Kai Hamburger & Florian Röser

The Meaning of Gestalt for Human Wayfinding –How Much Does it Cost to Switch Modalities?

1. Introduction

About 120 years ago the Austrian philosopher Christian von Ehrenfels introduced the principle of *Übersummativität* (superadditivity) as one characteristic for a *Gestalt* (von Ehrenfels 1890). The basic premise is that the entirety of an object has properties that are different from those of its parts, thus an object which is structurally composed of elementary features cannot be defined only by some isolated features (this is especially true for landmarks). Over the intervening years this perceptual property has been demonstrated repeatedly (e.g., Wertheimer 1912, phi motion; Koffka 1935, simultaneous color contrast; Steinman, Pizlo, et al. 2000; Sharps & Wertheimer 2000). The question of how our perceptual system creates *Übersummativität* has become increasingly relevant today since "[vision science is…] witnessing a physiology that aims to understand these principles [Gestalt factors] in terms of interactions within neural networks" (Spillmann & Ehrenstein 2004, 1573). Such a *Gestalt*-Neuroscience will have to address fundamental questions concerning the nature of the stimulus information that can be used by *Gestalt* organizing principles (Tse 2004). All this is true for a perceptual level, where vision scientists have repeatedly shown how we are tricked by visual illusions (e.g., Robinson 1972; Eagleman 2001; Hamburger 2007). But, what about *Gestalt* and its meaning on higher, more cognitive levels? How does our perception contribute to cognitive processes, namely spatial memory and spatial orientation? In order to provide insights into these questions, we are here concerned with *Übersummativität* and spatial information, namely landmarks. In our work we not only want to focus on the meaning of landmarks (objects) for human wayfinding but also the different modalities this information may be processed in.

We are all quite familiar with the situation of travelling through a new, unknown or at least unfamiliar environment. Which kind of information prevents us from getting lost? It is generally accepted that we use so-called landmark information, objects that pop out from their environment and therefore become salient. Classical definitions of landmarks and landmark salience assume that an object must have a high visual contrast to the immediate surround in order to be easily distinguishable and therefore to possess a high salience (e.g., Lynch 1960; Presson &

GESTALT THEORY
© 2011(ISSN 0170-057 X)
Vol. 33, No.3/4, 363-382

Montello 1988). But, is it really/solely visual information that we rely on for successful wayfinding? Furthermore, are single features/characteristics of an object responsible for being a useful landmark or is it rather the whole *Gestalt*? Then, in terms of figure-ground segregation we have to deal with the question of how a landmark pops out from its immediate surround and how different modalities are interconnected to provide such an extraction of landmarks. From research with visually impaired or blind people it is known that humans *can* also make use of acoustic or haptic information for wayfinding (e.g., Loomis, Golledge, et al. 1998; Habel, Kerzel, et al. 2010). Studies with unimpaired participants are rare and often focus on a single domain (mainly vision). In a previous study we were able to show that spatial orientation with landmarks in other modalities than vision is possible and sometimes even more efficient (Röser, Hamburger, et al. 2011). But, how much does it then cost (time, error) to switch between modalities if needed? For example, you are provided with a verbal description of a route containing a few landmarks. When you travel the route you will transform the verbal information into visual information to be able to recognize a certain building as a relevant (or irrelevant) part of the route description. In this case you have to switch from verbal to visual information (further details are provided in the methods section). An extensive body of literature on task-switching is available (different tasks within a single modality; e.g., Abruthnott & Woodward 2002), but modality switching is rather new in the field of landmark research. Here, we provide first insights into possible costs of modality switching for landmark information and furthermore challenge some of the classical findings on landmark information processing. In summary, we expect that landmark information can successfully be processed in different modalities (visual, verbal, acoustic), with an advantage for visual information. Therefore, a landmark as a whole is present in more than a single perceptual modality (but maybe at different degrees of abstraction). Other modalities and cognitive processes (e.g., memory) cause a landmark to become a useful *Gestalt* for wayfinding. Furthermore, switching between modalities should be accompanied by decreased recognition/wayfinding performance and increased decision times compared to trials in which the modality remains the same, since the *Gestalt* needs to be established and made available in other modalities.

In wayfinding research different forms of landmark salience have been differentiated and established in computational models. The different saliences are *visual salience*, *semantic salience*, and *structural salience* (for an overview see Caduff & Timpf 2008; Hamburger & Knauff in press). David Caduff and Sabine Timpf (2008) recently introduced the term *cognitive salience* in order to emphasize that the information processing system (the human brain) also contributes to the meaning and importance of a landmark and not just the object features alone.

In previous studies we were able to show that the visual salience is overemphasized, since landmarks may also be successfully processed in other modalities (e.g., acoustic) (Hamburger, Röser, et al. in preparation; Röser et al. 2011). Therefore, we redefined the visual salience as a *perceptually based (contrast) salience*. This should be kept in mind for the remainder of this manuscript.

2. Experiments

To test the above research questions and hypotheses, we tested various conditions: congruent trials with the same modality (visual/visual; verbal/verbal; acoustic/acoustic) and incongruent ones with different modalities required (visual/verbal; verbal/visual; acoustic/visual; visual/acoustic). A comparison of verbal and acoustic was neglected here. We rather wanted to focus on modality switching with visual information since visual information is almost always present in everyday life (when we need to switch between modalities). We performed recognition as well as wayfinding experiments. As a side note, in the recognition experiments distractors were shown to the participants (objects that were not presented during the learning phase), which will not be presented and discussed in detail here, since they are only necessary for obtaining possible errors. In other experiments we could show that landmarks may be processed in other modalities, and additionally, even better than in the visual modality (Hamburger et al. in preparation). Therefore, our focus here is on the switching costs for making *Gestalt* characteristics or full *Gestalten* available in a different modality at retrieval than at learning.

2.1 Methods

The experimental design for the following experiments was a within-subject design with one factor. The dependent variables for each experiment were the performance (correct recognition; correct route decisions) and the decision times required for recognition and for route decisions.

Due to the fact that the material for the three experiments described in this article is the same, with variations in the landmark objects and modalities, it will be described here once, with only the additional variations in the appropriate sections. The samples will also be described within the appropriate experimental sections.

2.1.1 Material – Virtual Environment

The experiments were conducted in our 3D virtual maze (Squareland), which was programmed with the freeware graphics software Google SketchUp 6.4 © by Google ©. The maze consisted of a 10x10 block design with an edge length of 5.5x5.5m and a height for each block of 2.95m. The paths between the blocks were 2.75m wide. Since all "streets" between the blocks are in orthogonal direction to each other, it is somehow reminiscent of the layout of many major north-american cities. This simple structure provides the possibility of high experimental control, since –without additional objects or other types of information– all intersections and path sections look identical. Thus, only the landmarks/objects provided in the maze offer valuable information for successful wayfinding. Figure 1 gives an impression of the virtual environment used in this experimental series. For a more detailed description of the SQUARELAND virtual environment the reader is referred to the work by Kai Hamburger and Markus Knauff (in press).

For the current experiments the environment was designed as an indoor environment, with typical indoor structures such as white walls, a wooden floor and a ceiling. For the experiment a video clip was created with the freeware Fraps 3.1.1 (by Beepa Ltd.). The simulated eye-height of the participant was set to 1.70m which is close to the current average human eye-height. The simulated walking speed was 2.3m/s. The different routes that had to be travelled were identical across the different conditions and participants were randomly assigned to the conditions. The full path length was 176m; start and goal were highlighted by wooden doors.

To avoid that participants could see through the maze and perceive more than one intersection (or landmark) at a time, a haze was implemented in front of the participants. This haze had a distance of 10m from the participant. Thus, participants could only see one landmark (object) at a time. After one landmark disappeared from view, the following landmark and decision point could not be perceived yet. This was necessary to provide clear links between certain landmarks and their corresponding decision points.

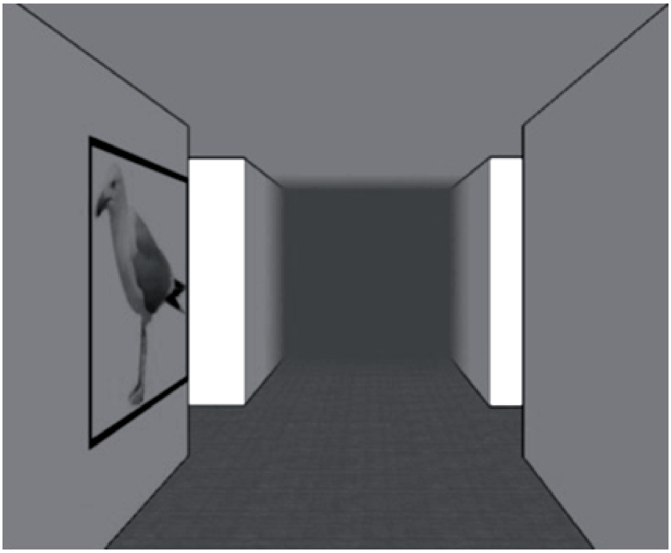


Fig. 1: Screenshot of the virtual environment (SQUARELAND) used in this experimental series. Shown is an example of the conditions using animal pictures (here: seagull). You find the coloured version of this figure at http://gestalttheory.net/gth/meaning.html

The projection screen subtended 60 deg in height and 67 deg of visual angle in width (170 x 230cm at a distance of 100cm).

2.1.2 Material – Landmark Objects

For the first two experiments we used pictures, names (words), and sounds of animals as landmarks. In a pre-experiment a total of 20 participants evaluated the stimulus material (37 animals) to verify that the sounds, the corresponding images, and names were attributed to the same animals, so that the experimental conditions were comparable (equally difficult). For this purpose, participants had to assign animal words to animal pictures, pictures to words, sounds to words, words to sounds, sounds to pictures, and pictures to sounds. We measured the correct assignments as well as the time required for each assignment. With this method we obtained the basis for reliably choosing the best 24 objects (out of 37) for Experiments 1 and 2.

For the third experiment we used pictures and names of famous buildings. It was determined in a pre-experiment (*N*=22) which of 98 given buildings from all over the world were judged as most famous by the participants and they also had to make assignments as described above. The 24 with the best values (highest famousness and the fewest incorrect assignments) were used in the experiment. During the recognition phase the images were surrounded by a black frame and were presented centrally on the screen without any perspective distortion as has been the case at the wall within the maze. We need to mention that distortions did not affect participants' performance (objects could perceptually equally well be recognized). The sounds were visually indicated by a pictogram of a loudspeaker on the screen. Thus, objects had to be recognized from a different perspective, which should not be more difficult for such material, since all animals were rather easy to identify.

2.1.3 Procedure

Participants were passively led through the virtual environment and had to learn a certain route including the visual or verbal information on the left wall. Presenting the material on the left had two reasons. First, when a word is presented on the left it is possible to read it in the 'natural' direction and in direction of the path, which is not the case when it is shown on the right, as then participants have to read the word 'backwards' (in the opposite direction of walking). Second, keeping the position constant prevented participants from only encoding the structural path information (e.g., the object is always in direction or in opposite direction of the turn). In the case of sounds the sounds were given when the visual and verbal information came into sight (presentation time was therefore equal for all stimuli with duration of 4s). In the subsequent test phases the participants had to pass a recognition task and afterwards a wayfinding task. In the recognition task they had to decide whether they were presented with a certain landmark in the learning phase or not (independent of modality). They had to indicate their decision via a key press on a response box RB-530 (Cedrus Corporation ©). The keys were labeled with the required answers: yes (means "I have seen this stimulus in the learning phase") and no ("I did not see this stimulus in the learning phase"). The stimuli in the recognition task (pictures) on the screen had an average size of 70 cm (35 deg) in width and 50 cm (27 deg) in height. Letters of the words were 10 cm in height for upper case letters and 7 cm for lower case letters. Words were easily readable and the full size of words differed due to word length. Sounds had a loudness of 35 decibel (db). During the wayfinding task participants had to make the appropriate decision at the intersection to follow the correct route. Stimulus size varied here due to the movement within the maze.

The experiments consisted of modality congruent and modality incongruent trials. In modality congruent trials the required sensory modality for processing the information in the learning phase remained the same in the test phase (e.g., visual/visual). In the incongruent trials there was a switch from one sensory modality in the learning phase to a different sensory modality in the test phase (e.g., visual/acoustic). We tested with a within-subject design, so that modality-switch and no modality-switch was tested for each participant.

3. Experiment 1: Visual and Acoustic

In the first experiment we compared visual (animal pictures) and acoustic (animal sounds) landmarks. Therefore, we could compare four different categories (types of processing; sequence of required modalities): visual and visual, visual and acoustic, acoustic and visual, and acoustic and acoustic. The material and procedure for this experiment are described above.

3.1. Method

3.1.1 Participants

The sample of this experiment consisted of 20 students from the University of Giessen (19 females, 1 male). The mean age was 25.7 years (*SD*=7.1). All of them were naïve with respect to the research questions and did not have any pre-experience with wayfinding experiments in virtual environments. All participants had normal or corrected-to-normal visual acuity. None of them suffered from epileptic seizures or motion sickness, which were exclusion criteria for this study. They received course credits for participation in the experiment. Participation was limited to one experiment, so that they were not included in any other wayfinding experiment of this series. All participants provided informed written consent.

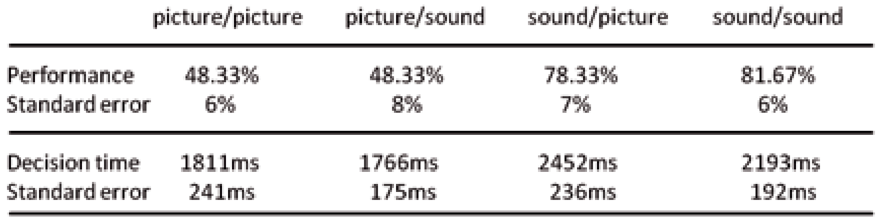
3.2. Results

All data reported in the following experiments represent correct responses in the form of correct assignments whether a stimulus was present in the learning phase or not (given in %) and decision times (given in milliseconds, ms). We calculated a one factorial F-test for repeated measures and a post-hoc t-test if the F-test provided any significant effects for all of the following comparisons between the four categories. For all other data we calculated a t-test for repeated measures. For all results the values are Greenhouse-Geisser corrected for the F-values and we used the Bonferroni-Holm correction for the post-hoc t-values.

3.2.1. Recognition

The mean performance for this experiment was 64.17%. For the landmarks the F-test for the performance showed differences between the four categories (*F*(3)=10.827, *p*<.001) (see table 1). Post-hoc tests revealed the following significant differences: picture to picture differed from sound to picture (*t*(19)=-4.414, p<.001); picture to picture differed from sound to sound (*t*(19)=-4.156, *p*=.001); picture to sound differed from sound to picture (*t*(19)=-3.757, *p*=.001); and picture to sound differed from sound to sound (*t*(19)=-3.249, *p*=.004).

For the decision time we had a mean value of 2053ms. For the landmarks the F-test revealed a significant result *F*(3)=4.376, *p*=.008 (see table 1). Post-hoc tests showed a significant difference between picture to sound and sound to picture (*t*(19)=-3.245, *p*=.004). The other pairwise comparisons were not statistically significant.



	picture/picture	picture/sound	sound/picture	sound/sound
Performance	48.33%	48.33%	78.33%	81.67%
Standard error	6%	8%	7%	6%
Decision time	1811ms	1766ms	2452ms	2193ms
Standard error	241ms	175ms	236ms	192ms

Table 1: Mean performance and decision times for the landmarks (animal pictures and sounds) in the recognition task across all four modality conditions.

3.2.2 Modality Switch

If we take a look at the two possible combinations of modalities we see that we can merge the change from one modality to another (picture to sound and sound to picture) and compare these results with the other values (picture to picture and sound to sound) (see figure 2). Here, we obtained neither for the performance nor for the decision time significant differences between the two groups (both t-values are smaller than 1).

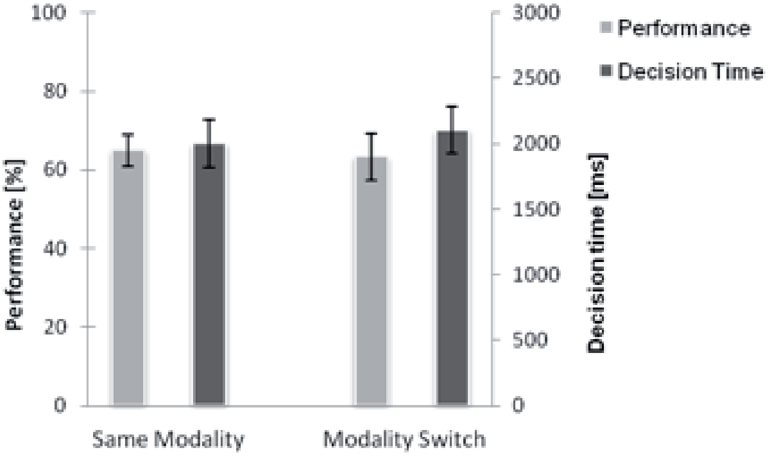


Fig. 2: Mean performance and decision times in the recognition task (animals) for the same modality (picture to picture and sound to sound) and modality switch (picture to sound and sound to picture). Error bars denote the standard error.

3.3 Discussion

Thus far we can see that a modality switch between visual and acoustic information is possible (which would be expected), but it comes at no additional switching costs (which is unexpected from the perception literature). The question therefore arises whether the information is automatically processed in different modalities at learning or is transformed into the appropriate modality at retrieval. Since we could not find any evidence for switching costs, we may assume that the different *Gestalt* information is initially processed within different modalities so that the information is represented in a modality independent format in the brain. This is in line with the dual coding theory of spatial cognition by Tobias Meilinger and Markus Knauff (2008). One further point we would like to address is the difference between pictures and sounds. When pictures had to be learned initially, performance in the recognition task was at chance level independent of the modality (48%). When sounds were learned, performance was much better (80%). These differences in favor of the encoding of sounds has been found earlier (Röser et al. 2011). This means that landmarks which are presented as sounds are easier to remember for participants compared to pictures. This, however, does not affect the results for the modality switch, because we had two combinations of stimuli for both conditions (sounds/sounds and pictures/pictures; pictures/sounds and sounds/pictures). Furthermore, it will be shown in the wayfinding task that participants were able to make the correct route decisions in combination with the stimulus material that had to be learned (e.g., wayfinding performance was above chance level in all conditions).

4. Experiment 2: Visual and Words (Semantic)

In the second experiment we investigated visual (animal pictures) and verbal (words; animal names) landmarks. The material and procedure are described above.

4.1. Method

4.1.1 Participants

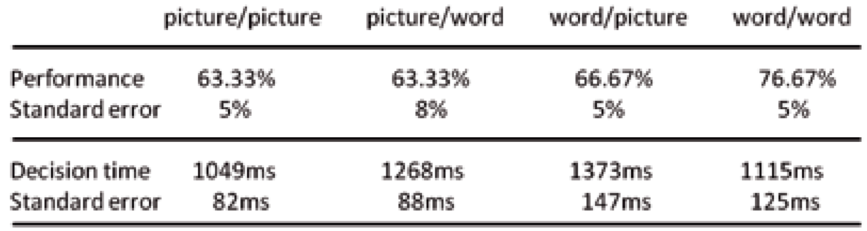
The sample consisted of 20 students from the University of Giessen (19 females, 1 male). The mean age was 22.05 years (*SD*=2.3). They all met the same criteria as the participants of the previous experiment and received the same compensation.

4.2. Results

4.2.1 Performance

Participants showed a mean performance of 67.5% correct responses. For the landmarks the F-test revealed no significant differences between the four categories (*F*(3)=1.441, n.s.) (see table 2).

For the decision time we obtained a mean of 1201ms. For the landmarks the F-test showed a significant result *F*(3)=3.029, *p*=.037 (see table 2). Yet, post-hoc tests revealed no significant differences for any of the pairwise comparisons.



	picture/picture	picture/word	word/picture	word/word
Performance	63.33%	63.33%	66.67%	76.67%
Standard error	5%	8%	5%	5%
Decision time	1049ms	1268ms	1373ms	1115ms
Standard error	82ms	88ms	147ms	125ms

Table 2: Mean performance and decision times for the landmarks (animal pictures and words) in the recognition task across all four modality conditions.

4.2.2 Modality Switch

The results for the comparison between modality switch and equal modalities are presented in figure 3. However, the t-test for the performance showed no significant difference between the two conditions (*t*(19)<1). But we found a significant result for the decision time in the recognition task. The modality switch led to a slightly slower decision time (*t*(19)=-2.610, *p*=.017).

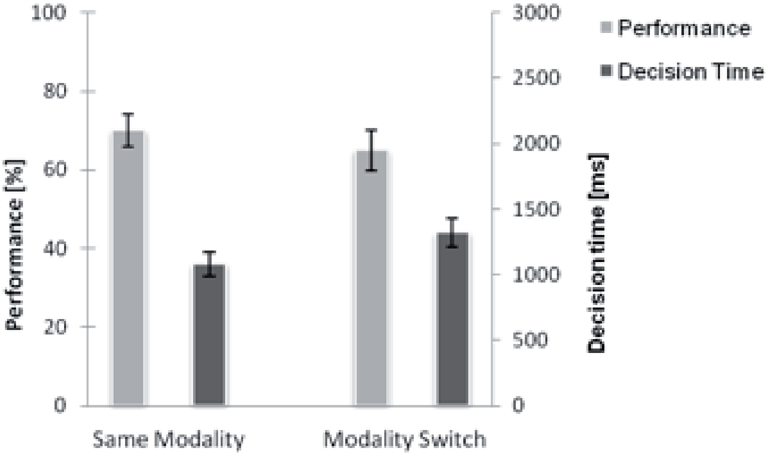


Fig. 3: Mean performance and decision times in the recognition task (animals) for the same modality (picture to picture and word to word) and modality switch (picture to word and word to picture). Error bars denote the standard error.

4.3 Discussion

In contrast to the first experiment the categories with a picture in the learning phase provided a higher performance here and are significantly above chance level (*t*(19)=2.491, *p*=.022). This could have happened due to the fact that pictures and words are more similar than pictures and sounds (initial processing in the visual system) possibly leading to the above mentioned overlapping effect. Regardless of whether they are different from chance level or not, we want to concentrate on the modality switch. A modality switch between visual and verbal information is also possible in either direction. This time there were at least some switching costs in the decision times present. This supports the notion that the required information is translated into the correct modality at retrieval. However, even though significant, these costs are within a range of 238ms. Thus, processing and retrieval of these stimuli is again very effective. Therefore, we again assume that the necessary information is processed independently of the modality in which it is initially presented. In other words, an image can be present in the form of a word, while a word may elicit a mental image of the stimulus. This assumption also favors a more global and holistic processing strategy, since processing single features first and then (sequentially) putting them together to build a whole (object) would probably require more time. We would also like to stress the notion of the dual coding theory of spatial cognition (Meilinger & Knauff 2008; Meilinger, Knauff, et al. 2008) here. This was true for animated content (animals). Since animated and unanimated objects are represented differently in the brain (e.g., Vidal, Ossandón, et al. 2010), it is also of interest to investigate a modality switch with unanimated objects (here buildings).

5. Experiment 3: Visual and Words (Buildings)

In the third experiment we compared visual and verbal (words) landmarks. The material and procedure is described above. In contrast to the previous experiments here the landmarks (24) were famous buildings from all over the world. They were either presented in pictorial or textual form.

5.1. Method

5.1.1 Participants

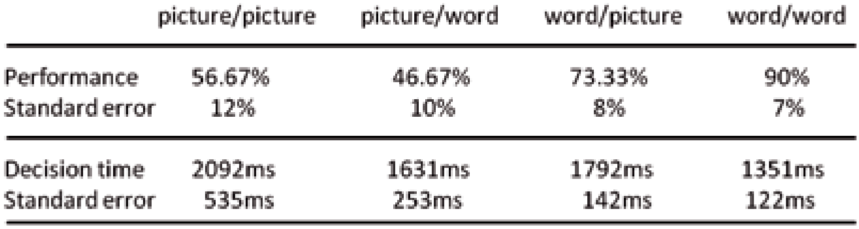
The sample of this experiment consisted of 10 participants (half of them students of the University of Giessen; 8 females, 2 males). The mean age was 28.2 years (*SD*=9.2). They all met the same criteria as the participants of the previous experiments and received the same compensation.

5.2. Results

5.2.1. Performance

The mean correct recognition for the landmarks was 66.67% and there was a significant difference between the four categories present (*F*(3)=4.593, *p*=.010) (see table 3). Post-hoc t-tests showed significant differences between pictures to words and words to words (*t*(19)=-3.545, *p*=.006).

The mean decision time for recognition was 1698ms and the F-test showed no significant differences between the four categories (*F*(3)=1.139, *p*=.351) (see table 3).



	picture/picture	picture/word	word/picture	word/word
Performance	56.67%	46.67%	73.33%	90%
Standard error	12%	10%	8%	7%
Decision time	2092ms	1631ms	1792ms	1351ms
Standard error	535ms	253ms	142ms	122ms

Table 3: Mean performance and decision times for the landmarks (pictures and names of famous buildings) in the recognition task across all four modality conditions.

5.2.2 Modality Switch

The results for the comparison between modality switch and equal modalities are presented in figure 4. Here, no significant differences could be found neither for the performance in the recognition task (*t*(19)<1) nor for the decision time (*t*(19)<1).

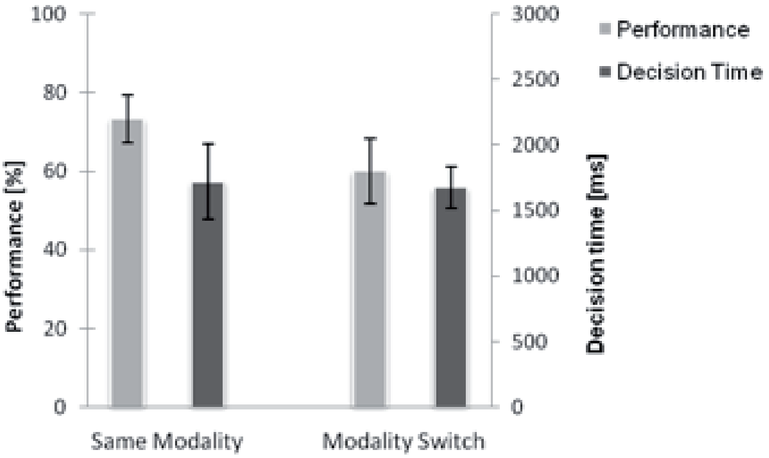


Fig. 4: Mean performance and decision times in the recognition task (buildings) for the same modality (picture to picture and word to word) and modality switch (picture to word and word to picture). Error bars denote the standard error.

5.3 Discussion

Again, a modality switch between learning and retrieval could be performed. A lower performance when switching from the visual to the verbal modality could have been expected, but the performance difference within the word modality comes as a surprise. There, recognition should have been the same, since pure recognition without a modality switch is required. More interesting is that again there are no additional costs for a modality switch. This indicates once more that the full *Gestalt* (with all the accompanying information; e.g., meaning) is processed right away at the moment of its first occurrence. If this was not the case, switching costs must have been present. These findings, this time in the semantic (verbal) domain, support the results of the previous experiments and the introduced theory.

So far, we presented results of rather simple recognition tasks for landmarks after learning the landmarks and route information (directions). It is important to note that we are capable of storing and retrieving such information in different modalities and that storage seems to be modality independent. However, what happens if we learned the information in one modality, but have to navigate through an unfamiliar environment and need this previously learned information? In other words, I learned landmarks and route information from a verbal description but have to recognize the buildings and the corresponding turns while navigating through the environment. Does that require additional cognitive resources and is this then accompanied by switching costs? Possible answers may be provided from our wayfinding (navigation) task which will be reported next.

As a side note, we again obtained performances around chance level for the picture condition (pictures in the learning phase). A possible explanation for this is different than that from the previous experiments. Here, an overlap effect due to the different sensory modalities could not occur. In contrast to the second experiment, more complex stimulus material was used (buildings). It seems as if it is more difficult for participants to remember pictures of buildings than the corresponding names. This is supported by the conditions including a modality switch, where a higher performance for the word to picture condition than for the picture to word condition was obtained.

Additionally, we here need to discuss the decision time differences between the three experiments. Possible explanations are the following. In the sound condition (Experiment 1) participants normally waited until the sound was terminated before they responded. This might result in the attitude that they do not need to respond very fast in the picture condition, which might in turn lead to longer decision times in the visual condition as well. Previous experiments testing only isolated modalities only showed an increased decision time in the sound conditions. Experiment 2 overall revealed the shortest decision times. A possible explanation is that simple, over-learned and comparable (pictures and words) stimulus material (animals) was used. There, the processing of the whole *Gestalt* occurs rather fast and more or less automatically (without any conscious percept). In Experiment 3, however, we used more complex stimulus material (names and pictures of famous buildings) that requires conscious processing (since it is not over-learned), resulting in longer decision times.

6. Wayfinding Task

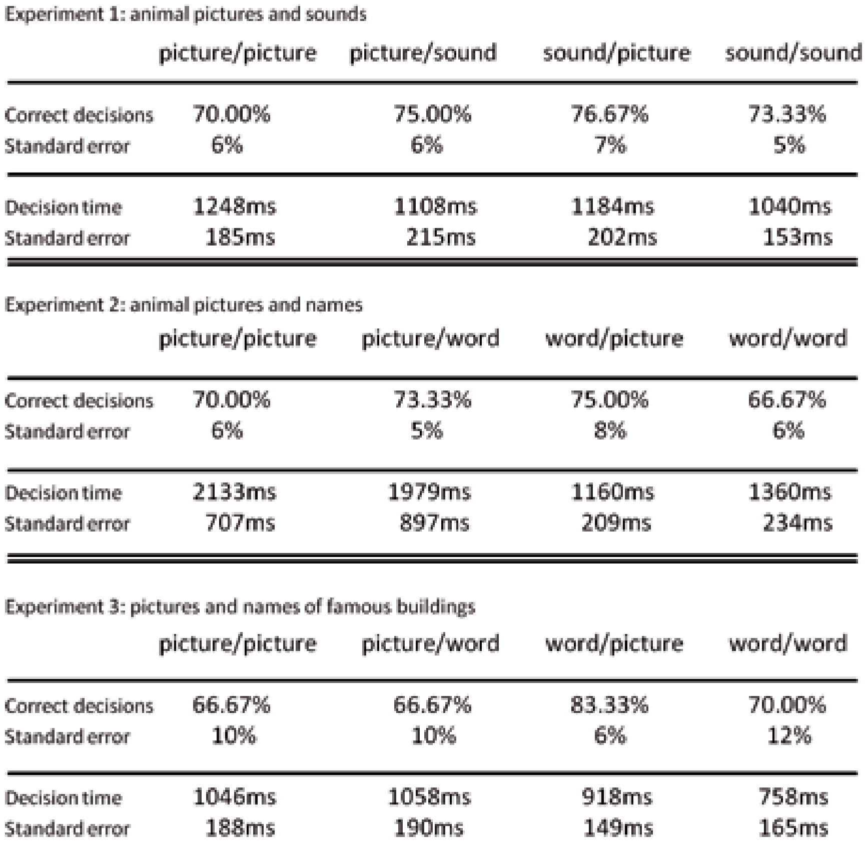
In the following section we describe the results for the wayfinding task, which was part of all previously reported experiments.

6.1 Method

Participants and material were already described above for each experiment. In each experiment, following the recognition task, participants were again presented the video sequence of the learning phase. All of them were similarly designed, namely the participants read a short instruction with the task to find the correct path they had previously learned. Then they saw the video sequence again starting from the beginning. This time, the movie stopped at each intersection and they had to indicate the direction in which the movie sequence/path would proceed (only left and right turns) via key presses. Dependent variables here were the number of correct route decisions (at the intersections) and the decision time.

6.2 Results

A detailed presentation of the number of correct responses and the decision times for the four different categories in the three experiments are presented in table 4.



Experiment 1: animal pictures and sounds

	picture/picture	picture/sound	sound/picture	sound/sound
Correct decisions	70.00%	75.00%	76.67%	73.33%
Standard error	6%	6%	7%	5%
Decision time	1248ms	1108ms	1184ms	1040ms
Standard error	185ms	215ms	202ms	153ms

Experiment 2: animal pictures and names

	picture/picture	picture/word	word/picture	word/word
Correct decisions	70.00%	73.33%	75.00%	66.67%
Standard error	6%	5%	8%	6%
Decision time	2133ms	1979ms	1160ms	1360ms
Standard error	707ms	897ms	209ms	234ms

Experiment 3: pictures and names of famous buildings

	picture/picture	picture/word	word/picture	word/word
Correct decisions	66.67%	66.67%	83.33%	70.00%
Standard error	10%	10%	6%	12%
Decision time	1046ms	1058ms	918ms	758ms
Standard error	188ms	190ms	149ms	165ms

Table 4: Mean correct decisions and decision times for the wayfinding task of the three experiments.

For the comparison between animal sounds and pictures (Experiment 1) the mean for correct direction decisions was 73.75% (*SEM*=6%). The mean decision time was 1145ms (*SEM*=189ms). Neither for the direction decisions (*F*(3)<1) nor for the decision time did the F-tests show a significant difference (*F*(3)<1) between the four categories. This also accounts for the comparison of modality switch and equal modality for the direction decisions (*t*(19)<1) and the decision time (*t*(19)<1).

The mean wayfinding performance for Experiment 2 (the comparison between pictures and words of animals) was 71.25% (*SEM*=6%) for the correct direction decisions and 1658ms for the decision time (*SEM*=512ms). Here again, no significant differences between the four categories, neither for the correct direction decisions (*F*(3)<1) nor for the decision time (*F*(3)<1) occurred. Just as for the previous experiments, the comparison between switching modalities and equal modalities were not significant neither for the correct direction decisions (*t*(19)<1) nor for the decision time (*t*(19)<1).

In Experiment 3 (comparison of pictures and names of famous buildings) the mean performance for the wayfinding task was 71.67% (*SEM*=9%) with a mean decision time of 1658ms (*SEM*=173ms). Again, there are no significant differences between the four categories, neither for the performance (*F*(3)<1) nor for the decision time (*F*(3)<1). The results for the comparison between modality switch and equal modalities did not reveal any significant differences for the correct direction decisions (*t*(19)<1) and the decision time (*t*(19)<1).

6.3 Discussion

What we have obtained for the recognition tasks is also true for the wayfinding task. Here, as well, a modality switch did not have any influence or additional processing costs for the landmarks or direction information. If we need to orientate ourselves in or have to find our way through an environment, the spatial information is automatically (unconsciously) processed within different modalities.

7. General Discussion and Conclusion

First, we may notice that our participants were able to process (and make use of) landmark information in different modalities in a recognition and wayfinding task. Thus, healthy participants are capable of using *Gestalt* (landmark) information from modalities other than just vision. This challenges the classical view concerning the importance of visual information in landmarks. We may therefore think about reducing the amount of visual information in route descriptions/route information, since the visual modality is already occupied by other information (e.g., paying attention to traffic and traffic lights, the environment and/or the general wayfinding and walking task). The processing of a landmark *Gestalt* seems also to be easier in other modalities (Hamburger et al. in preparation; Röser et al. 2011).

In general, we also find that participants are able to successfully switch between the modalities at learning and retrieval. But, this is not the whole story. They also switch the modality at no additional cost, meaning that the performance is stable no matter whether the modality needs to be switched or not. This is an interesting finding since the opposite would be expected from the literature. Such challenging results and the accompanying discussion should be integrated into future research. Above we stated that the visual domain should be relieved. We also have evidence for this assumption from the modality switching results, since participants performed better when acoustic or verbal information was presented during the learning phase but were later confronted with visual information. When they were presented with visual information during learning, performance was worse during test.

To summarize, we did not find evidence for systematic costs when it comes to modality switching in recognition of landmarks and wayfinding with landmark information. At a first glance the absence of any statistical differences (null-effects) might appear to be disappointing but this seems to be wrong as we will try to point out in the following.

If we had found systematic evidence for modality switching costs, this would mean that our brain has to deal with an increased workload in order to process the available information to prepare and then execute a relevant action (here: to make a turn or stay on the path). However, this is not what we found with our recognition and wayfinding paradigms. Participants were performing the tasks equally well no matter whether the learning conditions (modalities) were congruent with the retrieval situation or not. Thus, our brain is excellently capable of dealing with *Gestalt* information in different modalities (transformation or adjustment) without additional costs. Some supporting evidence may be found in the theory of Meilinger and Knauff (2008) who assume that much of the relevant information, no matter in which form it was initially presented, is automatically made available in different forms, e.g., propositional and visual. It is therefore possible that at the moment of information retrieval (from working memory or long-term memory), the necessary *Gestalt* information is already available to the different modalities without the necessity of any further processing.

As a side note, one might argue that the verbal information is also a kind of visual information. This is correct but other processing steps are involved, since for example a word needs to be mentally transformed in order to be comparable to the real visual information. Therefore, the words are just an abstract visual representation of visual content. The word "frog" may result in different interpretations by participants of how it looks like and it does not necessarily have to be green as was the case in our visual condition. Furthermore, in the first instance a word belongs to the semantic domain whereas a picture does not.

These assumptions on modality switching and landmark information are thus far quite speculative since they are based on just a few empirical findings from human landmark and wayfinding experiments. However, they provide first valuable insights and should encourage us to further investigate this field. It may also prove worthwhile to use navigational relevant information in modalities other than just vision, not only in visually impaired or blind people (e.g., Loomis et al. 1998) but also for *normal* people in order to reduce the amount of perceptual and cognitive load in a single modality.

Our assumptions in the discussion sections to each experiment are for example in line with functional imaging studies, reporting that an acoustic stimulation (animal sounds) also leads to activation increases in the visual cortex (Tranel, Damasio, et al. 2003). It seems as if we automatically perform (conscious or unconscious) mental imagery. Thus, the *Gestalt*/landmark information is indeed processed in different modalities. This finding may not explain why a visual stimulation (animal pictures) does not result in additional activations in auditory areas of the brain (Finney, Clementz, et al. 2003). This latter finding would again suggest switching costs which could not be found in our experiments. However, it again highlights the importance of and the reliance on the visual system, even though other modalities are capable of taking over the job. Here, we provided first behavioral evidence for the absence of switching-costs in human wayfinding. We also provided some evidence for holistic *Gestalt* processing in this research field. Future research may then address questions of whether landmarks are processed in the same brain areas as simple objects without any context information or additional activations due to the linkage of landmark/*Gestalt* with its surround (context). It might also be that other brain areas turn out to be of importance that we might not have on our "list of expected candidates" so far. In order to find conclusive answers, more brain imaging studies on the neural correlates of landmarks are required, which is one of the very next points on our research agenda.

Finally, we pointed out that our brain automatically processes *Gestalt* information in different modalities, even if they were exclusively presented in a single modality. Thus, our knowledge and our experiences –retrieved from long-term memory– are highly involved in the processing of a *Gestalt* in spatial cognition. In previous experiments we were able to demonstrate that single (visual) features do not represent a full landmark (Röser et al. 2011). Instead, meaning arises from the combination of visual features (or other modalities), our semantic knowledge about these objects, or experiences with them, etc. Initially we cited Christian von Ehrenfels with the words that "the whole is more than just the sum of its parts" and now we may complete the circle: landmarks and direction information are linked together as a *Gestalt*, e.g., at the church I have to turn left, etc. Every single information/feature on its own does not contain much valuable information for wayfinding, but together they provide us with all the necessary information that we need to successfully navigate or differentiate environments. We showed that the concept of a *Gestalt* is also an important one in the research domain of spatial cognition and we hope that this contribution motivates to move on in this direction.

Summary

How much does it cost to switch between different modalities when we have to process *Gestalt* information in form of landmarks in wayfinding (navigation)? In a series of experiments we show that in recognition and wayfinding tasks with landmarks there is no evidence for costs in modality switching (lower performance, increased decision times). For example, learning visual information and retrieving visual information is as effective as learning visual information but retrieving this information in the acoustic domain and vice versa. This has been tested for visual, verbal, and acoustic material (images, words/names, and sounds of various animals and images and words/names of famous buildings). Our results challenge the notion of switching-costs in the domain of human wayfinding. We assume that the human brain already integrates the relevant *Gestalt* information in different modalities so that no additional costs occur at the time of information retrieval. Furthermore, the connections between the path and the landmarks seem to be of great relevance, since wayfinding performance was as good as performance in the recognition task. Our conclusion is that a modality switch is possible at no additional costs so that landmarks may also be useful in more (different) modalities than just the visual one. Furthermore, we have evidence that our cognitive system processes more information during the learning of landmarks and spatial information than is physically present, so that the whole is in this cognitive domain (spatial cognition) as well is much more than just the sum of its parts.

Keywords: Landmarks, wayfinding, modality switching, modality independent processing.

Zusammenfassung

Wie viel kostet es, zwischen verschiedenen Verarbeitungsmodalitäten zu wechseln, wenn wir *Gestalt*information in Form von Landmarken beim Wegfinden (Navigation) verarbeiten müssen? In einer Reihe von Experimenten zeigen wir auf, dass es beim Wiedererkennen und Wegfinden von und mit Landmarken keine Evidenzen für sogenannte Wechselkosten gibt (niedrigere Leistung, höhere Entscheidungszeiten). Beispielsweise erwiesen sich das Lernen und der Abruf von visueller Information als gleich effektiv wie das Erlernen von visueller Information, die dann in der akustischen Domäne abgerufen werden musste. Dies wurde für visuelles, verbales und akustisches Material untersucht (Bilder, Wörter/Namen und Geräusche von Tieren sowie Bilder und Wörter/Namen von berühmten Gebäuden). Unsere Befunde widersprechen der Auffassung von Wechselkosten beim Wegefinden. Wir gehen davon aus, dass das menschliche Gehirn sämtliche relevante *Gestalt*information bereits in die verschiedenen Modalitäten integriert, so dass es in der Abrufsituation zu keinen zusätzlichen Verarbeitungsschritten und damit Wechselkosten kommt. Darüber hinaus scheinen die Verknüpfungen zwischen Weginformation und Landmarke von großer Bedeutung zu sein, da die Wegfindeleistung mindestens genauso gut war wie das reine Wiedererkennen. Unsere Schlussfolgerung lautet, dass ein Modalitätswechsel möglich ist und dieser ohne zusätzliche Verarbeitungskosten stattfindet, so dass Landmarken auch durchaus in anderen Modalitäten als der visuellen hilfreich sein können. Zusätzlich gibt es Evidenzen, die für eine Verarbeitung von mehr Landmarken- und Routeninformationen sprechen als tatsächlich während der Präsentation gegeben sind, so dass das Ganze auch in diesem kognitiven Bereich (Raumkognition) mehr ist als die Summe seiner Teile.

Schlüsselwörter: Landmarken, Wegfinden, Modalitätswechsel, modalitätsunspezifische Verarbeitung.

Acknowledgement

This study was supported by the German Research Foundation (DFG HA5954/1-1). We thank Karl Kopiske and Malina Traub for their help with the data collection and Nadine Jung for proofreading the manuscript.

References

Abruthnott, K. D. & Woodward, T. S. (2002): The influence of cue-task association on switch costs and alternating-switch costs. *Canadian Journal of Experimental Psychology 56*, 18 – 29.

Caduff, D. & Timpf, S. (2008): On the assessment of landmark salience for human navigation. *Cognitive Processing 9,* 249 – 257.

Eagleman, D. M. (2001): Visual illusions and neurobiology. *Nature Reviews Neuroscience 2*, 920 – 926.

Finney, E. M., Clementz, B. A., Hickok, G. & Dobkins, K. R. (2003): Visual stimuli activate auditory cortex in deaf subjects: evidence from MEG. *NeuroReport 14*, 1425 – 1427.

Habel, C., Kerzel, M. & Lohmann, K. (2010): Verbal assistance in tactile-map explorations: A case for visual representations an reasoning, in McGreggor, K. & Kunda, M. (eds., cochairs) (2010): *Papers from the 2010 AAAI Workshop Visual Representation and Reasoning.* Technical Report WS-10-07. Menlo Park, California: The AAI Press.

Hamburger, K. (2007): *Visual illusions – Perception of luminance, color, and motion in humans*. Saarbrücken: VDM Verlag Dr. Müller.

Hamburger, K. & Knauff, M. (in press): Squareland: A virtual environment for investigating cognitive processes in human wayfinding. *PsychNology*.

Hamburger, K., Röser, F. & Knauff, M. (in preparation): Can we hear famous landmarks? – A comparison of visual, verbal, and acoustic landmarks and the influence of familiarity.

Koffka, K. (1935): *Principles of Gestalt Psychology.* New York: Hardcourt, Brace.

Loomis, J. M., Golledge, R. G. & Klatzky, R. L. (1998): Navigation system for the blind: Auditory display modes and guidance. *Presence 7,* 193 – 203.

Lynch, K. (1960): *The Image of the City.* Cambridge, MA: MIT Press.

Meilinger, T. & Knauff, M. (2008): Ask for directions or use a map: A field experiment on spatial orientation and wayfinding in an urban environment. *Journal of Spatial Science 53,* 13 – 24.

Meilinger, T., Knauff, M. & Bülthoff, H. H. (2008): Working memory in wayfinding – A dual task experiment in a virtual city. *Cognitive Science 32,* 755 – 770.

Presson, C. C. & Montello, D. R. (1988): Points of reference in spatial cognition: Stalking the elusive landmark. *British Journal of Developmental Psychology 6,* 378 – 381.

Robinson, J. O. (1972): *The Psychology of Visual Illusions*. London: Constable and Company.

Röser, F., Hamburger, K. & Knauff, M. (2011): The Giessen virtual environment laboratory: human wayfinding and landmark salience. *Cognitive Processing 12*, 209 – 214.

Sharps, M. J. & Wertheimer, M. (2000): Gestalt perspectives on cognitive science and on experimental psychology. *Review of General Psychology 4*, 315 – 336.

Spillmann, L. & Ehrenstein, W. H. (1996): From neuron to Gestalt: mechanisms of visual perception, in Greger, R. & Windhorst, U. (eds.) (1996): *Comprehensive Human Physiology*, Vol. 1., 861 – 893. Heidelberg: Springer.

Steinman, R. M., Pizlo, Z. & Pizlo, F .J. (2000): Phi is not beta, and why Wertheimer's discovery launched the Gestalt revolution. *Vision Research 40*, 2257 – 2264.

Tranel, D., Damasio, H., Eichhorn, G. R., Grabowski, T., Ponto, L. L. B. & Hichwa, R. D. (2003): Neural corre-lates of naming animals from their characteristic sounds. *Neuropsychologia 41,* 847 – 854.

Tse, P. U. (2004): Unser Ziel muss eine Gestalt-Neurowissenschaft sein. *Gestalt Theory 26*, 287 – 292.

Vidal., J. R., Ossandón, T., Jerbi, K., Dalal, S. S., Minotti, L., Ryvlin, P., Kahane, P. & Lachaux, J.-P. (2010): Category-specific visual responses: An intercranial study comparing gamma, beta, alpha, and ERP response selectivity. *Frontiers in Human Neuroscience 4*, 195.

von Ehrenfels, C. (1890): Über Gestaltqualitäten. Vjschr. wiss. Philos. 14, 249 – 292. Translation in *Foundations of Gestalt Theory* (B. Smith Ed.), Munich: Philosophia-Verlag, 1988. (pp 82 – 120)

Wertheimer, M. (1912): Experimentelle Studien über das Sehen von Bewegung. *Zeitschrift für Psychologie 61*, 161 – 265.

Kai Hamburger, born in 1977, received his diploma in Psychology from the University of Frankfurt in 2004 (former Max Wertheimer chair). Since 2003 he is collaborating with Prof. Dr. Lothar Spillmann (Freiburg). 2005-2007 scholarship in the graduate program "Neural Representation and Action Control – NeuroAct (DFG 885/1) and PhD student in the research group Experimental Psychology University Giessen (Prof. Karl R. Gegenfurtner, PhD); graduation in 2007 (dr. rer. nat.). Currently he is assistant professor in the research group Experimental Psychology and Cognitive Science at the University of Giessen (Prof. Dr. Markus Knauff). Main research topics are spatial cognition (human wayfinding) and visual illusions.

Adress: Experimental Psychology and Cognitive Science, Justus Liebig University Giessen,
Otto-Behaghel-Str. 10F, 35394 Giessen, Germany.
E-Mail: kai.hamburger@psychol.uni-giessen.de

Florian Röser, born in 1982, received his diploma in Psychology from the University of Trier in 2009. Since 2010 he is PhD student in the research group Experimental Psychology and Cognitive Science at the University of Giessen (Prof. Dr. Markus Knauff). His research topics include landmarks, spatial orientation, and wayfinding (in a project of Dr. Kai Hamburger and Prof. Dr. Markus Knauff).
E-Mail: Florian.Roeser@psychol.uni-giessen.de