modality-Shifting or Modality-Switch Effect (hereafter MSE).

The MSE with language has been extensively explored with both behavioral (Marques, 2006; Pecher, Zeelenberg & Barsalou, 2004; Scerrati, Baroni, Borghi, Galatolo, Lugli & Nicoletti, 2015; van Dantzig, Pecher, Zeelenberg & Barsalou, 2008; see

also Vermeulen, Niedenthal, Luminet, 2007 for a similar result with emotional concepts) and ERP studies (Collins, Pecher, Zeelenberg, Coulson, 2011; Hald, Marshall, Janssen, Garnham, 2011; Hald, Hocking, Vernon, Marshall & Garnham, 2013). Whether the finding of a purely perceptual phenomenon during conceptual processing is just the epiphenomenal result of spreading activation or evidence supporting the assumption that perceptual information is engaged in conceptual processing is debated. On the one hand, it has been argued that the conceptual system is separated from sensory information (disembodied cognition hypothesis, see Mahon & Caramazza, 2008; Mahon & Hickok, 2016 for discussions). On this account, the MSE would reflect the way in which activation spreads throughout the system, therefore it would not reveal anything about semantic processing. On the other hand, it has been assumed that the perceptual and conceptual systems are tightly interwoven and share the same processing mechanisms. Proponents of the grounded accounts of knowledge (Barsalou, 2008; for recent reviews see Borghi and Caruana, 2015; Pecher, 2013) assume that knowledge representation and processing is achieved by reactivating aspects of experience. In particular, conceptual processing would imply constructing a sensorimotor simulation of the objects or events concepts refer to. Such a simulation would involve the partial reactivation of those brain areas that were also active during our interaction with the concepts'

referents. For example, on processing the concept DOG, brain areas that represent visual, auditory, tactile, olfactory, gustatory, affective, and motor information about dogs would be liable to partial reactivation. Importantly, simulations are sketchy records of experience that can be flexibly adapted to the context and task at hand (Barsalou, 1999; Gallese, 2009).

Recently, Scerrati et al. (2015) obtained evidence that sensorimotor simulations can also be triggered by a perceptual, linguistically described stimulus presented in a sensory modality different from vision (i.e., the auditory modality). Participants were presented with a prime sentence describing a light or a sound's perceptual property (e.g., "The light is flickering", "The sound is echoing") then they were required to perform a property-verification task on a target sentence with a vision-related or a hearing-related content (e.g., "Butter is yellowish", "Leaves rustle"). The sensory modality activated by the content of the prime sentence could be compatible with the target's content modality (e.g., vision-vision: "The light is flickering" followed by "Butter is yellowish") or not (e.g., vision-audition: "The light is flickering" followed by "Leaves rustle"). Crucially, the stimuli's presentation modality was manipulated such that half of the participants was faced with written prime and target sentences while the other half was faced with spoken prime and target sentences. Results showed that participants were faster at judging whether a certain property was true of a given concept when the

target's content modality corresponded to the one pre-activated by the content of the prime sentence with both visual and aural presentation of stimuli.

In the present study we were interested in examining whether switching between different mode of presentation (i.e., visual, aural) across prime and target sentences conveying a sensory content brings about a modality switching cost. Specifically, we aimed at understanding whether and how the conceptual MSE is modulated by the mode of presentation of stimuli. To our knowledge, no previous study has explored this issue in regard to the MSE. Interestingly, however, different studies found that sentence processing can be affected by mode of presentation. Kaschak, Zwaan, Aveyard and Yaxley (2006, Experiment 2) showed that participants were faster in making sensibility judgements on target sentences when the direction of motion implied by the sentence with a hearing-related content (e.g., "The commuter had just arrived on the platform when the subway roared into the station") and the direction of motion depicted by a concurrent auditory stimulus were the same, provided that both the sentence and the stimulus were aurally presented. In a different yet related study, Vermeulen, Corneille and Niedenthal (2008) showed that asking people to store three visual or auditory items (i.e., pictures or sounds) in short-term memory for a subsequent memory task resulted in a worse performance in an intervening property verification task when the latter concerned

sentences involving properties in the same modality as that of the stored items (interference hypothesis). Vermeulen et al. (2008) suggested that the general attentional load imposed upon participants together with the high complexity of the dual-task paradigm used in their study moderated switching costs. On the basis of this previous evidence, we expect that the mode of presentation of sentences might be relevant in modulating the MSE. Specifically, given that neither we manipulate attentional load, nor we use a dual task paradigm, we expect to observe a facilitation when the prime and the target share the same presentation and content modality as in prior studies where switching costs were found.

Whether and how the conceptual MSE is affected by the mode of presentation of stimuli may hinge on task demands. Connell & Lynott (2014) found that task-specific implicit perceptual attention preactivates modality-specific systems leading to facilitated representation of semantic information related to those modalities. That is, preactivating the visual system through the presentation of strongly visual words (e.g., "cloudy") facilitated performance in the lexical decision task, whereas preactivating the auditory system through the presentation of strongly auditory words (e.g., "noisy") facilitated performance when the task was reading aloud. In the present research, we used two different tasks: the property verification and the lexical decision task (LDT, McNamara, 1992). We believe that the mode of

presentation of stimuli might differently impact the conceptual MSE on the basis of the depth of processing required by the task. With the property verification task we predict to observe a better performance when the presentation and the content modalities of target sentences are congruent (e.g., "Butter is yellowish" presented visually) compared to when they are incongruent (e.g., "Butter is yellowish" presented aurally) due to the depth of processing required by the task. With a less conceptually engaging task such as the LDT, instead, we expect that the mode of presentation might feature prominently compared to the content modality of sentences.

Method

Participants

One hundred twenty-eight students from the University of Bologna (79 females; mean age: 21.45, SD: 2.37) participated in the experiment in exchange for course credit. Sixty-five participants were randomly assigned to the property verification task condition whereas 63 participants were randomly assigned to the LDT condition. All participants were Italian native speakers, had normal or corrected-to-normal vision and hearing by self-report, and were naïve as to the purpose of the experiment. The experiment was approved by the Psychology Department's ethical committee of the University of Bologna. Written informed consent was obtained from all individual

participants included in the study. Minors did not take part in the study.

Materials

Twenty-four prime sentences and forty-eight target sentences were used in this experiment. Stimuli were the same as in Scerrati et al. (2015). Half of the prime sentences had a vision-related content (e.g., "the *LIGHT is flickering*") whereas the other half had a hearing-related content (e.g., "the *SOUND is echoing*"). Properties in the visual and auditory prime sentences were taken from the norming study by Lynott and Connell (2009) and from a rating of 50 Italian adjectives (see Appendix A in Scerrati et al., 2015). Each of the 24 prime sentences was repeated four times throughout the experiment, twice they were aurally presented over closed-ear headphones and twice they were visually presented on the screen.

Target sentences were taken from the van Dantzig et al.'s study (2008). Twenty-four had a vision-related content (e.g., "a *WALNUT is brown*") and twenty-four had a hearing-related content (e.g., "a *BEE buzzes*"). In these critical pairs the property was always true of the concept. Each pair was used only once. Two properties were repeated once across the pairs, although paired with different concepts (i.e., "a *BEE buzzes*", "a *FLY buzzes*"; "*BROCCOLI is green*", "*SPINACH is green*"). For an overview of the visual and auditory prime and target

sentences see Appendix B in Scerrati et al. (2015). Prime and target sentences were the same across tasks.

As for the property verification task, an additional set of 48 filler sentences, always taken from van Dantzig et al. (2008) was used. In the filler sentences the property was always false of the concept. Twelve filler sentences had a false visual property (e.g., "the *WATER is opaque*"), 12 had a false auditory property (e.g., "the *COMB sings*"), whereas the remaining 24 filler sentences had a false property that did not belong to any modality (e.g., "the *BED is sleepy*"). This latter type of fillers was used in order to avoid participants from basing their answers on a superficial word-association strategy, rather than on deeper conceptual-processing (see Solomon & Barsalou, 2004).

As for the LDT, an additional set of 48 filler sentences featuring a non-word was used. In half of the filler sentences the non-word was the concept word, whereas in the other half the non-word was the property word. Non-words were generated altering two of the consonants or the double consonant keeping unchanged the vowels so as to preserve the phonotactic rules of Italian.

Each participant was presented with 96 prime sentences followed by 96 target sentences (48 critical and 48 filler) throughout the experimental session. Prime and target sentences were randomly combined to form four modality conditions, that

is: different-different (DD, when both the presentation and the
content modalities do switch from prime to target sentence),
different-same (DS, when the presentation modality does switch
but the content modality does not), same-different (SD, when
the content modality does switch but the presentation modality
does not) and same-same (SS, when the prime and the target
sentences share the same presentation and content modalities).
For example, a visually presented prime sentences with a
vision-related content (e.g., "the *LIGHT is flickering*") could be
combined with: 1) an aurally presented target sentence with a
hearing-related content (e.g., "a *BEE buzzes*", DD); 2) an
aurally presented target sentence with a vision-related content
(e.g., "a *WALNUT is brown*", DS); 3) a visually presented prime
sentences with a hearing-related content (e.g., "a *BEE buzzes*",
SD); 4) a visually presented prime sentences with a vision-
related content ("a *WALNUT is brown*", SS). Each target
sentence appeared in all modality conditions, counterbalanced
across participants.

Procedure

   The stimuli were presented on a 17 inch monitor (1.6 Ghz
refresh rate). The participants sat at a viewing distance of about
60 cm from the monitor in a dimly-lit room. They were invited
to wear a pair of headband headphones before starting the
experiment. Each trial started with the presentation of a fixation
cross (0.5 cm x 0.5 cm) for 500 milliseconds (ms). Immediately

after the fixation the prime sentence appeared on the screen or

was delivered through headphones for 2000 ms. Then, the target

sentence was displayed on the screen or delivered through

headphones until a response was given or until 4000 ms had

elapsed. Visually presented prime and target sentences ranged

from 5.9 cm to 17.3 cm (from 9 to 29 characters) which resulted

in a visual angle range between 5.6° and 16.5°. All sentences

were bold lowercase Courier new 18 and were presented in

black in the center of a white background. Participants were

instructed to read or to listen to the prime and target sentences

and then judge, as quickly and accurately as possible, whether

in each target sentence the property was true of the concept

(property verification task condition), or whether in each target

sentence there was a non-word or not (LDT condition). In both

task conditions, half of the participants pressed the "s" key of a

"qwerty" keyboard when either the property was true of the

concept or there was a non-word in the target sentence and the

"k" key when either the property was false of the concept or the

target sentence did not contain a non-word. The other half of the

participants was assigned to the reverse mapping.

The order of presentation of each prime-target sentence was

completely randomized across participants. Participants

underwent a short practice session of 32 stimuli (different from

those used in the experimental blocks) before starting the

experiment. The experiment consisted of one block of 96 prime-target pairs and lasted approximately 15 minutes.

Results

In the property verification task condition, five participants (all females) were excluded from the analysis: Four of these participants failed to reach an accuracy score of 65% while the other participant responded 35% of the trials in less than 300 ms, indicating that she may have misconceived the task and tried to respond on the prime sentence also. Sixty participants therefore remained for further analysis. Responses to filler sentences were discarded. Omissions (5.93%), Incorrect responses (21.42%) and RTs faster/slower than the overall participant mean minus/plus 2 standard deviations (2.19%) were excluded from the analyses. In the LDT condition, three participants (two females) failed to reach an accuracy score of 65%. Their data were removed, leaving sixty participants for further analysis. Responses to filler sentences were discarded. Omissions (5.03%), Incorrect responses (7.04%) and RTs faster/slower than the overall participant mean minus/plus 2 standard deviations (2.60%) were excluded from the analyses.

Mean Response Times (RTs) of the correct responses were submitted to a Repeated Analysis of Variance (ANOVA) with *Mode of Presentation* (different vs. same), *Content Modality* (different vs. same) and *Target Congruency* (incongruent vs.

congruent) as the within-subject factors for the two *tasks* (property verification vs. lexical decision) separately. Data are shown in Table 1.

In the property verification task condition there was a main effect of *Mode of Presentation*, F(1, 59) = 4.582, MS_e = 75789.90, p < .05, η_p^2 = .072, that is, decision latencies were faster when the *Mode of Presentation* was the same across prime and target sentences rather than different (M: 2036 ms vs. 2090 ms). The analysis also revealed a main effect of *Target Congruency* (F(1, 59) = 18.633, MSE = 65906.25, p < .001, η_p^2 = .240), that is, decision latencies were faster when the *Mode of Presentation* and the *Content Modality* of the target were congruent rather than incongruent (M: 2013 ms vs. 2114 ms). No other main effect or interaction turned out to be significant, F_s < 2.66, p_s > .108.

In the LDT condition, there was a main effect of *Mode of Presentation*, F(1, 59) = 6.544, MS_e = 59889.45, p < .05, η_p^2 = .1, that is, decision latencies were faster when the *Mode of Presentation* was the same across prime and target sentences rather than different (M: 1942 ms vs. 1999 ms). No other main effect or interaction turned out to be significant, F_s < 1.420, p_s > .238.

Mean of the incorrect responses were submitted to an ANOVA with the same factors as those of the RTs analysis. In the property verification task condition no main effect or

interaction turned out to be significant, $F_s < 2.247$, $p_s > .139$. In the LDT condition there was a significant interaction between *Mode of Presentation* and *Target Congruency*, $F(1,59) = 4.484$, $MS_e = 140.56$, $p < .05$, $\eta_p^2 = .071$. Paired-sample *t*-tests showed that percentage of ERs was higher when the Mode of Presentation was the same across prime and target sentences and the target was congruent compared to different and incongruent (9.5% vs. 6.8%), $t(59) = -2.075$, $p < .05$ same and incongruent (9.5% vs. 5.8%), $t(59) = -2.225$, $p < .05$ and different and congruent (9.5% vs. 5.9%), $t(59) = -2.327$, $p < .05$. No other main effect or interaction turned out to be significant, $F_s < 2.683$, $p_s > .107$.

General Discussion

The present research investigated whether and to what extent switching between different modes of presentation (i.e., visual, aural) across prime and target sentences affects the conceptual MSE. Although previous studies investigated how sentence processing can be affected by mode of presentation of linguistic stimuli, such relationship had not previously been studied in the context of the MSE. Given that the impact of the mode of presentation of stimuli on language processing may be modulated by task demands (see Connell and Lynott, 2014 for a similar result in a different context), we compared performance on a property verification priming paradigm with performance

on a lexical decision priming paradigm, each involving different

levels of conceptual processing.

In keeping with our hypothesis, we found evidence for the

involvement of the mode of presentation of stimuli in both the

property verification and the lexical decision task. Crucially,

results from both tasks showed that the presentation-driven

effect weakens the conceptual MSE. Indeed, a conceptual MSE

was observed in the property verification task, but not in LDT,

as expected; however, it did not reach significance[1].

Interestingly, the property verification task highlighted an

effect of the target congruency. That is, we found that

participants were slower in deciding whether a certain property

was true of the concept when the presentation and the content

modality were incongruent for the target (e.g., "a *BEE buzzes*"

presented visually) compared to when they were congruent.

Such a within-target MSE is in line with the results of van

Dantzig et al. (2008) showing that when a perceptual stimulus

(i.e., a light flash, a tone or a vibration) and a subsequent target

sentence were in a different sensory modality decision latencies

were slower compared to when they were in the same modality.

Our results broaden their finding showing such an effect within

the same stimulus, that is, when the processing of perceptual

and conceptual information overlap in time. It is worth noting

---

[1] The conceptual MSE in the property verification task was 34 ms, p = .168, whereas in the LDT it was 4 ms, p = .838.

that such interference only occurred with the property verification task. Therefore, it seems likely that since the lexical decision task did not emphasize conceptual processing, it only recruited the semantic system to a certain extent insufficient to generate interference between the two systems.

In sum, our findings show that conceptual processing is not only affected by switching between sensory modalities on a semantic level (i.e., content modality of stimuli) but also by switching between sensory modalities on a purely perceptual level (i.e., mode of presentation of stimuli). Interestingly, our results also demonstrate a task-dependent, complex interplay of perceptual and semantic information taking place within the target. These findings question the view according to which the MSE does not reveal anything about semantic processing as claimed by the critic of the grounded accounts of knowledge (Mahon & Caramazza, 2008).

We conclude that the MSE is a task-related, multilevel effect which can occur on two different levels of information processing, i.e., perceptual and semantic. We interpret these results as further evidence supporting the likelihood that the perceptual and conceptual systems are tightly interwoven and share the same processing mechanisms as claimed by the simulation account of conceptual processing (Barsalou, 1999, 2003, 2008).

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577–660.

Barsalou, L. W. (2003). Situated simulation in the human conceptual system. *Language and Cognitive Processes*, 18:5-6, 513-562.

Barsalou, L. W. (2008). Grounded cognition. *The Annual Review of Psychology*, 59, 617–645.

Borghi, A. M., & Caruana, F. (2015). "Embodiment theory", in *International Encyclopedia of the Social & Behavioral Sciences*, 2nd edition, ed. J. D. Wright (Oxford: Elsevier), Vol 7.

Collins, J., Pecher, D., Zeelenberg, R., Coulson, S. (2011). Modality switching in a property verification task: an ERP study of what happens when candles flicker after high heels click. *Front. Psychology* 2:10. doi: 10.3389/fpsyg.2011.00010

Connell, L., & Lynott, D. (2014). I see/hear what you mean: semantic activation in visual word recognition depends on perceptual attention. *Journal of Experimental Psychology: General, 143*, 527-533.Gallese, V. (2009). Motor abstraction: a neuroscientific account of how action goals and intentions are mapped and understood. *Psychological Research*, 73, 486–498.

Hald L.A., Marshall J-A, Janssen D.P., Garnham, A. (2011). Switching modalities in a sentence verification task: ERP evidence for embodied language processing. *Front. Psychology* 2:45. doi: 10.3389/fpsyg.2011.00045

Hald L.A., Hocking I., Vernon D., Marshall J-A, and Garnham, A. (2013). Exploring modality switching effects in negated sentences: further evidence for grounded representations. *Front. Psychol*. 4:93. doi: 10.3389/fpsyg.2013.00093

Kaschak, M.P., Zwaan, R.A., Aveyard, M., Yaxley, R.H. (2006). Perception of auditory motion affects language processing. *Cognitive Science*, 30, 733-744.

Lynott, D. & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, 41 (2), 558-564.

Mahon B. Z., & Hickok, G. (in press). Arguments about the nature of concepts: Symbols, embodiment, and beyond. *Psychonomic Bullettin & Review*.

Mahon B. Z., & Caramazza, A. (2008). A critical look at the embodied cognition hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology*, 102, 59-70.

Marques, J. F. (2006). Specialization and semantic organization: Evidence for multiple semantics linked to sensory modalities. *Memory & Cognition*, 34 (1), 60-67.

McNamara, T. P. (1992). Theories of priming: I. Associative distance and lag. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 18, 1173-1190.

Pecher, D. (2013). The perceptual representation of mental categories. In D. Reisberg (Ed.). *The Oxford Handbook of Cognitive Psychology*. New York: Oxford University Press.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2003). Verifying the properties of object concepts across different modalities produces switching costs. *Psychological Science*, 14, 119-124.

Pecher, D., Zeelenberg, R., & Barsalou, L. W. (2004). Sensorimotor simulations underlie conceptual representations: Modality-specific effects of prior activation. *Psychonomic Bulletin and Review*, 11, 164-167.

Scerrati, E., Baroni, G., Borghi, A.M., Galatolo, R., Lugli, L., Nicoletti, R. (2015). The modality-switch effect: Visually and aurally presented prime sentences activate our senses. *Front. Psychol*. 6:1668. doi: 10.3389/fpsyg.2015.01668

Solomon, K. O., Barsalou, L. W. (2004). Perceptual simulation in property verification. *Memory & Cognition*, 32 (2), 244-259.

Spence, C., Nicholls, M. E. R., & Driver, J. (2001). The cost of expecting events in the wrong

      sensory modality. *Perception & Psychophysics*, 63, 330-336.

van Dantzig, S., Pecher, D., Zeelenberg, R., Barsalou, L. W. (2008). Perceptual processing

      affects conceptual processing. *Cognitive Science*, 32, 579-590.

Vermeulen, N., Niedenthal, P. M., & Luminet, O. (2007). Switching between sensory and

      affective systems incurs processing costs. *Cognitive Science*, 31, 183-192.

Vermeulen, N., Corneille, O., Niedenthal, P. M., (2008). Sensory load incurs conceptual

      processing costs. *Cognition*, 109, 287-294.

Table 1

Mean Response Times (in Milliseconds) and Percentage of Errors with Standard Deviations in parenthesis as a Function of *Mode of Presentation* (MoP: Different, Same), *Content Modality* (CM: Different, Same) and *Target Congruency* (TC: incongruent, congruent) for each tasks separately.

	Property verification				Lexical decision			
	RT		ERs		RT		ERs	
	D	**S**	**D**	**S**	**D**	**S**	**D**	**S**
MoP	2090 (338.6)	2036 (340.2)	21 (17)	21 (15.9)	1999 (337.9)	1942 (268.8)	6 (10.4)	7 (10.9)
CM	2080 (357.2)	2046 (321.6)	20 (16.4)	22 (16.5)	1972 (309.7)	1968 (297)	7 (10.9)	6 (10.4)
	I	**C**	**I**	**C**	**I**	**C**	**I**	**C**
TC	2114 (340.5)	2013 (338.3)	22 (16.2)	20 (16.7)	1972 (304.7)	1968 (302)	6 (10)	7 (11.3)