

# **UNIVERSITÄTSKLINIKUM HAMBURG-EPPENDORF**

Medizinische Fakultät der Universität Hamburg  
Institut für systemische Neurowissenschaften

Prof. Dr. med. C. Büchel

## **Semantic Incongruency Interferes With Endogenous Attention in Cross-Modal Integration of Semantically Congruent Objects**

### **Publikationspromotion**

zur Erlangung des Grades eines Doktors der Medizin  
an der Medizinischen Fakultät der Universität Hamburg.

vorgelegt von:

Julia Spilcke-Liss  
aus Hannover

Hamburg 2021

**(wird von der Medizinischen Fakultät ausgefüllt)**

**Angenommen von der  
Medizinischen Fakultät der Universität Hamburg am: 05.12.2022**

**Veröffentlicht mit Genehmigung der  
Medizinischen Fakultät der Universität Hamburg.**

**Prüfungsausschuss, der/die Vorsitzende: PD Dr. Gregor Leicht**

**Prüfungsausschuss, zweite/r Gutachter/in: Prof. Dr. Christian Büchel**

## **Inhaltsverzeichnis**

<b>1. Artikel in gedruckter Originalversion .....</b>	<b>1</b>
<b>2. Zusammenfassende Darstellung der Publikation .....</b>	<b>14</b>
<b>3. Zusammenfassung in deutscher und englischer Sprache .....</b>	<b>28</b>
<b>4. Literaturverzeichnis .....</b>	<b>30</b>
<b>5. Abkürzungsverzeichnis .....</b>	<b>33</b>
<b>5. Abbildungs- und Tabellenverzeichnis .....</b>	<b>33</b>
<b>7. Erklärung des Eigenanteils an der Promotion .....</b>	<b>34</b>
<b>8. Danksagung .....</b>	<b>35</b>
<b>9. Lebenslauf .....</b>	<b>36</b>
<b>10. Eidesstattliche Erklärung .....</b>	<b>38</b>



# Semantic Incongruency Interferes With Endogenous Attention in Cross-Modal Integration of Semantically Congruent Objects

Julia Spilcke-Liss<sup>1</sup>, Jun Zhu<sup>2\*</sup>, Sebastian Gluth<sup>3</sup>, Michael Spezio<sup>1,4</sup> and Jan Gläscher<sup>1\*</sup>

<sup>1</sup> Institute for Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg, Germany, <sup>2</sup> Department of Computer Science, Tsinghua University, Beijing, China, <sup>3</sup> Department of Psychology, University of Basel, Basel, Switzerland, <sup>4</sup> Psychology and Neuroscience, Scripps College, Claremont, CA, United States

## OPEN ACCESS

### Edited by:

Xiaolan Fu,  
Chinese Academy of Sciences, China

### Reviewed by:

Yong Gu,  
Shanghai Institutes for Biological  
Sciences (CAS), China  
Riccardo Brunetti,  
Università Europea di Roma, Italy

### \*Correspondence:

Jun Zhu  
dcszj@mail.tsinghua.edu.cn  
Jan Gläscher  
glaescher@uke.de

**Received:** 16 June 2019

**Accepted:** 29 August 2019

**Published:** 11 September 2019

### Citation:

Spilcke-Liss J, Zhu J, Gluth S,  
Spezio M and Gläscher J (2019)  
Semantic Incongruency Interferes  
With Endogenous Attention  
in Cross-Modal Integration  
of Semantically Congruent Objects.  
*Front. Integr. Neurosci.* 13:53.  
doi: 10.3389/fnint.2019.00053

Efficient multisensory integration is often influenced by other cognitive processes including, but not limited to, semantic congruency and focused endogenous attention. Semantic congruency can re-allocate processing resources to the location of a congruent stimulus, while attention can prioritize the integration of multi-sensory stimuli under focus. Here, we explore the robustness of this phenomenon in the context of three stimuli, two of which are in the focus of endogenous attention. Participants completed an endogenous attention task with a stimulus compound consisting of 3 different objects: (1) a visual object (V) in the foreground, (2) an auditory object (A), and (3) a visual background scene object (B). Three groups of participants focused their attention on either the visual object and auditory sound (Group VA,  $n = 30$ ), the visual object and the background (VB,  $n = 27$ ), or the auditory sound and the background (AB,  $n = 30$ ), and judged the semantic congruency of the objects under focus. Congruency varied systematically across all 3 stimuli: All stimuli could be semantically incongruent (e.g., V, ambulance; A, church bell; and B, swimming-pool) or all could be congruent (e.g., V, lion; A, roar; and B, savannah), or two objects could be congruent with the remaining one incongruent to the other two (e.g., V, duck; A, quack; and B, phone booth). Participants exhibited a distinct pattern of errors: when participants attended two congruent objects (e.g., group VA: V, lion; A, roar), in the presence of an unattended, incongruent third object (e.g., B, bath room) they tended to make more errors than in any other stimulus combination. Drift diffusion modeling of the behavioral data revealed a significantly smaller drift rate in two-congruent-attended condition, indicating slower evidence accumulation, which was likely due to interference from the unattended, incongruent object. Interference with evidence accumulation occurred independently of which pair of objects was in the focus of attention, which suggests that the vulnerability of congruency judgments to incongruent unattended distractors is not affected by sensory modalities. A control analysis ruled out the simple explanation of a negative response bias. These findings implicate that our perceptual system is highly sensitive to semantic incongruencies even when they are not endogenously attended.

**Keywords:** cross-modal integration, semantic congruency, exogenous attention, endogenous attention, drift diffusion model

## INTRODUCTION

Cross-modal, multi-sensory integration is one of the most remarkable achievements of perceptual processing as it enables the binding of information from different sensory modalities into a single coherent percept [see, e.g., (Senkowski et al., 2008) for a review]. Yet the efficiency of integration is influenced by several modulating factors including, but not limited to, spatial and temporal proximity (Meredith and Stein, 1986a,b), and semantic congruency (Taylor et al., 2006; Doehrmann and Naumer, 2008; Steinweg and Mast, 2017). By varying these modulating factors and observing their effects on multi-sensory integration, we can study how the brain accomplishes the requisite binding processes, along with the role of endogenous attention. To capture these dynamics requires a design that engages endogenous attention in selecting at least two objects for comparison, in the presence of at least one distractor, and controlling for modality.

Several studies point to the notion that attention is likely critical for the advantage that semantic congruence confers upon cognitive processes of cross-modal integration. For instance, recent accounts demonstrate a performance advantage for semantically congruent multisensory stimuli during visual search (Iordanescu et al., 2008, 2010), but only under low cognitive load (Matusz et al., 2015). Furthermore, semantic congruency of multi-modal stimuli facilitates perceptual processing of unrelated material at the same location of the congruent multisensory prime (Mastroberardino et al., 2015). The implication of this first line of research is that semantic congruency facilitates attentional selection at the location of the congruent stimuli and boosts perceptual processing and performance. This attentional focusing is not directly linked to the stimuli *per se* and therefore cannot be classified as “bottom-up” or “stimulus-driven” (Corbetta and Shulman, 2002; Koelewijn et al., 2010; Talsma et al., 2010). Rather, it is enhanced by the semantic congruency of the stimuli. The facilitation is therefore due to learned semantic associations and as such must be classed as a “top-down” process. Yet in these studies, the attentional engagement is exogenously controlled via semantic priming, and voluntary, endogenous attention was not investigated.

Contrasting with the previous literature, a more recent second line of research investigating the same cognitive processes arrived at a different conclusion. A recent study found that task performance involving two cross-modal objects diminished in the presence of a third modality if that task-irrelevant object was semantically congruent with one, but not both, of the two task-relevant objects, especially when the task-relevant objects were themselves incongruent (Misselhorn et al., 2016). A similar effect was observed in two other studies. When participants attended to one of two laterally presented visual streams of letters while performing a sequential matching task, their response times (RT) were significantly longer, when incongruent, task-irrelevant letter sounds were presented as well. The increase in RTs on these trials coincided with increased fMRI activation in the anterior cingulate cortex and over fronto-central EEG sensors (Zimmer et al., 2010a,b). These findings suggest that semantically incongruent stimuli induce a cognitive conflict between the components of a multi-modal stimulus and

subsequently, likely exogenously, recruit executive attentional resources to resolve the conflict, thus reducing the efficiency of multi-sensory integration of semantically congruent stimuli. Thus, this line of research suggested that the voluntary allocation of attentional resources in processing semantically congruent stimuli can be disrupted by endogenously unattended, task-irrelevant semantically incongruent stimuli.

These two lines of research imply different mechanisms for the interaction of semantic congruency and attentional selection. While the former suggests that congruent stimuli at an attended location boosts performance, the latter implies that incongruent and unattended stimuli recruit exogenous attention, and so divert resources from processing the congruent stimuli in the attentional focus, which reduces behavioral performance.

Here, we aimed to address these conflicting findings by investigating the interaction of attentional focus and semantic congruency in greater detail. We systematically varied the semantic congruency of three objects (a visual object, an auditory sound, and a visual background scene) in single- and cross-modal combinations, under different attentional foci and under conditions of explicit semantic congruence processing. Participants in three different groups directed their attention to two of three objects in the stimuli and made semantic congruency judgments for two attended stimuli. This allowed us to observe behavioral performance under conditions that replicated and extended critical features of the two lines of research yielding conflicting evidence. We were able to evaluate whether performance for attended congruent stimuli is increased or diminished in the presence of a distracting unattended and incongruent stimulus.

## MATERIALS AND METHODS

### Participants

Participants ( $n = 87$ , mean age 25.53 years, SD 3.71, and 43 male) were recruited from the student population of the University Hamburg and participated for a small payment. They all had normal hearing and normal or corrected-to-normal vision. The study was approved by the ethics committee of the German Psychological Society (JG072015) and was conducted in accordance with the principles of the Declaration of Helsinki on human subject research.

### Experimental Design and Stimuli

To investigate the interaction of semantic congruence and attentional focus on the processing of multi-sensory stimuli, we created 3-object-component stimuli consisting of a visual object in the foreground (V), and typical auditory object (A) associated with the visual object (V), and a visual background scene (B). Semantic congruence was designed as a within-subject factor and varied between stimulus components, giving rise to the following 5 experimental conditions<sup>1</sup> (see also **Table 1**): (1) none of the components are semantically congruent (coded as III), (2) V and

<sup>1</sup>The three letters in the condition codes always refer to 1, visual; 2, auditory; 3, background in that order.

**TABLE 1** | Experimental conditions and example objects comprising the cross-modal stimuli conditions.

Condition code	Visual	Auditory	Background
III	Door bell	Ambulance	Sky
CCI	Duck	Duck	Phone booth
CIC	Fire truck	Church bell	Burning house
ICC	Vacuum cleaner	Door bell	Door with bell
CCC	Lion	Roar	Savannah

Condition codes are listed in the order visual-auditory-background. C indicates semantic congruency between the respective components, I indicates incongruency with the other components.

A are congruent (coded as CCI, 1st and 2nd components are congruence), (3) V and B are congruent (coded as CIC, 1st, and 3rd component are congruent), (4) A and B are congruent (coded as ICC, 2nd and 3rd components are congruence, and (5) all components are congruent (coded as CCC).

Endogenous attention was manipulated as a between-subject factor in 3 groups: visual object and auditory sound (VA,  $n = 30$ , 16 males); visual object and background (VB,  $n = 27$ , 13 males); and auditory sound and background (AB,  $n = 30$ , 14 males). Participants in each group were instructed to focus their attention on the two object components of their group and judge these 2 components accordingly (see section “Experimental Task and Procedure” below).

Visual objects were pictures of animals and everyday items, auditory objects were typical sounds of these visual objects, and background scenes depicted typical contexts in which the visual or auditory object could be found (see **Figure 1** for an example). Incongruent combinations were created by randomly pairing an indoor object with an outdoor background (or sound) and vice versa. Upright pictures of the visual objects were scaled to a height of 250 pixels (px) [7.6 degrees of visual angle (dva)], horizontal pictures were scaled to a width of 510 px (16.13 dva, mean height 252.82 px, SD 52.41 px, mean width 315.48 px, and SD 99.98 px). The background pictures were scaled to  $768 \times 1024$  px ( $25.36 \times 33.4$  dva) and presented with a gray frame on a Samsung SyncMaster 2443DW screen. The sounds were presented via

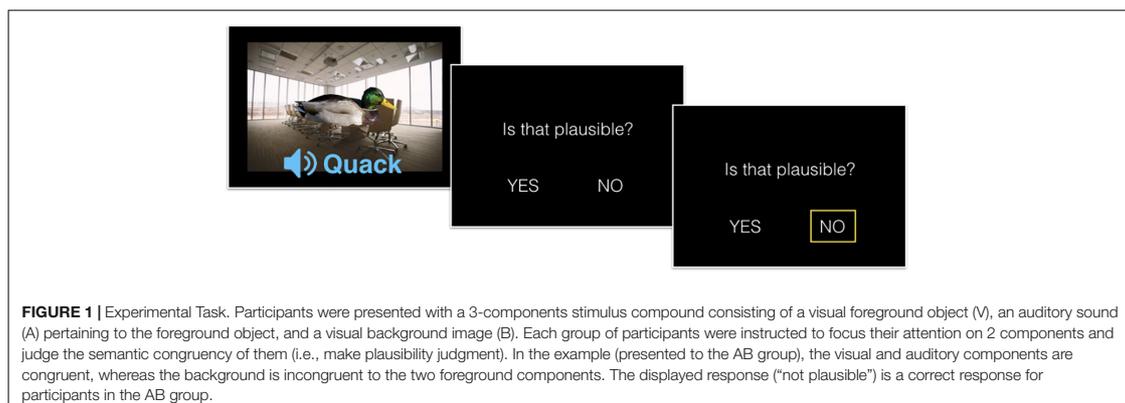
headphones with a volume of  $\sim 65$  DB. All stimulus aspects were presented simultaneously with the foreground picture centered on the background (see **Figure 1** for an example).

## Experimental Task and Procedure

After obtaining informed consent from the subject the experimenter instructed the participants about the goals of the study, the 3-component nature of the stimuli, and the attentional focus that they should maintain throughout the experiment. Participants were instructed to evaluate the congruency of the two components in their attentional focus, which was framed as a judgment of plausibility. Initial pilot data suggested that participants understood the term “plausible” better than “congruent.” In the main experiment it was explained to the participants that their plausibility judgment referred to the semantic congruence of the two components in question. They were told to respond as quickly and as accurately as possible with either the left and right arrow key representing a “YES” or a “NO” answer. The assignment of the response to the two response keys was counterbalanced across participants who responded with their index and middle finger of their dominant hand. Prior to the main experiment each participant completed a few training trials from the CCC and III condition until they responded correctly in 5 consecutive trials. In the main experiment, participants completed 150 trials (30 in each condition). Each stimulus was presented for a variable duration (depending on the duration of the sound clip (mean duration 1.53 s). Participants had to respond within 4 s. Failure to respond in this window resulted in a missing trial. Trials were separated by 1.4 s. The experiment lasted around 15 min.

## Data Preprocessing

Response accuracy and RT were collected as experimental data. All missing trials were removed from the data. Outliers were defined as 2 SD above the mean of the square-root transformed RT data and also removed from the experimental data. Finally, the effect of stimulus duration was removed using a regression approach: RT data were log-transformed and regressed onto the stimulus duration (general linear model with stimulus duration



**FIGURE 1** | Experimental Task. Participants were presented with a 3-components stimulus compound consisting of a visual foreground object (V), an auditory sound (A) pertaining to the foreground object, and a visual background image (B). Each group of participants were instructed to focus their attention on 2 components and judge the semantic congruency of them (i.e., make plausibility judgment). In the example (presented to the AB group), the visual and auditory components are congruent, whereas the background is incongruent to the two foreground components. The displayed response (“not plausible”) is a correct response for participants in the AB group.

and intercept as predictor variables). Duration-predicted RTs were removed by subtractions, and residuals from this regression were back projected into the original RT space and subjected to exponential transformation.

### Statistical Analysis

Response accuracy and RT were analyzed with linear mixed-effects models using the nlme package in R. Specifically, we set up omnibus mixed-effects repeated measures ANOVAs with the within-subject factor *Condition* (III, CCI, CIC, ICC, and CCC) and the between subject-factor *Focus* (VA, VB, and AB) using the formula:

$$DV \sim \text{condition} * \text{focus}, \text{random} = \sim 1 | \text{id}/\text{condition}$$

where DV is the dependent variable “percent error,” “RT (correct trials),” or “RT (incorrect trials).” *Post hoc* Bonferroni-adjusted contrasts were carried out using the multcomp package in R.

### Cognitive Computational Modeling

Hierarchical Bayesian parameter estimation with the drift diffusion model (DDM) yielded group and individual participant estimates the drift rate ( $v$ ), the boundary separation ( $a$ ), and the non-accumulation time ( $t$ ), also called the non-decision time (Ratcliff and McKoon, 2008). The DDM treats a binary decision as the result of an evidence accumulation process, in which the gathering of evidence for one or the other option is modeled as a Gaussian random walk that drifts at a certain rate toward one of two decision boundaries representing the two alternative options [see Ratcliff and McKoon (2008) for a schematic of the model]. Once one of these boundaries is crossed, a decision for this option is made. There are 4 primary free parameters in the DDM, whose optimized values are determined during model fitting: (1) the drift rate  $v$  governs the speed of evidence accumulation, corresponds to the slope of the random walk, and reflects choice difficulty, (2) the boundary separation  $a$  represents the distance between both decision boundaries and models how cautious a decision maker is with higher caution corresponding to a larger boundary separation, (3) the starting point  $z$  is the point between both decision boundaries at which the evidence accumulation starts. Although this parameter is unused in this study (i.e., is set to  $a/2$ ) it can model general biases toward one or the other option, (4) the non-decision time  $t$  captures all aspects of the RT that are not related to evidence accumulation, i.e., stimulus-encoding, feature selection, action-planning, and action-execution time.

Model fitting with the HDDM package in Python (Wiecki et al., 2013) offered a Hierarchical Bayesian workflow using Markov Chain Monte Carlo (MCMC) techniques. In most cases this yields more stable results than traditional Maximum Likelihood estimation and includes measures of estimation uncertainty in the form of posterior distributions of parameters. In addition, subject-specific parameter values are sampled from an overarching group distribution, which is updated using the data from all participants. This usually leads to more stable optimized parameter solutions, while also allowing for individual variability in these estimates.

The package offers a model parameterization depending on the experimental factors, e.g., one could model different drift rates for all conditions or for all groups or any combination of them. We compared these different model variants using the deviance information criterion (DIC), a model comparison index similar to the Bayesian information criterion (BIC), but applicable for Bayesian analysis using MCMC sampling. A difference in DIC scores of 15 and above is considered meaningful (Spiegelhalter et al., 2002).

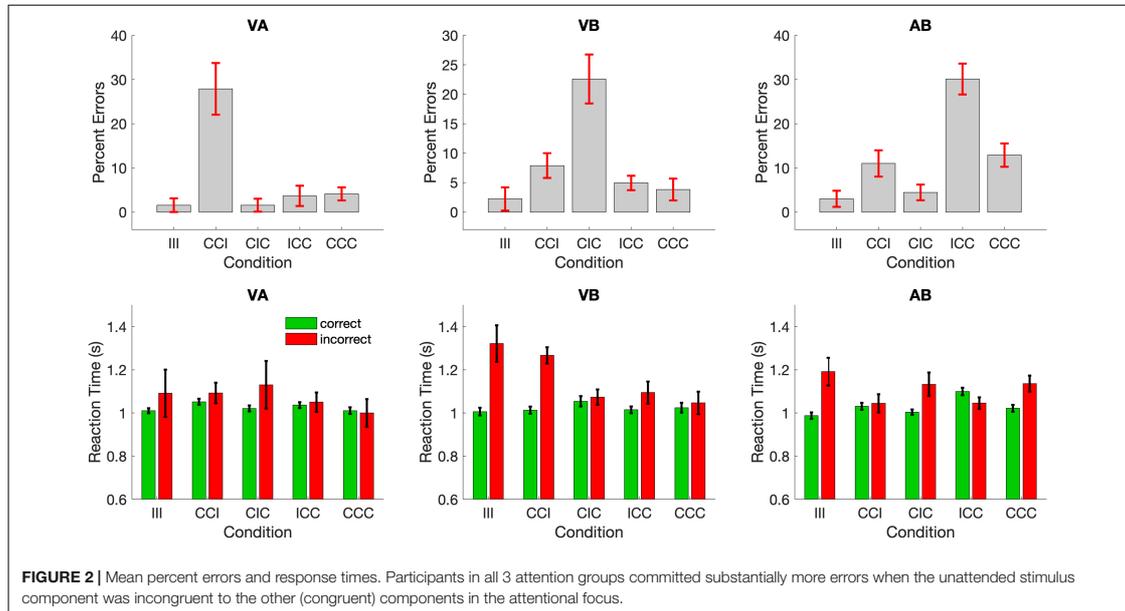
Each model variant was fit using the HDDM package (Wiecki et al., 2013) with 4 chains and 7000 samples following a burn-in phase of 500 samples to reduce the dependencies on initial values and to reach a steady state of the chain. Convergence was tested through visual inspection of the chains and by calculating the  $\hat{R}$  statistic (Gelman and Rubin, 1992), which compares within-chain and between-chain variance. The threshold for non-convergence was set at 1.05. We used the HDDM defaults as group-level priors, namely the drift rate was modeled as a group-level normal distribution  $[N(\mu, \sigma^2)]$ , whose parameters  $\mu_v$  and  $\sigma_v^2$  were modeled as  $N(2,3)$  and half-normal distribution  $HN(2)$  (2 being the variance parameter). The boundary separation was modeled as a Gamma (G) distribution, whose parameters  $\mu_a$  and  $\sigma_a^2$  were modeled as  $G(1.5, 0.75)$  and  $HN(0.1)$  distributions. Finally, the non-decision time was also modeled as a normal distribution, whose parameters  $\mu_t$  and  $\sigma_t^2$  were modeled as  $N(2,3)$  and  $HN(1)$  distributions. We compared parameter estimates for the different levels of each factor by mean of the group posterior distribution.

## RESULTS

### Analysis of Errors and Response Times

We first inspected the percent errors in all 3 groups of subjects with different attentional foci across all 5 stimulus conditions. To be counted as an error, the participant would have to (a) respond “not plausible” to two congruent components in the attentional focus (e.g., in group VA visual: lion, auditory: roar, background: swimming pool) or (b) respond “plausible” to two incongruent components in the attentional focus (e.g., in group VA visual: fire truck, auditory: church bell, background: burning house). Overall, participants only made few errors on the task (overall percentage of errors: group VA: 8.1% incorrect, 91.2% correct, 0.7% missing trials, group VB: 9.1% incorrect, 90.5% correct, 0.4% missing trials, group AB: 13.2% incorrect, 86.3% correct, and 0.5% missing trials). However, despite the overall low number of errors the different groups made substantially more errors in different, yet specific conditions in the task (see **Figure 2**): whenever the unattended component was incongruent to the two congruent components in the attentional focus (i.e., in group VA – CCI, in group VB – CIC, in group AB – ICC), the error rate was substantially higher, than in all other conditions.

A mixed effects ANOVA with the within-subject factor *Condition* and between-subject factor *Focus* confirmed a significant main effect of *Condition* ( $F_{4,336} = 10.48, p < 0.0001$ ) and a significant *Condition*  $\times$  *Focus* interaction effect ( $F_{8,336} = 14.42, p < 0.0001$ ). Subsequent, Bonferroni-adjusted



**FIGURE 2 |** Mean percent errors and response times. Participants in all 3 attention groups committed substantially more errors when the unattended stimulus component was incongruent to the other (congruent) components in the attentional focus.

contrasts between the different stimulus conditions revealed that the interaction effect was driven in each group by a significant difference between the critical condition (in group VA – CCI, in group VB – CIC, in group AB – ICC) and all conditions (all  $z$ -values  $> 3.9$ ,  $p < 0.001$ ).

Across all conditions a “NO” response (not plausible) was more frequently correct (for instance in conditions III, CIC, and ICC for the VA group) than a “YES” response (namely in condition CCI and CCC for the group VA). Thus, it is conceivable that participants learned about this subtle response bias and that they committed more errors in the critical conditions. The possibility of such a response bias is detectable, if the data are sorted according to the response itself instead of the response accuracy. If a response bias was present in the data, we would expect to see higher frequency of “NO” response across all conditions in all groups. **Figure 3** demonstrates that this is not the case. In fact, the pattern found in this analysis mirrors the finding from **Figure 2**: in the critical conditions there were a significant number of “NO” responses (i.e., and incorrect decision), whereas in the non-critical condition there were mostly “NO” and “YES” responses (correct responses depending on the condition). Importantly, this figure reveals that there was no overall bias toward “NO” responses.

In contrast, analyses of the RTs did not yield an equally systematic pattern of findings despite a significant effect for *Condition* ( $F_{4,331} = 7.11$ ,  $p < 0.001$ ) and for *Condition*  $\times$  *Focus* ( $F_{8,331} = 5.53$ ,  $p < 0.0001$ ) for RT in *correct trials* and a significant effect for *Condition* ( $F_{4,170} = 2.78$ ,  $p = 0.029$ ) and a trend-level *Condition*  $\times$  *Focus* interaction ( $F_{8,170} = 1.86$ ,  $p = 0.069$ ) for RT in *error trials*. Subsequent Bonferroni-adjusted *post hoc* contrasts revealed that for RTs in *correct trials* only, condition ICC in

group AB was significantly longer than all other conditions (all  $z$ -values  $> 4.1$ ,  $p < 0.001$ ). In addition, for RTs in error trials, conditions III, and CCI in group VB were significantly larger than all other conditions (all  $z$ -values  $> 3.16$ ,  $p < 0.05$ ). There were no additional RT effects in any of the other groups.

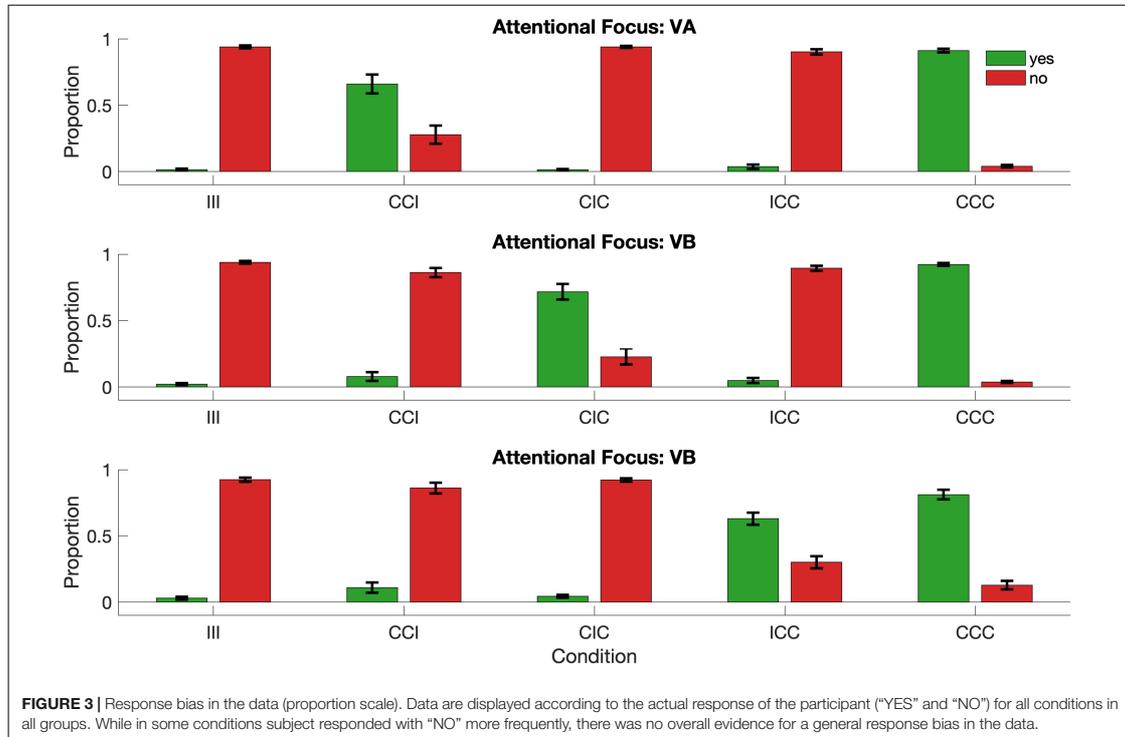
## CONCLUSION

In conclusion, participants made significantly more errors whenever the unattended stimulus was semantically incongruent with the two congruent stimuli in the focus of endogenous attention. However, a corresponding increase in RT on those error trials could not be found.

## Computational Cognitive Modeling

In the next step we applied cognitive computational modeling to these data to gain additional insights into the cognitive processes governing the responses in this task. The drift diffusion model (DDM) (Ratcliff and McKoon, 2008) is particularly well-suited for modeling the decisions in this task. A decision in the DDM is the result of an evidence accumulation process, which “drifts” at a specific rate to one of two decision boundaries representing the two decision options. In our case we defined the two options as “correct” and “incorrect” responses as this form of data coding has provided fruitful insights into the speed-accuracy trade-off present in most behavioral decision-making paradigms (Ratcliff and Rouder, 1998; Steinweg and Mast, 2017).

We compare variants of the DDM with different configurations of free parameters. Each of the 3 selected parameters (drift rate  $v$ , boundary separation  $a$ , and non-decision



time  $t$ ) could be modeled as a single parameter across all stimulus conditions or as a single parameter across all groups. In contrast, each parameter could be also modeled separately for each stimulus condition and for each attention group. We systematically compared all possible variants of the DDM using their DIC score (see **Figure 4**).

This model comparison analysis reveals that the model variants in which all three parameters are modeled separately for each condition provided the best model fit, but there are no meaningful differences between these regarding the group impact (DIC differences  $< 15$ ). Nevertheless, model 1 (the model with the lowest DIC score) also provided separate parameter distributions for each group, which allowed us to compare parameter distribution for each condition in each group. Given that the critical condition corresponded to different stimulus configurations in each group, model 1 thus provides the granularity to detect the effect of critical conditions in the parameter distributions. We show the group posterior distributions for all parameters in **Figure 5**.

Interestingly, the drift rate parameter (**Figure 5, top**) for the critical condition is always smaller than all other conditions in each group. This resembles the patterns of errors seen in the behavioral analysis above: whenever the unattended stimulus component was incongruent to the two other components in the attentional focus, we observed a reduced drift rate parameter. Similarly, the boundary separation parameter for the critical

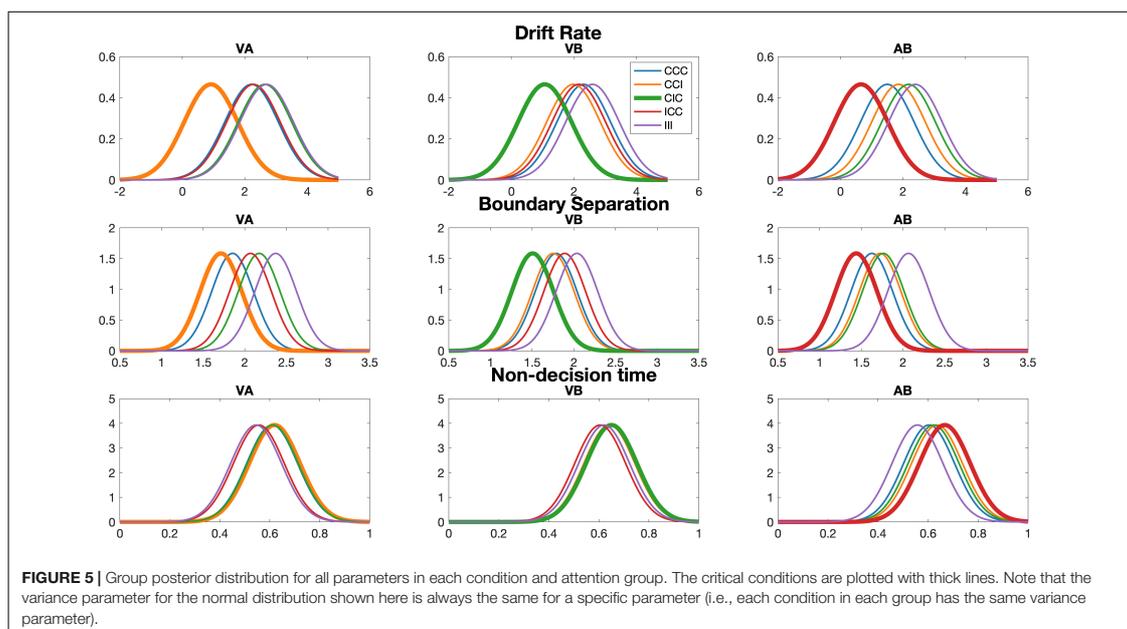
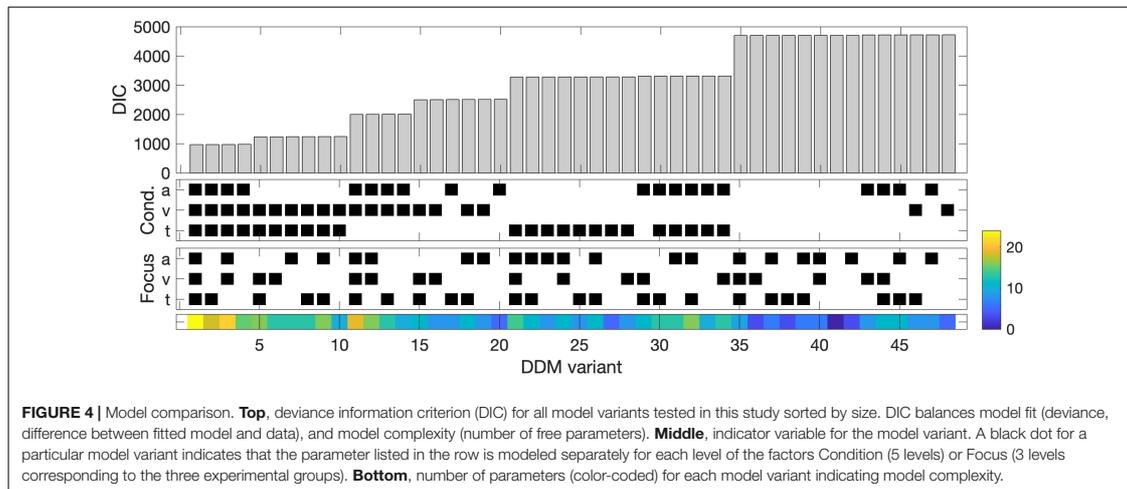
condition is also the smallest compared with all other conditions, but this pattern is less clear than for the drift rate. Finally, no such pattern of the critical conditions was observed for the non-decision time.

Having selected the best-fitting model from within a family of model variants does not insure that the model actually fits the data. This can be tested using posterior predictive checks (PPC), in which the model generates new data using the fitted parameters. These data are then compared to the original data. Below, we show the PPC findings for our selected Model 4, which simulated 500 new data points for the same number of subjects in each attention group. The response accuracy and correct and error RTs were then compared to the original data (see **Figure 2**). **Figure 6** shows the findings from this posterior predictive check.

The correct and incorrect responses of the PPC match the original data with high accuracy. However, the simulated RTs do not fit with the subtle differences in correct and incorrect RTs in the data. In fact, it seems that in the PCC simulations all conditions in all groups are modeled with essentially the same mean RT.

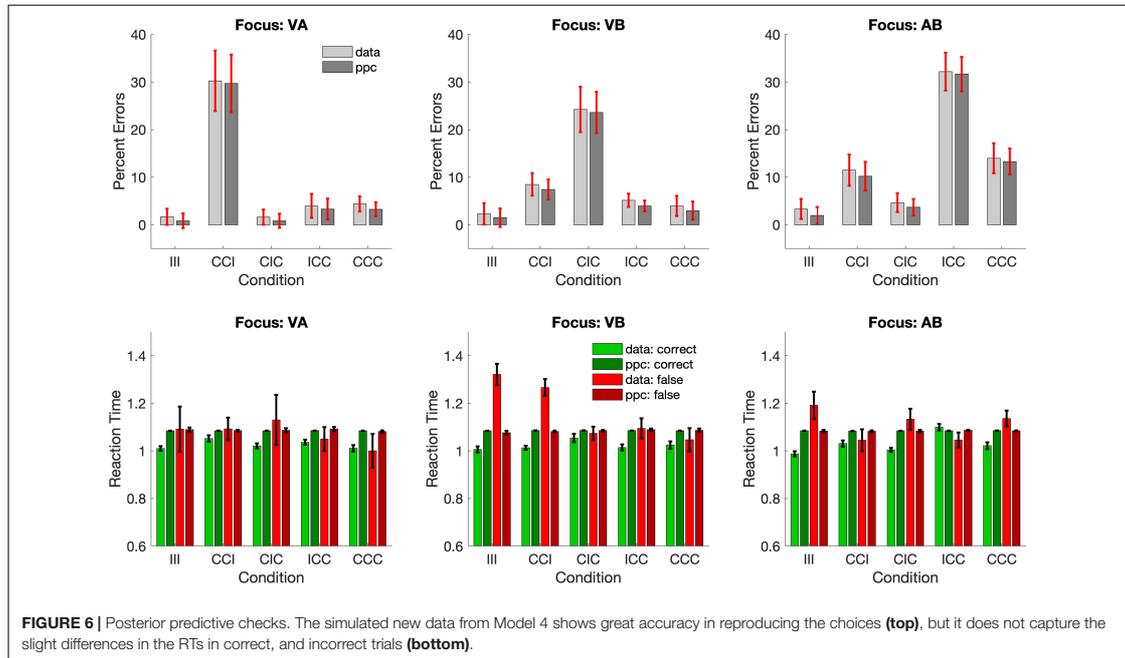
## DISCUSSION

We found a specific effect of attentional focus on the processing of our multi-sensory stimuli. Whenever the unattended



stimulus is incongruent with the two others in the attentional focus, participants made significantly more errors in semantic congruency judgments than in any other stimulus condition. This effect is paralleled by a significantly reduced drift rate parameter in these stimulus conditions as revealed in our drift diffusion modeling. RTs do not show a similar increase in RTs in error trials in these specific conditions. Rather, the pattern in error RTs seems to be driven by non-systematic increases in specific stimulus conditions, but unrelated to the attentional manipulation.

Our findings are in line with those studies demonstrating that semantically incongruent stimuli outside the focus of attention can capture those processing resources and disrupt the processing and evaluation of the attended stimuli (Zimmer et al., 2010a,b; Misselhorn et al., 2016). Indeed, it seems that in our data semantically incongruent stimuli induce a re-focusing of attention, such that the incongruency of the unattended stimulus is then considered leading to an incorrect (incongruent) judgment. If this was the case, then from the perspective of the participants, they would be making correct responses. This could



be the reason why the RTs between correct and incorrect trials in these critical conditions are almost identical (see **Figure 2** bottom, VA – CCI, VB – CIC, AB – ICC). Such an account would still be consistent with the neuroimaging findings from earlier studies demonstrating higher activations in anterior cingulate cortex (ACC) implying a processing of the conflict between semantically incongruent stimuli (Zimmer et al., 2010b). Other previous studies that also investigated semantic congruency in a multimodal context also observed higher ACC activations during the processing on incongruent stimuli (Weissman et al., 2004) reminiscent of the findings on conflict detection in the Stroop task (Fan et al., 2003). The brains in our subjects could be detecting the incongruity between the one of the previously attended congruent stimuli and the incongruent previously unattended, but now re-focused stimuli and yet still make an incongruent (but from their perspective correct) judgment. Of note, our primary finding of attentional capture of semantically incongruent stimuli occurs irrespective of the modality of the stimuli suggesting that we observed a general effect between attentional selection that is influenced by semantic (in) congruency.

Nevertheless, previous studies investigating semantic congruency with multi-modal stimuli also observed modulation of brain activity in the primary uni-sensory areas. In general, activation in primary sensory cortices is boosted if the modality is task-relevant (Weissman et al., 2004) and (semantically) congruent with other modality in stimulus compound (van Atteveldt et al., 2004), although an active encoding task might alleviate the advantage for congruent stimulus compounds (van Atteveldt et al., 2007). In fact, other studies have also

reported increased activity in higher activations for incongruent stimuli in primary sensory areas of the target modality (Weissman et al., 2004). These neural findings generally support the influence of endogenous and exogenous attention on the processing of multi-modal semantic congruency: as attention is directed toward a target modality (endogenous attention) the activation in those primary uni-sensory is increased, but if the target modality is incongruent with an unattended modality, processing resources are also recruited (exogenous attention) and activation in related brain regions is also increased. Our findings support the exogenous attention recruitment hypothesis: that is, participants committed significantly more errors whenever the unattended stimulus was incongruent to the two stimuli in the attentional focus, irrespective of the sensory modality of the stimuli. This points toward a general attentional bias for the processing of semantic incongruency.

In our study, the differences between the critical conditions mentioned above and the other conditions involving one incongruent stimulus is that in the critical conditions, the participants are initially primed to process a congruent stimulus combination because it is in the attentional focus. The incongruent stimulus then captures attentional resources leading to a refocusing of attention and prompting the participants to make more “incongruent” judgments, which are counted as “incorrect” here from the standpoint of an all-knowing observer, who knows what the participant should focus on. That is, in the critical condition the congruent stimulus pair comes first, whereas in the other conditions (e.g., for VA – CIC and ICC) the attention is already focused on an incongruent

stimulus pair, which is in most cases correctly detected through an “incongruent” judgment. This could be a potential reason for the lack of a systematic response time difference between the conditions: participants make the identical “incongruent” judgment, which could take approximately the same amount of processing time, but in the critical condition these responses are counted as incorrect.

Our drift diffusion modeling revealed that the observed increases in error rate in the critical conditions involving an unattended incongruent stimulus were paralleled by a significantly lower drift rate (Figure 5). The drift rate in diffusion models describes the speed of evidence accumulation until a decision is reached, when the diffusion process hits one of the two decision boundaries (Ratcliff and McKoon, 2008). In terms of cognitive processing, a lower drift rate in the presence of constant boundary separation means that participants take longer to accumulate evidence over the same boundary. This is commonly an indicator of difficulty induced by task condition or some other variable. The situation in our critical conditions would qualify as increased difficulty of evidence accumulation if exogenous attention engaged by incongruent, task-irrelevant stimuli interfered with endogenous attention. This implication has further evidence in that our model fitting of the DDM resulted in a significantly reduced boundary separation parameter in the critical condition, meaning that the representation of the two task options of congruent vs. incongruent was less stably separate, likely due to interference from exogenous attention to the incongruent distractor. The combination of lower drift diffusion rates and reduced boundary separation is consistent with our observation that RTs were not reduced in the critical conditions, and with the increased error rates in decisions in the critical conditions. Thus, the cognitive computational modeling revealed cognitive dynamics that a more conventional analysis of RTs would have missed.

Our RT data (Figure 2) also revealed a small number of significant RT differences between correct and incorrect trials in some conditions in the VB and AB groups. However, there appears to be no systematic pattern in these differences that can be related to the experimental manipulation. A potential reason for these non-systematic effects could be that the overall error rate in the experiment is quite low leaving only a few error trials for computing an average error RT. It is therefore likely that some of these high error RTs are driven by outlying data points that were not detected in our preprocessing steps.

The low number and unsystematic occurrence of error trials is also the likely reason that the classic DDM failed to replicate the observed differences in RT in the posterior predictive check (Figure 6), while at the same time reproducing the pattern of errors quite accurately. In fact, the synthetic data generated from the fitted parameters of the classic DDM exhibited no difference in mean RTs for any condition in any group, which could be interpreted that the observed RT differences are unsystematic and cannot be accurately modeled by the classic DDM. One way of accounting for different RT distributions of correct and incorrect responses is to add parameters that model inter-trial variability of drift rate, starting point and non-decision time. We did not include these parameters, because our main interest was on the

core DDM parameters such as drift rate and boundary separation, and estimating the latter can be compromised by adding the former (Boehm et al., 2018). However, this does not mean that the classic DDM is not suitable for modeling the data in our experiment. In fact, by tuning drift rate, boundary separation and non-decision time independently for each condition, the model is capable of reproducing the pattern of correct, and incorrect responses with a high degree of accuracy (Figure 6). This reinforces the interpretation from above that a lower drift rate in the critical condition indicates an increased processing demand due to the refocusing of the attentional focus to include the (formerly) unattended, incongruent stimulus.

Semantic congruency is a powerful amplifier of multi-sensory integration leading to higher brain activation (Doehrmann and Naumer, 2008) and better performance (Taylor et al., 2006; Steinweg and Mast, 2017). In addition, it can focus non-voluntary, “stimulus-driven” attention toward congruent stimuli and can boost perceptual processing resources at their location (Iordanescu et al., 2008, 2010). Moreover, semantic incongruency can disrupt the perceptual processing of stimuli in the attentional focus (Zimmer et al., 2010a,b). The findings of the present study are in line with these previous findings as we were able to show that semantic incongruency – independent of the stimulus modality – led to a re-focusing of attention to include the previous unattended (and incongruent) stimulus. It thus seems that our perceptual system is finely attuned to detect semantic incongruencies, even at a pre-attentive state. From a predictive coding perspective (Rao and Ballard, 1999; Friston, 2005), such incongruencies constitute prediction errors (violations of our expectations), which prompts the reallocation of processing resources via exogenous attention in implicitly attempting to resolve the incongruency of the percept. This would imply that the behavioral performance of our subjects in the critical condition is not erroneous, but rather adaptive to the needs for further cognitive processing independently of modality.

## DATA AVAILABILITY

The datasets generated for this study are available on request to the corresponding author.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the German Psychological Society. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

JG designed the research. JS-L collected the data. JG, JS-L, and MS analyzed the data. SG helped with computational modeling. All authors wrote the manuscript.

## FUNDING

JG and JZ were supported by the Collaborative Research Center TRR 169 “Crossmodal Learning” funded by the German Research Foundation (DFG)

and the National Science Foundation of China (NSFC). MS and JG were supported by a Computational Neuroscience grant from the United States National Science Foundation and the German Ministry of Education and Research (BMBF, 01GQ1603).

## REFERENCES

- Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., et al. (2018). Estimating across-trial variability parameters of the diffusion decision model: expert advice and recommendations. *J. Math. Psychol.* 87, 46–75. doi: 10.1016/j.jmp.2018.09.004
- Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755
- Doehrmann, O., and Naumer, M. J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain Res.* 1242, 136–150. doi: 10.1016/j.brainres.2008.03.071
- Fan, J., Flombaum, J. I., McCandliss, B. D., Thomas, K. M., and Posner, M. I. (2003). Cognitive and brain consequences of conflict. *Neuroimage* 18, 42–57. doi: 10.1006/nimg.2002.1319
- Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622
- Gelman, A., and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7, 457–472. doi: 10.1214/ss/1177011136
- Iordanescu, L., Grabowecy, M., Franconeri, S., Theeuwes, J., and Suzuki, S. (2010). Characteristic sounds make you look at target objects more quickly. *Atten. Percept. Psychophys.* 72, 1736–1741. doi: 10.3758/APP.72.7.1736
- Iordanescu, L., Guzman-Martinez, E., Grabowecy, M., and Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychon. Bull. Rev.* 15, 548–554. doi: 10.3758/PBR.15.3.548
- Koelewijn, T., Bronkhorst, A., and Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: a review of audiovisual studies. *ACTPSY* 134, 372–384. doi: 10.1016/j.actpsy.2010.03.010
- Mastroberardino, S., Santangelo, V., and Macaluso, E. (2015). Crossmodal semantic congruency can affect visuo-spatial processing and activity of the fronto-parietal attention networks. *Front. Integr. Neurosci.* 9:45. doi: 10.3389/fnint.2015.00045
- Matusz, P. J., Broadbent, H., Ferrari, J., Forrest, B., Merkley, R., and Scerif, G. (2015). Multi-modal distraction: insights from children's limited attention. *Cognition* 136, 156–165. doi: 10.1016/j.cognition.2014.11.031
- Meredith, M. A., and Stein, B. E. (1986a). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res.* 365, 350–354. doi: 10.1016/0006-8993(86)91648-3
- Meredith, M. A., and Stein, B. E. (1986b). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J. Neurophysiol.* 56, 640–662. doi: 10.1152/jn.1986.56.3.640
- Misselhorn, J., Daume, J., Engel, A. K., and Fries, U. (2016). A matter of attention: crossmodal congruence enhances and impairs performance in a novel trimodal matching paradigm. *Neuropsychologia* 88, 113–122. doi: 10.1016/j.neuropsychologia.2015.07.022
- Rao, R. P. N., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87. doi: 10.1038/4580
- Ratcliff, R., and McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* 20, 873–922. doi: 10.1162/neco.2008.12.06.420
- Ratcliff, R., and Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychol. Sci.* 9, 347–356. doi: 10.1111/1467-9280.00067
- Senkowski, D., Schneider, T. R., Foxe, J. J., and Engel, A. K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci.* 31, 401–409. doi: 10.1016/j.tins.2008.05.002
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. B* 64, 583–639. doi: 10.1111/1467-9868.00353
- Steinweg, B., and Mast, F. W. (2017). Semantic incongruity influences response caution in audio-visual integration. *Exp. Brain Res.* 235, 349–363. doi: 10.1007/s00221-016-4796-0
- Talsma, D., Senkowski, D., Soto-Faraco, S., and Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci.* 14, 1–11. doi: 10.1016/j.tics.2010.06.008
- Taylor, K. I., Moss, H. E., Stamatakis, E. A., and Tyler, L. K. (2006). Binding crossmodal object features in perirhinal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8239–8244. doi: 10.1073/pnas.0509704103
- van Atteveldt, N., Formisano, E., Goebel, R., and Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron* 43, 271–282. doi: 10.1016/j.neuron.2004.06.025
- van Atteveldt, N. M., Formisano, E., Goebel, R., and Blomert, L. (2007). Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage* 36, 1345–1360. doi: 10.1016/j.neuroimage.2007.03.065
- Weissman, D. H., Warner, L. M., and Woldorff, M. G. (2004). The neural mechanisms for minimizing cross-modal distraction. *J. Neurosci.* 24, 10941–10949. doi: 10.1523/JNEUROSCI.3669-04.2004
- Wiecki, T. V., Sofer, I., and Frank, M. J. (2013). HDDM: hierarchical bayesian estimation of the drift-diffusion model in python. *Front. Neuroinform.* 7:14. doi: 10.3389/fninf.2013.00014
- Zimmer, U., Itthipanyanan, S., Grent't-Jong, T., and Woldorff, M. G. (2010a). The electrophysiological time course of the interaction of stimulus conflict and the multisensory spread of attention. *Eur. J. Neurosci.* 31, 1744–1754. doi: 10.1111/j.1460-9568.2010.07229.x
- Zimmer, U., Roberts, K. C., Harshbarger, T. B., and Woldorff, M. G. (2010b). Multisensory conflict modulates the spread of visual attention across a multisensory object. *Neuroimage* 52, 606–616. doi: 10.1016/j.neuroimage.2010.04.245

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Spilcke-Liss, Zhu, Gluth, Spezio and Gläscher. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## **2. Zusammenfassende Darstellung der Publikation**

### **2.1 Einleitung**

Im Alltag sind wir mit vielen Reizen konfrontiert, die uns über unsere verschiedenen Sinne erreichen und die dabei oft mehrere Sinne gleichzeitig ansprechen. Sehen wir beispielsweise einen Hund in unserer nahen Umgebung, können wir ihn auch bellen hören, sein Fell anfassen und seinen Geruch einatmen. Der Eindruck des Hundes ist somit multisensorisch, die kognitive Reizverarbeitung läuft multimodal bzw. crossmodal. Gleichzeitig wirken viele weitere Reize auf uns ein. Wir sehen die Umgebung mit den Blumen und den Bäumen, hören die Kinder, die auf dem naheliegenden Spielplatz toben und riechen das frisch gemähte Gras, auf dem der Hund spielt. Die Herausforderung, dabei zusammengehörige Reize gemeinsam zu verarbeiten, ist eine der herausragendsten Leistungen unserer Wahrnehmung. Wenn einzelne Sinneseindrücke zu einem zusammenhängenden Bild zusammengefügt werden, spricht man von crossmodaler oder multimodaler Integration. Diese beeinflusst unser Verhalten und kann unsere Verhaltensleistung in vielen Fragestellungen verbessern (vgl. hierzu Spilcke-Liss, 2015).

Welche Reize dabei zusammengehören, werden durch zeitliche, räumliche und semantische Faktoren bestimmt. Der Erfolg der multimodalen Integration hängt dabei nicht nur von einer zeitlichen und örtlichen Nähe, sondern auch von einer Übereinstimmung der inhaltlichen Bedeutung ab (vgl. Stein and Meredith, 1993; Taylor et al., 2006; Doehrmann and Naumer, 2008; Steinweg and Mast, 2017).

#### **2.1.1 erste Pilotstudie**

Bei der Planung dieser Arbeit stand die inhaltliche Komponente im Vordergrund. In ersten Verhaltensexperimenten untersuchten wir den Einfluss der inhaltlichen Übereinstimmung gemeinsam mit dem Einfluss der räumlichen Komponente. In unserer ersten Pilotstudie präsentierten wir Probanden auf einer Bildschirmhälfte ein Bild und zeitgleich über einen Kopfhörer entweder auf dem linken oder dem rechten Ohr einen inhaltlich zum Bild passenden oder nicht passenden Ton. Dabei präsentierten wir Bild und Ton entweder seitengleich oder auf unterschiedlichen Seiten. Die Probanden sollten die dargebotenen Stimuli nach der Lebhaftigkeit als Objekt oder Lebewesen einteilen. Bei der Auswertung zeigten sich, im Gegensatz zu den Ergebnissen aus anderen Studien (Macaluso and Driver, 2005; Calvert, 2001; Meredith and Stein, 1986), keine signifikanten Effekte bezüglich des räumlichen Einflusses. Dies führten wir darauf zurück, dass der Abstand zwischen der linken und rechten Bildschirmhälfte zu klein war, um ausreichende Unterschiede zu bewirken. Daher konzentrierten wir uns auf die Untersuchung des inhaltlichen Einflusses.

#### **2.1.2 zweite Pilotstudie**

Inwiefern sich inhaltlicher Einfluss im Sinne semantischer Kongruenz auf crossmodale Integration auswirkt, wurde weitgehend mit bimodalen Reizen untersucht. So wurden mit audiovisuellen Stimuli beispielsweise die Bearbeitung von Aufgaben zur Farberkennung (Laurienti et al., 2004) und Einordnung als „lebendig“ oder „nicht lebendig“ (Steinweg and Mast, 2016) bzw. Musikinstrument oder Werkzeug (Noppeney et al., 2010) analysiert. Dabei ergab eine inhaltlich übereinstimmende (semantisch kongruente) Präsentation z.B. kürzere Antwortzeiten als eine unimodale Präsentation, eine unstimmgige Kombination (semantisch inkongruent) hingegen verlängerte die Antwortzeit.

Für eine ausführliche Zusammenstellung zur Rolle von semantischer Kongruenz bei audiovisueller Stimulation vgl. Spilcke-Liss, 2015.

Nun besteht unser Alltag nicht aus rein bimodalen Reizen und wir meistern ständig die Aufgabe, multisensorische Reize zu identifizieren und aus ihnen relevante Informationen zu filtern. Diese Reize stehen meist in einem Kontext, zu dem sie passen können oder nicht. Wir stellten uns daher die Frage, welche Auswirkung inhaltliche Kongruenz bzw. Inkongruenz auf einen im Vordergrund dargestellten audiovisuellen Stimulus hat und wie dieser von einem kongruenten bzw. inkongruenten visuellen Hintergrund beeinflusst wird.

So veränderten wir unser Studiendesign auf audiovisuelle Stimuli aus drei Bestandteilen (visueller Vordergrund, auditiver Ton, visueller Hintergrund) und ließen die unterschiedlichen Kombinationen auf Plausibilität bewerten. Dabei erfolgten die Kombinationen in fünf unterschiedlichen Bedingungen: Vordergrund, Ton und Hintergrund kongruent (CCC), Vordergrund und Ton kongruent mit inkongruentem Hintergrund (CCI), Vordergrund und Hintergrund kongruent mit inkongruentem Ton (CIC), Ton und Hintergrund kongruent mit inkongruentem Vordergrund (ICC) und Vordergrund, Ton und Hintergrund inkongruent (III).

In der Auswertung zeigte sich insbesondere, dass die Kombination eines visuellen Vordergrunds mit kongruentem Ton und inkongruentem visuellen Hintergrund (CCI) signifikant häufiger als nicht plausibel bewertet wurde (vgl. Abbildung 1).

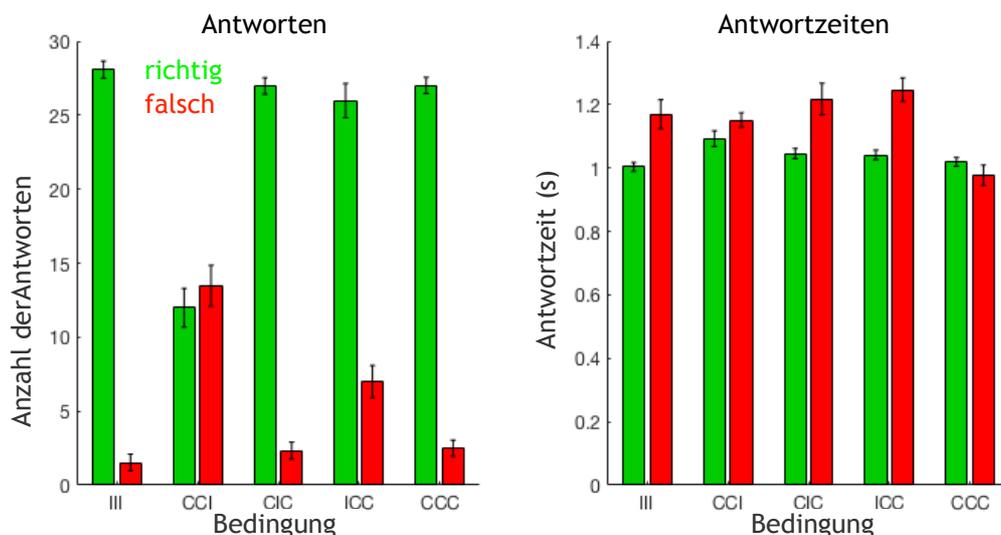


Abb. 1: Fehlerrate und Antwortzeiten Pilotstudie ohne gerichtete Aufmerksamkeit

Um diesem Effekt weiter zu nachzugehen und den Einfluss semantischer Kongruenz bzw. Inkongruenz bei der Verarbeitung multimodaler Reize genauer zu untersuchen, konkretisierten wir unsere Fragestellung durch Lenkung der Aufmerksamkeit auf je zwei der drei dargestellten Stimuluskomponenten.

### 2.1.3 Stimulus-getriebene und zielgerichtete Aufmerksamkeit

Verschiedene Studien legen nahe, dass Aufmerksamkeit eine wesentliche Voraussetzung für den Nutzen von semantischer Kongruenz bei multimodaler Verarbeitung ist. So ist die Worterkennung bei audiovisueller Präsentation mittels vorgelesener Wörter und gezeigter Lippenbewegung bei bimodal gerichteter

Aufmerksamkeit und inhaltlich kongruenter Darstellung beschleunigt. Wird der passende Ton zur visuellen Präsentation allerdings nicht gezielt beachtet, konnte kein Vorteil der kongruenten Darstellung nachgewiesen werden (Morís Fernández et al., 2015).

In der Lehre der selektiven Aufmerksamkeit unterscheidet man typischerweise Stimulus-getriebene und zielgerichtete Aufmerksamkeit. Letztere nutzt dabei kognitive Fähigkeiten und lenkt diese gezielt auf die Aufgabenbearbeitung. Sie wird auch als endogen charakterisiert. Wird unsere Aufmerksamkeit unterbewusst durch unerwartete Veränderungen in Anspruch genommen, spricht man von Stimulus-getriebener Lenkung. Diese von außen beeinflusste Aufmerksamkeit wird auch als exogen bezeichnet. (Corbetta and Shulman, 2002; Styles, 2006). Sind wir beispielsweise im Supermarkt und schauen uns die Apfelauslage an, weil wir einen Apfelkuchen backen wollen, machen wir dies zielgerichtet, unsere Aufmerksamkeit ist endogen auf das Apfelregal gelenkt. Sind wir hingegen aus anderen Gründen im Supermarkt und uns rollt aus dem Obstregal ein Apfel entgegen, wird unsere Aufmerksamkeit Stimulus-getrieben, exogen auf die Apfelauslage gelenkt.

Ähnlich werden zwei Wege bei der Informationsverarbeitung unterschieden, die in verschiedenen Arbeitsgebieten wie der Softwareentwicklung, Modellierung und auch den Neurowissenschaften Anwendung finden: den „top down“- und den „bottom up“-Ansatz. Startet ein Prozess im Kleinen und baut daraus den übergeordneten Kontext auf, spricht man von „bottom-up“. Wird er hingegen mit einer bestimmten Zielsetzung vom großen Ganzen herab ins Detail geleitet, nennt man dies „top-down“. Bezogen auf die kognitionswissenschaftlich untersuchte, kontrollierte Aufmerksamkeit wie bei der Reizdifferenzierung oder gezielten Aufgabenbearbeitung, wird diese zielgerichtete Aufmerksamkeit im Sinne einer „top-down“-Kontrolle bewusst auf Einzelheiten gelenkt. Mittels funktioneller Magnetresonanztomographie (fMRT) konnte gezeigt werden, dass dabei ein dorsales Netzwerk aus dorsalen, posterior-parietalen und frontalen Hirnregionen aktiv ist. Ist die Aufmerksamkeit hingegen nicht gezielt gelenkt, sondern wird durch einen sich in den Vordergrund stellenden Stimulus von der Grundlage des Dargestellten abgelenkt, wird die „bottom up“-Kontrolle ausgeübt. Hierbei spielen in der Bildgebung insbesondere das rechtsseitige temporoparietale Kreuzungsareal sowie der ventrale frontale Kortex eine Rolle. Dieses exogen geleitete, ventrale Netzwerk kann durch die Aufmerksamkeitslenkung laufende endogene Prozesse unterbrechen. Dabei überwiegen aus der gewohnten Umgebung herausstechende Reizimpulse der kognitiven Steuerung und unsere Aufmerksamkeit wird auf die auffallenden Stimuli gelenkt. (Corbetta and Shulman, 2002; Styles, 2006; Tang et al., 2016; Talsma et al., 2010)

#### **2.1.4. Aufmerksamkeit und multimodale Integration**

In der Literatur herrscht großer Diskurs über den Zusammenhang zwischen der multimodalen Reizverarbeitung als fraglich automatisch ablaufendem „bottom-up“-Prozess und der Aufmerksamkeitslenkung dagegen als „top-down“-Kontrolle (Tang et al., 2016; Morís Fernández et al., 2015; De Meo et al., 2015; Koelewijn et al., 2010; Talsma et al., 2010). Ein „top-down“-Einfluss wird dabei jedoch auch für die Verarbeitung eines multimodalen Stimulus beschrieben: die Verknüpfung der einzelnen Stimulusmodalitäten als zusammengehörigen Reiz hängt von erlernten Verbindungen ab, die bewusst zusammengeführt werden müssen. Damit ist die multimodale Integration nicht rein automatisch, sondern wird auch objektbezogen

und kongruenzabhängig gelenkt (Tang et al., 2016; ten Oever et al., 2016). Sie läuft parallel mit der Aufmerksamkeitslenkung ab und Integration und Aufmerksamkeit beeinflussen sich gegenseitig (Koelewijn et al., 2010; Talsma et al., 2010). Dabei wird vorsichtig ein überwiegender Effekt der Stimulus-getriebenen Lenkung beschrieben. So können multisensorische Prozesse „bottom-up“ ablaufen und die Wahrnehmung größtenteils unabhängig von „top-down“-Kontrolle beeinflussen (De Meo et al., 2015).

Bei der Nutzung von Aufmerksamkeitslenkung zur Leistungsverbesserung bei multimodaler Stimulusverarbeitung spielt inhaltliche Übereinstimmung eine wichtige Rolle. So war die Leistung von Probanden in einer visuellen Suchaufgabe signifikant schneller, wenn sie zeitgleich einen semantisch kongruenten Ton hörten (Iordanescu et al. 2008, 2010). Ebenso vereinfachte semantische Kongruenz den Wahrnehmungsprozess von einem nicht im Bezug stehendem Element am gleichen Ort wie ein kongruenter multisensorischer Prime (Mastroberardino et al., 2015). Die Leistungsverbesserung wird dabei durch die verstärkte, örtliche Aufmerksamkeit hervorgerufen, die durch semantische Kongruenz an dem Ort des kongruenten Stimulus erhöht wurde. Dabei war in diesen Studien die Einbindung der Aufmerksamkeit durch semantisches Priming Stimulus-getrieben und nicht zielgerichtet gelenkt.

Einen anderen Ansatz untersuchte den Einfluss irrelevanter Reize auf die Verarbeitung multimodaler Reize, die im zielgerichteten Aufmerksamkeitsfokus stehen.

Dazu wurde eine Untersuchung mit auditiver, visueller und taktiler Stimulation durchgeführt, wobei die Probanden die Kongruenz der Intensitätsveränderung zweier im Aufmerksamkeitsfokus stehender Stimulusmodalitäten bewerten sollten ohne die dritte Modalität zu beachten. Die Studie ergab, dass die Leistung bei der Aufgabenverarbeitung von im zielgerichteten Aufmerksamkeitsfokus stehender bimodaler Stimuli durch die Anwesenheit von einer dritten Modalität verschlechtert wird, wenn letztere - für die Aufgabe irrelevante Stimulation - semantisch kongruent zu einer, jedoch nicht beiden aufgabenrelevanten Modalitäten war. Dieser Effekt war besonders deutlich, wenn die beachteten, aufgabenrelevanten Stimuli inkongruent zueinander waren (Misselhorn et al., 2016). Ein ähnlicher Stör-Effekt wurde in einer Vergleichsaufgabe nachgewiesen, in der Probanden angewiesen waren, ihre Aufmerksamkeit auf eine von zwei am Bildschirmrand präsentierten Buchstabenfolgen zu richten und aufeinanderfolgende gleiche Buchstaben zu erkennen. Bei gleichzeitiger Präsentation von für die Aufgabe irrelevanten, inkongruenten ausgesprochenen Buchstaben verlängerten sich die Antwortzeiten signifikant (Zimmer et al., 2010a,b). So führt eine semantisch inkongruente Präsentation zu einem mentalen Konflikt zwischen den Komponenten des multimodalen Stimulus und nimmt exogenen Aufmerksamkeitsressourcen ein, um den Konflikt zu lösen. Dadurch wird die Effizienz der multimodalen Integration abgeschwächt (Misselhorn et al., 2016; Zimmer et al., 2010a,b).

Lenkt man die Aufmerksamkeit selektiv auf eine Modalität, schwächt sich ebenfalls die multimodale Performanceverbesserung ab, die bei geteilter Aufmerksamkeit auftritt. Der Nachteil durch crossmodale Ablenkung wird durch diese unimodale Aufmerksamkeitslenkung jedoch nicht signifikant reduziert (Mozolic et al., 2008). Die zielgerichtete Aufmerksamkeit beeinflusst also die multimodale Integration ebenso wie die Stimulus-getriebene.

Zusammengefasst werden zwei Zusammenhänge von semantischer Kongruenz und Aufmerksamkeitslenkung in Bezug auf unser Verhalten beschrieben: Einerseits verbessert semantische Kongruenz bei beachtetem Stimulus die Leistung, andererseits führen Inkongruenz und unbeachtete Störstimuli zu einer Leistungsver schlechterung.

Wie sich die Interaktion von gelenktem Aufmerksamkeitsfokus und semantischer Kongruenz bei der Verarbeitung audiovisueller Stimuli im visuellen Kontext auswirkt, untersuchten wir in dem im vorliegenden Paper beschriebenen Verhaltensexperiment. Ziel war es, zu evaluieren, ob ein nicht im Aufmerksamkeitsfokus stehender inhaltlich kongruenter bzw. inkongruenter Störstimulus positiven oder negativen Einfluss auf die Verarbeitung eines beachteten kongruenten bzw. inkongruenten Stimulus ausübt.

## 2.2 Studiendesign

Die Studie wurde durchgeführt mit 87 Studenten der Universität Hamburg (Durchschnittsalter 25,53 Jahre (SD 3,71), 43 Männer (49%)) mit gesundem bzw. korrigiertem Hör- und Sehvermögen.

### 2.2.1 Stimulusdesign und Aufgabenstellung

Wir entwarfen Stimuli aus folgenden drei Komponenten: visueller Vordergrund (V, im Beispiel aus Abbildung 2 eine Ente), auditiver Ton (A, im Beispiel aus Abbildung 2 Entengeschnatter, dargestellt durch „quak“) und visueller Hintergrund (B, im Beispiel aus Abbildung 2 eine Telefonzelle). Die verwendeten Bilder stellten Alltagsgegenstände oder Lebewesen dar und die Töne die dazu passenden Geräusche (vgl. Abbildung 2).



Abb. 2: Stimulusbeispiel

Den Kongruenz-Faktor konstruierten wir als *within-subject* Faktor und variierten diesen in den unterschiedlichen Komponenten des Stimulus, wodurch wir die bereits oben erwähnten und in Tabelle 1 aufgeführten fünf Bedingungen aufstellten. Semantische Inkongruenz erreichten wir durch zufällige Paarung von typischerweise drinnen bzw. draußen vorzufindenden Objekten/Lebewesen.

**Tabelle 1 Bedingungen und Beispiele für Stimuluskombinationen**

Bedingung	Vordergrund	Audio	Hintergrund
III	Türklingel	Krankenwagen	Himmel
CCI	Ente	Entengeschnatter	Telefonzelle
CIC	Feuerwehrauto	Glockenläuten	brennendes Haus
ICC	Staubsauger	Türklingel	Tür mit Klingel
CCC	Löwe	Löwengebrüll	Savanne

*Die Kodierung der Bedingungen sind mit C für kongruent und I für inkongruent in der Reihenfolge Vordergrund-Audio-Hintergrund erstellt.*

Wir ließen die Teilnehmer die Aufmerksamkeit auf zwei der drei Stimuluskomponenten richten und variierten sie als *between-subject* Faktor in drei Gruppen: eine Gruppe sollte sich auf Vordergrund und Ton (VA, n = 30, 16 (53%) männlich) konzentrieren, eine auf Vordergrund und Hintergrund (VB, n = 27, 13 (48%) männlich) und eine weitere auf Ton und Hintergrund (AB, n = 30, 14 (46%) männlich).

Die Probanden wurden angewiesen, die beachteten Stimuluskomponenten nach Kongruenz zu bewerten und diese mit Klicken der linken bzw. rechten Pfeiltaste einer vor ihnen liegenden Tastatur als passend bzw. nicht passend einzuordnen. So lautete zum Beispiel die richtige Bewertung des in Abbildung 2 dargestellten Stimulus je nach Aufmerksamkeitsfokus für die AV-Gruppe (Ente und Geschnatter) „ja, passend“ und für die VB-Gruppe (Ente und Telefonzelle) „nein, nicht passend. Die Bedeutung der Pfeilrichtung war unter den Probanden ausgeglichen zwischen links passend bzw. rechts passend. Die Probanden waren angewiesen, mit Zeige- und Mittelfinger ihrer dominanten Hand zu arbeiten sowie die Bewertung so schnell und so genau wie möglich vorzunehmen. Direkt vor dem Experiment durchliefen alle ein Training mit einigen Trials in den Bedingungen CCC und III, bis sie aufeinanderfolgend fünf richtige Bewertungen je Bedingung absolviert hatten. Das Experiment bestand aus 150 Trials (30 je Bedingung), wobei die Dauer eines Trials variierte (abhängig von der Länge des Soundclips ~2 Sek). War nach 4 Sekunden keine Bewertung erfolgt, wurde das Trial als nicht beantwortet gewertet. Zwischen den Trials lagen 1,4 Sekunden, insgesamt dauerte das Experiment etwa 15 Minuten.

### 2.2.2. Datenverarbeitung

Zur Auswertung wurden die Antwortzeiten sowie die Richtigkeit der Antworten genutzt. Die nicht beantworteten Trials wurden aussortiert ebenso wie Ausreißer (definiert als 2 SD über dem Mittelwert nach Wurzeltransformation der Antwortzeiten). Der Effekt der Stimulusdauer wurde herausgerechnet mittels eines generalisierten linearen Modells angewendet auf die logarithmustransformierten Antwortzeiten.

### 2.2.3 Modellierung

Zur weiteren Einordnung der Daten führten wir eine Modellierung mittels *hierarchical bayesian drift diffusion model* durch. Nach dem *drift diffusion model* (DDM) werden binäre Entscheidungsprozesse gefällt, indem Hinweise für die eine oder andere Entscheidungsoption gesammelt werden, bis ein Schwellenwert erreicht und damit eine Entscheidung getroffen wird. Dabei wird die Sammlung von Hinweisen durch einen Gausschen random walk modelliert (vgl. Ratcliff and McKoon, 2008).

Je eindeutiger die Hinweise sind, desto schneller kann die Entscheidung getroffen werden. Diese Geschwindigkeit wird mit der *drift rate*  $v$  angegeben. Soll eine Entscheidung sehr sicher sein, müssen viele Hinweise gesammelt werden, was durch einen großen Schwellenwert, im Modell die *boundary separation*  $a$ , wiedergegeben wird.

Ein DDM ist damit definiert durch folgende vier freie Parameter:

- (1) die *drift rate*  $v$ , die die Geschwindigkeit der Hinweissammlung/-häufung angibt und damit die Steigung des Zufallswegs bestimmt und die Schwierigkeit des Entscheidungsprozesses widerspiegelt

- (2) die *boundary separation*  $a$ , die den Abstand zwischen den beiden Entscheidungsmöglichkeiten und somit die für den Entscheidungsträger geforderte Sicherheit angibt
- (3) der Startpunkt  $z$  zwischen den beiden Entscheidungsmöglichkeiten, der bei vorbestehender Tendenz zu einer Entscheidung vom Mittelpunkt  $a/2$  abweichen kann
- (4) eine *non-decision time*  $t$ , die alle Anteile der Antwortzeit beinhaltet, die nicht direkt zum Entscheidungsprozess gehören (Stimuluserkennung, Handlungsplanung u.a.).

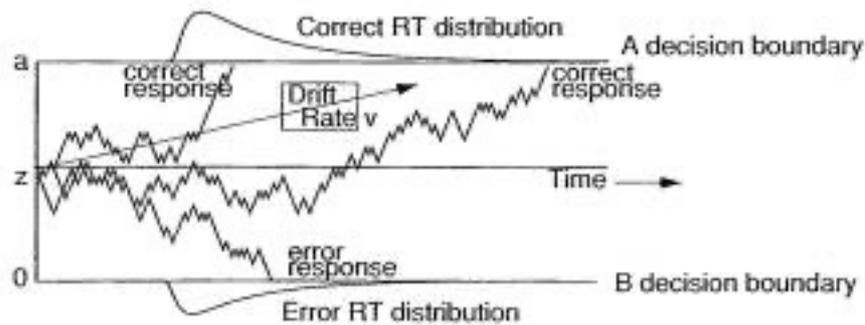


Abb. 3 Drift Diffusion Model aus Ratcliff, 2008

Das *HDDM package* in *Python* bietet eine Modellierung mittels hierarchischer Bayes Verfahren sowie Markov Chain Monte Carlo Methoden (Wiecki et al, 2013) und führt somit zu stabilen Ergebnissen bei gleichzeitig möglicher individueller Schwankung der Schätzungen.

Wir kodierten zur Modellberechnung unsere Antwortdaten abhängig von der Korrektheit, was zu guten Ergebnissen bei typischerweise in Verhaltensexperimenten auftretendem *speed-accuracy trade off* führt (Ratcliff and Rouder, 1998; Steinweg and Mast, 2017).

Die Parameterschätzung erfolgte abhängig von aus den Versuchsergebnissen gewählten Faktoren. So ließen wir mit dem *HDDM package* in *Python* diverse Modelle berechnen mit unterschiedlichen Parametern für die jeweiligen Stimulusbedingungen bzw. für die unterschiedlichen Gruppen bzw. für verschiedene Kombinationen dieser. Wir bewerteten die Modelle nach dem *deviance information criterion* (DIC). Dabei werden die Unterschiede der generierten Daten im Vergleich zu den erhobenen Daten unter Einbeziehung der Komplexität des Modells bewertet. Somit wird die Genauigkeit des Modells gegen die geforderte Rechenleistung abgewogen. Dabei gilt, je kleiner der Wert, desto besser das Modell. Als aussagekräftig gilt ein DIC-Unterschied von 15 (Spiegelhalter et al., 2002).

## 2.3 Studienergebnisse

### 2.3.1 Fehlerrate

In den Versuchsergebnissen untersuchten wir zuerst die Fehlerrate. Wir werteten eine Entscheidung als Fehler, wenn (a) bei inhaltlich übereinstimmenden im Aufmerksamkeitsfokus stehenden Stimuluskomponenten die Antwort „nicht passend“ lautete oder (b) inhaltlich nicht übereinstimmende im Aufmerksamkeitsfokus stehende Stimuluskomponenten als „passend“ bewertet wurden.

Insgesamt zeigte sich eine sehr geringe Fehlerrate: In der Gruppe mit Fokus auf Vordergrund und Ton (VA) lag sie bei 8.1% (91.2% korrekt, 0.7% missing trials), in der Gruppe mit Fokus auf Vorder- und Hintergrund (VB) bei 9.1% (90.5% korrekt, 0.4% missing trials) und in der Gruppe, die sich auf Ton und Hintergrund (AB) fokussierte bei 13.2% (86.3% korrekt, 0.5% missing trials).

Betrachtet man die Fehlerraten in den einzelnen Bedingungen, fallen in allen Gruppen erhöhte Fehlerraten in den Bedingungen auf, in denen die nicht beachtete Stimuluskomponente inkongruent zu kongruenten im Aufmerksamkeitsfokus stehenden Komponenten präsentiert wurde (vgl. Abbildung 4). Diese kritischen Bedingungen sind in der Gruppe VA Bedingung CCI, in Gruppe VB Bedingung CIC und in Gruppe AB Bedingung ICC.

Die statistische Auswertung ergab einen signifikanten Effekt der Bedingung sowie einen signifikanten Interaktionseffekt von Bedingung und Fokus, wobei letzterer in jeder Gruppe auf einem signifikanten Unterschied zwischen den beschriebenen kritischen und allen Bedingungen beruht.

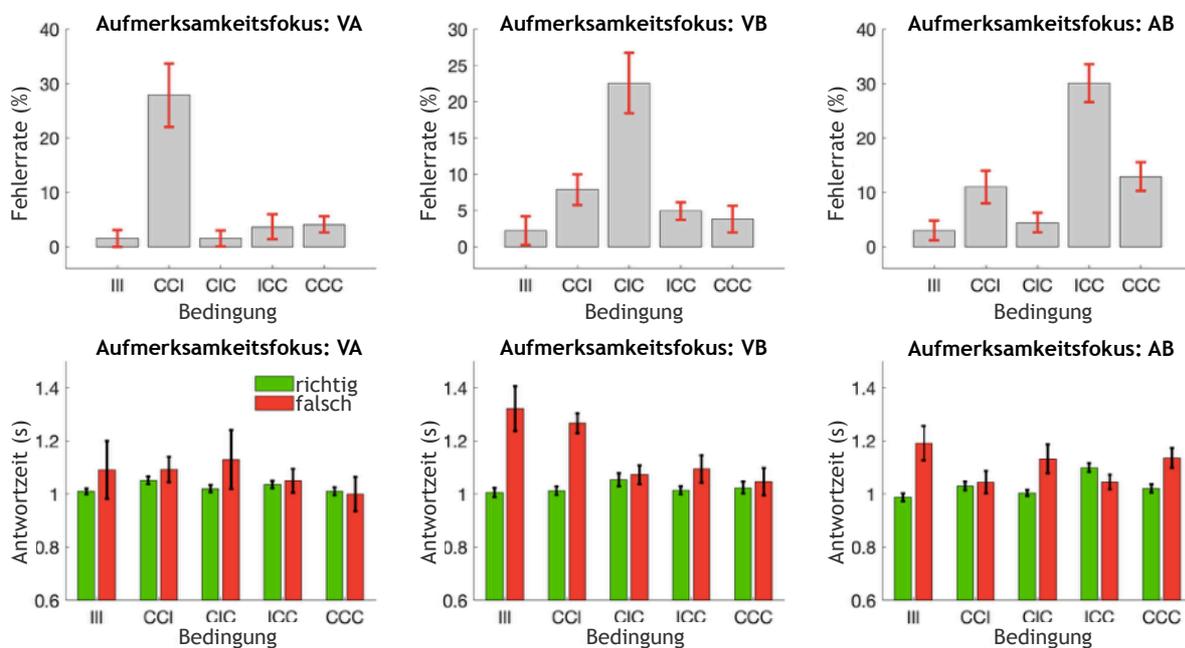


Abb. 4 Fehlerraten und Antwortzeiten in den drei Gruppen mit unterschiedlichem Aufmerksamkeitsfokus jeweils in den fünf Bedingungen

In den kritischen Bedingungen mit erhöhter Fehlerrate ist „Ja, passend“ die korrekte Antwort. Bedingt durch das Studiendesign ist in den meisten Bedingungen und damit während des Experiments insgesamt häufiger „nein, nicht passend“ die korrekte Antwort. So wäre eine Verzerrung in Richtung „nicht passend“ möglich.

Um diesen möglichen Bias zu untersuchen, betrachteten wir die Ergebnisse nicht nur nach Korrektheit sondern auch nach gegebener Antwort (vgl. Abbildung 5).

Wäre eine solche Verzerrung vorhanden, erwartete man die Antwort „Nein, nicht passend“ häufiger in allen Bedingungen. Dies war nicht der Fall. Wir sahen hingegen einen ähnlichen Effekt wie in Abbildung 4, dass nur in den kritischen Bedingungen „Nein“ als fehlerhafte Antwort gehäuft vorkommt. In den anderen Bedingungen wurde „ja“ und „nein“ gehäuft abhängig von der Korrektheit angegeben.

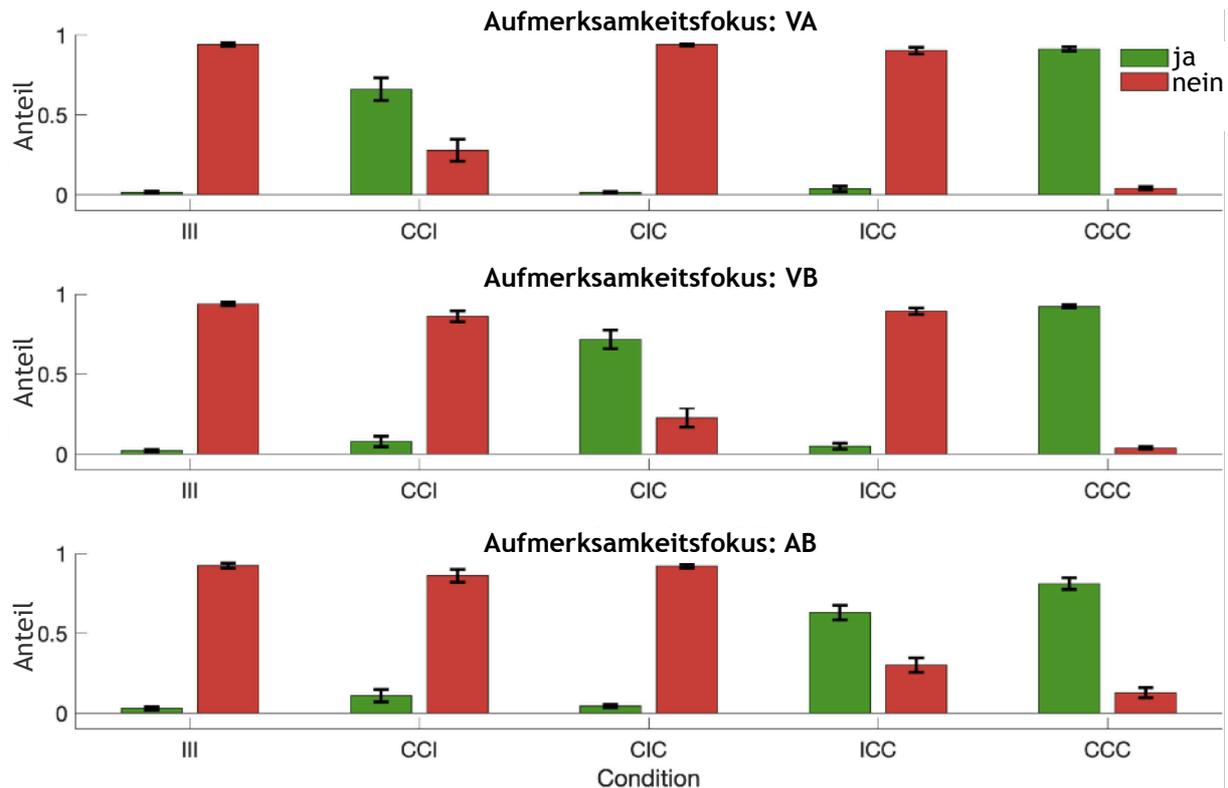


Abb. 5 Antworthäufigkeiten zur Untersuchung einer möglichen Verzerrung

### 2.3.2 Antwortzeiten

Bei den Antwortzeiten ließ sich hingegen kein klares Muster der signifikanten Effekte erkennen (vgl. Abbildung 4). Dabei spielen die Bedingungen unter den korrekten und fehlerhaften Antworten eine signifikante Rolle; der Interaktionseffekt von Bedingung und Fokus ist bei den Zeiten unter den korrekten Antworten sicher nachweisbar, unter den fehlerhaften scheint er wahrscheinlich. Es zeigt sich unter den korrekten Antworten eine signifikant längere Antwortzeit in der kritischen Bedingung ICC in der Gruppe AB. Unter den fehlerhaften Antworten ergaben sich in den Bedingungen III und CCI in der Gruppe VB signifikant längere Antwortzeiten als in den anderen Bedingungen. Weitere Effekte unter den Antwortzeiten ließen sich nicht nachweisen.

Zusammenfassend ergab sich in unserem Verhaltensexperiment eine erhöhte Fehlerrate bei Kongruenz der im zielgerichteten Aufmerksamkeitsfokus stehenden Stimuluskomponenten und inkongruentem Störstimulus ohne nachweisbaren Effekt der Antwortzeiten in diesen kritischen Bedingungen.

### 2.3.3 Modellierung mittels HDDM

Um zu untersuchen, welche Faktoren diesen Effekt begründen, führten wir anschließend eine Modellierung mittels HDDM durch. Wir stellten viele verschiedene Modelle mit unterschiedlichen Parametervariationen auf. Dabei variierten wir die *drift-rate*  $v$ , die die Geschwindigkeit angibt, mit der Entscheidungshinweise gesammelt werden, die *boundary separation*  $a$ , die die Mindestmenge an Hinweisen und damit Entscheidungssicherheit bestimmt sowie die *non-decision time*  $t$ . Diese drei Parameter ließen wir in einem einfachen Modell für alle Bedingungen gleich und auch für alle Gruppen gleich, und stellten systematisch verschiedene Modellvarianten auf, in denen die Parameter pro Bedingung bzw. pro Gruppe separat bestimmt wurden. Diese verglichen wir über die DIC Auswertung (vgl. Abbildung 6).

Dabei zeigten sich die Modelle, in denen alle drei Parameter für jede Bedingung einzeln bestimmt wurden, am geeignetsten. Diese sind die ersten vier Modelle in Abbildung 6. Ein klarer Nutzen der Bestimmung in Abhängigkeit des Aufmerksamkeitsfokus wurde nicht deutlich (DIC Unterschiede < 15).

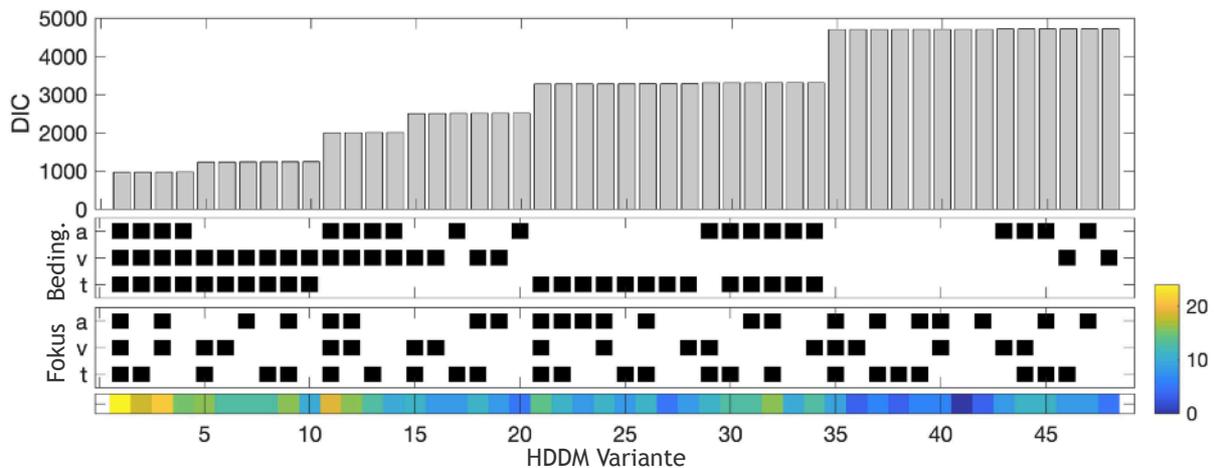


Abb. 6 HDD-Modell-Vergleich mittels DIC-Bewertung. Modellparameter abhängig von Bedingung (Mitte) bzw. Fokus (Unten)

Modell 1 (mit der niedrigsten DIC Auswertung), in dem die Modellparameter nicht nur für jede Bedingung sondern auch für jede Gruppe einzeln bestimmt wurden, ermöglichte uns dabei die Parameterverteilung für jeden Aufmerksamkeitsfokus und für jede Bedingung zu untersuchen. Da die kritische Bedingung in den einzelnen Gruppen jeweils eine andere ist, benötigten wir dies für eine detaillierte Auswertung. In der Gruppe mit VA-Fokus erfolgten die meisten Fehler in der Bedingung CCI, in der Gruppe mit VB-Fokus war es die Bedingung CIC und in der Gruppe mit Aufmerksamkeitsfokus AB war die Fehlerrate in der Bedingung ICC am höchsten. In Abbildung 7 sind die einzelnen Parameterschätzungen aus Modell 1 abgebildet, wobei die kritischen Bedingungen jeweils weiß umrandet sind.

Auffällig ist, dass die drift rate (Abbildung 7 links) in den kritischen Bedingungen in jeder Gruppe kleiner ist als in den anderen Bedingungen. Dies passt zu der erhöhten Fehlerrate in diesen Bedingungen, die wir wie oben beschrieben nachweisen konnten. Zudem sieht man in Abbildung 7, dass auch die boundary separation (mittig) in diesen kritischen Bedingungen jeweils am niedrigsten ist. Der Unterschied zu den anderen Bedingungen ist bei diesem Parameter jedoch nicht so deutlich wie bei der drift rate. Für die non-decision time (Abbildung 7 rechts) ergaben sich keine signifikanten Unterschiede für die kritischen Bedingungen.

Als weiteren Schritt zur Auswertung des Modells im Vergleich zu den erhobenen Daten führten wir post predictive checks (PPC) durch. Dabei berechneten wir für die Modelle mit niedrigen DICs 500 Datenpunkte, die die Versuchsergebnisse wiedergeben. Beispielhaft sind diese generierten Ergebnisse aus Modell 4 (Parameter  $v$ ,  $a$ ,  $t$  abhängig von der Bedingung, unabhängig von der Gruppe) in Abbildung 8 im Vergleich mit den im Experiment erhobenen Daten dargestellt. Die Fehlerraten werden dabei gut modelliert, die generierten Antwortzeiten geben die leichten Unterschiede in den Antwortzeiten der im Experiment erhobenen Daten nicht wieder. Bei den durch die Modellierung generierten Daten fällt auf, dass die Antwortzeiten über alle Bedingungen und alle Gruppen nicht signifikant variiert.

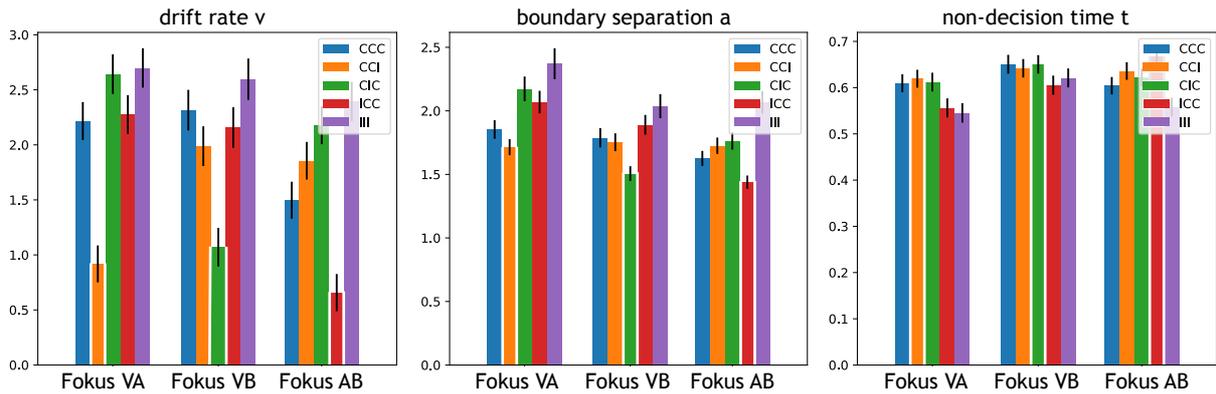


Abb. 7 Parameterschätzung aus Modell 1

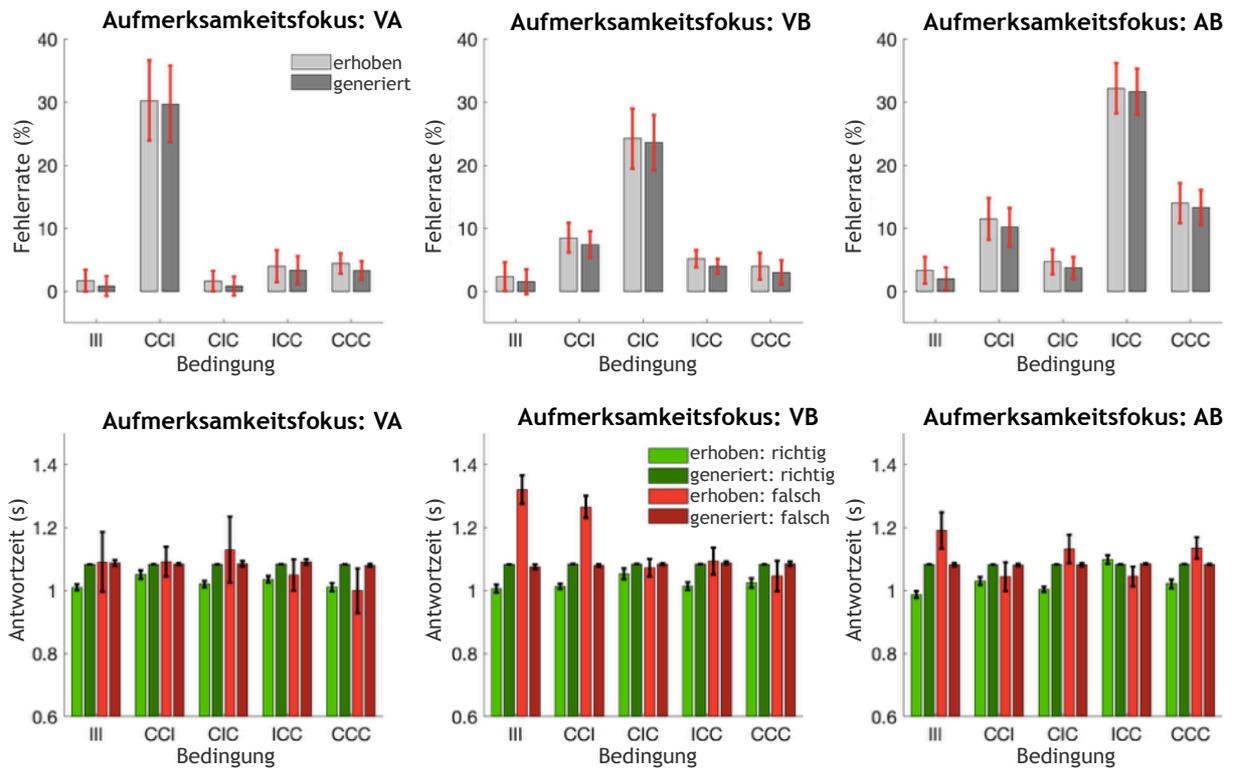


Abb. 8 PPC mit generierten Daten aus Modell 4 im Vergleich mit erhobenen Daten

## 2.4 Diskussion

In unseren Verhaltensdaten wiesen wir bei der Verarbeitung multisensorischer Reize einen konkreten Effekt nach. Sobald ein nicht beachteter Störstimulus inhaltlich nicht mit zwei im Aufmerksamkeitsfokus stehenden, semantisch kongruenten Stimuli übereinstimmt, machten die Teilnehmer signifikant mehr Fehler bei der Bewertung der Kongruenz als unter allen anderen Bedingungen. Dies drückt sich bei unserer Modellierung mittels HDDM in reduzierten *drift rates* in diesen Bedingungen aus. Bei der Auswertung der Antwortzeiten lässt sich kein systematischer Effekt nachweisen, sie scheinen nicht von der Aufmerksamkeitslenkung beeinflusst zu werden.

Unsere Ergebnisse stehen somit im Einklang mit den Studien, in denen gezeigt wurde, dass unbeachtete inkongruente Reize Verarbeitungsressourcen einnehmen und so die Bewertung von beachteten Reizen stören können (vgl. Zimmer et al., 2010a,b, Misselhorn et al., 2016; Corbetta and Shulman, 2002). In unserer Studie

heißt das, dass ein inkongruenter Störstimulus die Aufmerksamkeit so auf sich zieht, dass die Probanden von ihren beachteten Stimuli abgelenkt werden. Dadurch richten sie ihre Aufmerksamkeit auf den Störstimulus und machten in ihren Augen mit der Antwort „Nein, nicht passend“ eine korrekte Entscheidung. So ließen sich die fehlenden Unterschiede in den Antwortzeiten bei korrekten und inkorrekten Antworten in den kritischen Bedingungen erklären.

Einen stärkeren Ablenkungseffekt durch auditive Störreize im Gegensatz zu visuellen, wie er in anderen Studien nachgewiesen wurde (Leiva et al., 2015; Li et al. 2020), konnten wir nicht belegen. Dies kann im unterschiedlichen Studiendesign begründet sein, da in unserer Studie die Stimuli aus zwei visuellen und einer auditiven Komponente bestanden.

Den Effekt der Aufmerksamkeitslenkung bei Erkennen der Inkongruenz zwischen initial beachteten (kongruenten) Stimuli und initial nicht beachtetem (inkongruenten) Stimulus und daraus resultierender Entscheidung zur Inkongruenzantwort konnten wir in allen Gruppen und damit unabhängig von der Modalität nachweisen. Es scheint also eine generelle Auswirkung semantischer Kongruenz und Inkongruenz auf die Aufmerksamkeitslenkung zu geben.

Bei der Wahrnehmung von Konflikten ist in bildgebenden Studien eine stärkere Aktivierung des anterioren cingulären Kortex nachgewiesen. In diesem Bereich zeigt sich auch eine stärkere Aktivierung bei der Verarbeitung semantisch inkongruenter Stimuli (Weissman et al., 2004; Zimmer et al., 2010b; Littman et al., 2019). Bei der crossmodalen Reizverarbeitung wurde zudem eine Aktivitätsveränderung in den Bereichen der unimodalen Verarbeitung beschrieben. Dabei zeigte sich eine stärkere Aktivierung der primären sensorischen Kortexareale, wenn die Sinnesmodalität aufgabenrelevant war (Weissman et al., 2004) und (inhaltlich) mit anderen Stimulusmodalitäten übereinstimmte (van Atteveldt et al., 2004). Eine aktive Aufgabenstellung verringerte im Vergleich zur rein passiven Präsentation den Kongruenzeffekt (van Atteveldt et al., 2007). In anderen Studien konnte bei inkongruenten Stimuli eine stärkere Aktivierung der primären sensorischen Areale der beachteten Modalität nachgewiesen werden (Weissman et al., 2004).

Aus diesen neuronalen Ergebnissen lässt sich ein Einfluss sowohl der zielgerichteten als auch der Stimulus-getriebenen Aufmerksamkeit auf die Verarbeitung multimodaler inhaltlicher Kongruenz ableiten: Richtet man die Aufmerksamkeit bewusst auf eine Ziel-Modalität (endogene Aufmerksamkeit), lässt sich eine erhöhte Aktivität in den unimodalen sensorischen Arealen nachweisen; ist aber die Ziel-Modalität inkongruent zu einer nicht beachteten Störmodalität, werden weitere Verarbeitungsressourcen einbezogen (exogene Aufmerksamkeit) und die Aktivität in den dazu gehörigen Hirnarealen ist erhöht.

Auch in aktuellen Studien wird weiterhin das Zusammenspiel und die Abhängigkeit von multimodaler Integration und Aufmerksamkeitslenkung diskutiert. Einerseits wird eine frühe crossmodale Integration als „bottom-up“-Verarbeitung beschrieben, wobei selektive Aufmerksamkeit exogen die Rolle übernimmt, herausstechende multisensorische Stimuli zu erkennen. Durch erlernte Zusammenhänge und gespeicherte Schablonen wird zudem ein „top-down“ Einfluss ausgeübt (Fu et al., 2020). Dabei können kongruente audiovisuelle Stimuli auch ohne relevante

Aufmerksamkeitsressourcen im Sinne einer multimodalen Integration verarbeitet werden (Li et al., 2020; Spence and Frings, 2020). Sobald Inkongruenz auftritt, wird ein „top-down“ Prozess gemäß des aktuellen Ziels und der relevanten Modalitäten begonnen (Fu et al., 2020). Bei niedriger kognitiver Last wird so sowohl ein auditiver, als auch ein visueller Ablenkungseffekt deutlich. Werden jedoch bei aufgabenbedingt hoher kognitiver Auslastung Aufmerksamkeitsressourcen erschöpft, können irrelevante, nicht übereinstimmende Informationen automatisch herausgefiltert und von der weiteren Verarbeitung ausgeschlossen werden (Li et al., 2020).

Unsere Ergebnisse stimmen mit der Hypothese der Hinzunahme von Ressourcen durch Stimulus-getriebene Aufmerksamkeitslenkung überein. In unserer Studie wurden mehr Fehler gemacht, wenn der Störstimulus inkongruent zu den zwei im Aufmerksamkeitsfokus stehenden Stimulusanteilen war, unabhängig von den jeweiligen Modalitäten. Dies spricht für einen generellen Einfluss der Aufmerksamkeit bei der Verarbeitung semantischer Inkongruenz.

Bei der Auswertung unserer *drift diffusion* Modellierung zeigte sich, dass die höhere Fehlerrate in den kritischen Bedingungen mit einer niedrigeren *drift rate* einhergeht (vgl. Abbildung 5). Die *drift rate* beschreibt die Geschwindigkeit, mit der Entscheidungshinweise gesammelt werden, bis eine Grenze erreicht und eine Entscheidung getroffen wird (Ratcliff and McKoon, 2008). Überträgt man dies auf kognitive Verarbeitung, bedeutet eine niedrigere *drift rate* bei gleich bleibender Entscheidungsgrenze, dass Probanden mehr Zeit benötigen, um Hinweise zu sammeln und eine Entscheidung mit gleicher Sicherheit treffen zu können. Dies spiegelt typischerweise die Schwierigkeit der Aufgabenbedingung wider. In unseren kritischen Bedingungen liegt eine solche erschwerte Situation vor. Die Sammlung von Entscheidungshinweisen ist dadurch verlangsamt, dass die durch inkongruente Störstimulusanteile hervorgerufene Stimulus-getriebene Aufmerksamkeit mit zielgerichteter Aufmerksamkeit konkurriert. Diese Schlussfolgerung wird durch die errechnete niedrigere *boundary separation* in den kritischen Bedingungen ebenfalls gestützt. Die Entscheidungsgrenzen sind vermindert und die Antwortmöglichkeiten „passend“ bzw. „nicht passend“ so jeweils unsicherer, was durch den Einfluss der Stimulus-getriebenen Aufmerksamkeit begründet werden kann. Die Kombination aus niedriger *drift rate* und verminderter *boundary separation* drückt unsere Ergebnisse mit unbeeinträchtigten Antwortzeiten und höheren Fehlerraten in den kritischen Bedingungen aus. Durch die kognitive Modellbildung stellen sich Zusammenhänge dar, die sich bei der konventionellen Analyse der Antwortzeiten wohlmöglich nicht ergeben hätten.

## 2.5 Limitation

In unseren Antwortzeiten konnten wir einige signifikante Unterschiede zwischen richtig und falsch beantworteten trials in den Gruppen VB und AB nachweisen (vgl. Abbildung 3). Dabei ist jedoch kein klares Muster im Vergleich zu den Aufgabenanforderungen zu erkennen. Ein möglicher Grund für diese nicht nachvollziehbaren Effekte könnte die insgesamt niedrige Fehlerrate in dem Experiment sein. Dadurch gibt es nur wenige fehlerhaft beantwortete Trials, die als Grundlage zur Auswertung und Berechnung der Mittelwerte der Antwortzeiten genommen wurden. Lange Antwortzeiten können somit in einzelnen Ausreißerdaten begründet liegen, die durch unsere Datenaufbereitung nicht miterfasst wurden.

Die niedrige Anzahl und das fehlende Muster der Antwortzeiten ist wahrscheinlich auch der Grund, dass wir diese in unserer Modellierung mittels HDDM im posterior predictive check nicht nachahmen konnten, obwohl wir die Fehlerraten gut abbilden konnten (vgl. Abbildung 7). In den errechneten Daten lässt sich sogar kein einziger Effekt bei den Antwortzeiten nachweisen, was dafür spricht, dass die beobachteten Effekte unsystematisch sind und durch klassische drift diffusion Modellierung nicht modelliert werden können. Zur weiteren Untersuchung könnte man zusätzliche Parameter hinzunehmen, um beispielsweise für die *drift rate*, *starting point* oder *non-decision time* eine inter-trial Variabilität zu modellieren. Wir verzichteten hierauf, da wir uns auf die Kernparameter der DDM wie *drift rate* und *boundary separation* konzentrierten und die Schätzung dieser durch Hinzunahme der anderen beeinträchtigt werden kann (Boehm et al., 2018).

Dies heißt nicht, dass klassische drift diffusion Modellierung nicht geeignet ist, die Daten unseres Experiments zu modellieren. Durch genaues Abschätzen der drift rate, boundary separation und non-decision time einzeln für jede Bedingung konnten wir das Muster der richtigen und fehlerhaften Antworten mit hoher Genauigkeit modellieren (vgl. Abbildung 7). Dies bestärkt die Interpretation, dass eine niedrigere *drift rate* in den kritischen Bedingungen einen höheren Verarbeitungsanspruch bedingt durch die Aufmerksamkeitsverschiebung auf den zuvor nicht beachteten inkongruenten Stimulus widerspiegelt.

## 2.6 Schlussfolgerung

Semantische Kongruenz ist ein mächtiger Verstärker der multisensorischen Integration, der zu höherer Gehirnaktivität (Doehrmann and Naumer, 2008) und besserer Leistung (Taylor et al., 2006; Spilcke-Liss, 2017; Steinweg and Mast, 2017) führt. Zudem kann dadurch unwillkürlich Stimulus-getriebene Aufmerksamkeit auf einen kongruenten Reiz gelenkt und Wahrnehmungsvorgänge beschleunigt werden (Iordanescu et al., 2008, 2010). Überdies kann semantische Inkongruenz Wahrnehmungsprozesse im Aufmerksamkeitsfokus stehender Stimuli unterbrechen (Zimmer et al., 2010a,b, Tang et al., 2016).

Die Ergebnisse dieser Studie stimmen mit den beschriebenen Erkenntnissen überein. Wir konnten zeigen, dass semantische Inkongruenz unabhängig von der Stimulusmodalität zu einer Refokussierung der Aufmerksamkeit auf einen zuvor unbeachteten, inkongruenten Reiz führt. So scheint es, dass unser Wahrnehmungssystem unbewusst darauf ausgerichtet ist, inhaltliche Nicht-Übereinstimmung aufzuspüren. Im Sinne des *predictive coding* (Rao and Ballard, 1999; Friston, 2005) stellen solche Inkongruenzen fehlerhafte Vorhersagen (*prediction errors*) dar, die zu einer Neuordnung von Verarbeitungsressourcen durch Stimulus-getriebene Aufmerksamkeitslenkung führen, um unbedingt den inkongruenzbedingten Wahrnehmungskonflikt aufzulösen. Dies würde bedeuten, dass die Verhaltensleistung unserer Probanden in den kritischen Bedingungen nicht fehlerhaft, sondern angepasst war an die Anforderungen für weitere Verarbeitungsprozesse unabhängig von der Modalität.

### 3. Zusammenfassung in deutscher und englischer Sprache

Werden multimodale Reize gemeinsam als zusammengehörig verarbeitet, spricht man von multimodaler Integration. Dieser Zusammenschluss wird durch viele Faktoren beeinflusst, wie zum Beispiel von der inhaltlichen Bedeutung (Semantik) der einzelnen Reize und Beachtung dieser mit zielgerichteter (endogener) Aufmerksamkeit. Semantische Kongruenz kann dabei Verarbeitungsressourcen zu einem kongruenten Stimulus umlenken, während Aufmerksamkeit die Integration beachteter multimodaler Stimuli begünstigt. Wie sich diese Verhältnisse unter Beachtung von drei Reizen, von denen zwei im zielgerichteten Aufmerksamkeitsfokus stehen, verhalten, untersuchten wir in dieser Studie. Wir erstellten Stimuli aus drei Komponenten: 1. ein visuelles Objekt oder Lebewesen im Vordergrund (V), 2. ein Ton (A) und 3. ein visueller Hintergrund (B). Unsere Studienteilnehmer wiesen wir an, sich auf zwei dieser Komponenten zu konzentrieren und deren Kongruenz zu bewerten. Dazu teilten wir die Probanden in drei Gruppen ein: eine sollte das Bild im Vordergrund und den Ton beachten (Gruppe VA, n=30), die zweite das Bild im Vordergrund und den Hintergrund (Gruppe VB, n=27) und die dritte den Ton und den Hintergrund (Gruppe AB, n=30). Die Kongruenz variierten wir systematisch zwischen den drei Stimuluskomponenten: es konnten alle Komponenten inhaltlich übereinstimmen (z.B. V: Löwe, A: Gebrüll, B: Savanne), alle zueinander inkongruent sein (z.B. V: Krankenwagen, A: Glockengeläut, B: Schwimmbad) oder zwei der drei Komponenten kongruent sein mit dem dritten inkongruent dazu (z.B. V: Ente, A: Geschnatter, B: Telefonzelle). Als Ergebnis konnten wir einen Effekt in allen Gruppen nachweisen: Wenn die beiden beachteten Stimuluskomponenten inhaltlich zueinander passten, die unbeachtete Komponente aber semantisch inkongruent war, wurden deutlich mehr Fehler als in allen anderen Stimuluskombinationen gemacht. Zur weiteren Einordnung dieses Resultats führten wir eine hierarchische Drift-diffusion-Modellierung durch, die in den kritischen Bedingungen mit Kongruenz der beachteten Reize und inkongruentem Störreiz eine signifikant kleinere drift rate ergab. Die drift rate gibt bei dieser Modellierung von Entscheidungsprozessen die Ansammlung von Entscheidungshinweisen an. Ist sie niedrig, spricht dies für erschwerte Aufgabenanforderung. In unserer Studie wird also die Entscheidung in der kritischen Bedingung durch die Beeinflussung durch den inkongruenten Reiz erschwert. Dieser Einfluss auf die Hinweisansammlung war unabhängig davon, welche Objektpaare beachtet wurden. Eine Auswirkung der Sinnesmodalität auf die Anfälligkeit der Kongruenzbewertung durch inkongruente, unbeachtete Störreize scheint es somit nicht zu geben. Durch eine Kontrollanalyse konnten wir schnell eine mögliche Verzerrung ausschließen, die die erhöhte Fehlerrate durch insgesamt häufigere korrekte „nein, nicht passend“-Antworten begründen könnte. Wir konnten in dieser Studie nachweisen, dass unsere Wahrnehmung sehr empfindlich für semantische Inkongruenz ist, auch wenn diese nicht im bewussten, zielgerichteten Aufmerksamkeitsfokus steht.

Efficient multisensory integration is often influenced by other cognitive processes including, but not limited to, semantic congruency and focused endogenous attention. Semantic congruency can re-allocate processing resources to the location of a congruent stimulus, while attention can prioritize the integration of multi-sensory stimuli under focus. Here, we explore the robustness of this phenomenon in the context of three stimuli, two of which are in the focus of endogenous attention.

Participants completed an endogenous attention task with a stimulus compound consisting of 3 different objects: (1) a visual object (V) in the foreground, (2) an auditory object (A), and (3) a visual background scene object (B). Three groups of participants focused their attention on either the visual object and auditory sound (Group VA,  $n = 30$ ), the visual object and the background (VB,  $n = 27$ ), or the auditory sound and the background (AB,  $n = 30$ ), and judged the semantic congruency of the objects under focus. Congruency varied systematically across all 3 stimuli: All stimuli could be semantically incongruent (e.g., V, ambulance; A, church bell; and B, swimming-pool) or all could be congruent (e.g., V, lion; A, roar; and B, savannah), or two objects could be congruent with the remaining one incongruent to the other two (e.g., V, duck; A, quack; and B, phone booth). Participants exhibited a distinct pattern of errors: when participants attended two congruent objects (e.g., group VA: V, lion; A, roar), in the presence of an unattended, incongruent third object (e.g., B, bath room) they tended to make more errors than in any other stimulus combination. Drift diffusion modeling of the behavioral data revealed a significantly smaller drift rate in two-congruent-attended condition, indicating slower evidence accumulation, which was likely due to interference from the unattended, incongruent object. Interference with evidence accumulation occurred independently of which pair of objects was in the focus of attention, which suggests that the vulnerability of congruency judgments to incongruent unattended distractors is not affected by sensory modalities. A control analysis ruled out the simple explanation of a negative response bias. These findings implicate that our perceptual system is highly sensitive to semantic incongruencies even when they are not endogenously attended.

#### 4. Literaturverzeichnis

- Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., et al. (2018). Estimating across-trial variability parameters of the diffusion decision model: expert advice and recommendations. *J. Math. Psychol.* 87, 46–75. doi: 10.1016/j.jmp.2018.09.004
- Calvert, G. A. (2001). Crossmodal Processing in the Human Brain: Insights from Functional Neuroimaging Studies. *Cerebral Cortex* 11 (12), 1110–1123. doi: 10.1093/cercor/11.12.1110
- Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755
- De Meo, R., Murray, M. M., Clarke, S., & Matusz, P. J. (2015). Top-down control and early multisensory processes: chicken vs. egg. *Frontiers in integrative neuroscience*, 9, 17. doi: 10.3389/fnint.2015.00017
- Doehrmann, O., and Naumer, M. J. (2008). Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. *Brain Res.* 1242, 136–150. doi: 10.1016/j.brainres.2008.03.071
- Friston, K. (2005). A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* 360, 815–836. doi: 10.1098/rstb.2005.1622
- Iordanescu, L., Grabowecky, M., Franconeri, S., Theeuwes, J., and Suzuki, S. (2010). Characteristic sounds make you look at target objects more quickly. *Atten. Percept. Psychophys.* 72, 1736–1741. doi: 10.3758/APP.72.7.1736
- Iordanescu, L., Guzman-Martinez, E., Grabowecky, M., and Suzuki, S. (2008). Characteristic sounds facilitate visual search. *Psychon. Bull. Rev.* 15, 548–554. doi: 10.3758/PBR.15.3.548
- Koelewijn, T., Bronkhorst, A., and Theeuwes, J. (2010). Attention and the multiple stages of multisensory integration: a review of audiovisual studies. *ACTPSY* 134, 372–384. doi: 10.1016/j.actpsy.2010.03.010
- Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental brain research*, 158(4), 405–414. doi: 10.1007/s00221-004-1913-2
- Leiva, A., Parmentier, F. B., & Andrés, P. (2015). Distraction by deviance: comparing the effects of auditory and visual deviant stimuli on auditory and visual target processing. *Experimental psychology*, 62(1), 54–65. doi: 10.1027/1618-3169/a000273
- Li, Q., Wu, Q., Yu, Y., Wu, F., Takahashi, S., Ejima, Y., Yang, J., & Wu, J. (2020). Semantic Congruency Modulates the Effect of Attentional Load on the Audiovisual Integration of Animate Images and Sounds. *i-Perception*, 11(6), 2041669520981096. doi: 10.1177/2041669520981096
- Macaluso, E. and Driver, J. (2005). Multisensory spatial interactions: a window onto functional integration in the human brain. *Trends Neurosci.* 28 (5), 264–271. doi: 10.1016/j.tins.2005.03.008
- Mastroberardino, S., Santangelo, V., and Macaluso, E. (2015). Crossmodal semantic congruence can affect visuo-spatial processing and activity of the fronto-parietal attention networks. *Front. Integr. Neurosci.* 9:45. doi: 10.3389/fnint.2015.00045
- Matusz, P. J., Broadbent, H., Ferrari, J., Forrest, B., Merkley, R., and Scerif, G. (2015). Multi-modal distraction: insights from children's limited attention. *Cognition* 136, 156–165. doi: 10.1016/j.cognition.2014.11.031
- Meredith, M. A., and Stein, B. E. (1986). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Res.* 365, 350–354. doi: 10.1016/0006-8993(86)91648-3
- Misselhorn, J., Daume, J., Engel, A. K., and Fries, U. (2016). A matter of attention: crossmodal congruence enhances and impairs performance in a novel trimodal matching paradigm. *Neuropsychologia* 88, 113–122. doi: 10.1016/j.neuropsychologia.2015.07.022

- Morís Fernández, L., Visser, M., Ventura-Campos, N., Ávila, C., & Soto-Faraco, S. (2015). Top-down attention regulates the neural expression of audiovisual integration. *NeuroImage*, *119*, 272–285. doi: 10.1016/j.neuroimage.2015.06.052
- Mozolic, J.L., Hugenschmidt, C.E., Peiffer, A.M., Laurienti, P.J. (2008). Modality-specific selective attention attenuates multisensory integration. *Exp. Brain Res.* *184*, 39-52. doi: 10.1007/s00221-007-1080-3
- Noppeney, U., Ostwald, D., & Werner, S. (2010). Perceptual decisions formed by accumulation of audiovisual evidence in prefrontal cortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *30*(21), 7434–7446. doi: 10.1523/JNEUROSCI.0455-10.2010
- Rao, R. P. N., and Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* *2*, 79–87. doi: 10.1038/4580
- Ratcliff, R., and McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* *20*, 873–922. doi: 10.1162/neco.2008.12-06-420
- Ratcliff, R., and Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychol. Sci.* *9*, 347–356. doi: 10.1111/1467-9280.00067
- Senkowski, D., Schneider, T. R., Foxe, J. J., and Engel, A. K. (2008). Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci.* *31*, 401–409. doi: 10.1016/j.tins.2008.05.002
- Spence, C., & Frings, C. (2020). Multisensory feature integration in (and out) of the focus of spatial attention. *Attention, perception & psychophysics*, *82*(1), 363–376. <https://doi.org/10.3758/s13414-019-01813-5>
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. B* *64*, 583–639. doi: 10.1111/1467-9868.00353
- Spilcke-Liss, J. (2017). Die Rolle von semantischer Kongruenz bei audio-visueller Integration. Studienarbeit am Universitätsklinikum Hamburg Eppendorf
- Stein B.E. and Meredith M.A. (1993). *Merging of the senses*. Cambridge, MA: MIT Press.
- Steinweg, B., and Mast, F. W. (2017). Semantic incongruity influences response caution in audio-visual integration. *Exp. Brain Res.* *235*, 349–363. doi: 10.1007/s00221-016-4796-0
- Styles, E. (2006). *The Psychology of Attention*. New York, NY: Psychology Press
- Talsma, D., Senkowski, D., Soto-Faraco, S., and Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends Cogn. Sci.* *14*, 1–11. doi: 10.1016/j.tics.2010.06.008
- Tang, X., Wu, J., Shen, Y. (2016). The interactions of multisensory integration with endogenous and exogenous attention. *Neurosci. Biobehav. Rev.* *61*, 208-224, doi: 10.1016/j.neubiorev.2015.11.002
- Taylor, K. I., Moss, H. E., Stamatakis, E. A., and Tyler, L. K. (2006). Binding crossmodal object features in perirhinal cortex. *Proc. Natl. Acad. Sci. U.S.A.* *103*, 8239–8244. doi: 10.1073/pnas.0509704103
- ten Oever, S., Romei, V., van Atteveldt, N., Soto-Faraco, S., Murray, M.M., Matusz, P.J. (2016). The COGs (context, object, and goals) in multisensory processing. *Exp. Brain Res.* *234*, 1307-1323. doi: 10.1007/s00221-016-4590-z
- van Atteveldt, N., Formisano, E., Goebel, R., and Blomert, L. (2004). Integration of letters and speech sounds in the human brain. *Neuron* *43*, 271–282. doi: 10.1016/j.neuron.2004.06.025
- van Atteveldt, N. M., Formisano, E., Goebel, R., and Blomert, L. (2007). Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. *Neuroimage* *36*, 1345–1360. doi: 10.1016/j.neuroimage.2007.03.065
- Weissman, D. H., Warner, L. M., and Woldorff, M. G. (2004). The neural mechanisms for minimizing cross-modal distraction. *J. Neurosci.* *24*, 10941–10949. doi: 10.1523/JNEUROSCI.3669-04.2004

- Wiecki, T. V., Sofer, I., and Frank, M. J. (2013). HDDM: hierarchical bayesian estimation of the drift-diffusion model in python. *Front. Neuroinform.* 7:14. doi: 10.3389/fninf.2013.00014
- Zimmer, U., Itthipanyanan, S., Grent-'t-Jong, T., and Woldorff, M. G. (2010a). The electrophysiological time course of the interaction of stimulus conflict and the multisensory spread of attention. *Eur. J. Neurosci.* 31, 1744–1754. doi: 10.1111/j.1460-9568.2010.07229.x
- Zimmer, U., Roberts, K. C., Harshbarger, T. B., and Woldorff, M. G. (2010b). Multisensory conflict modulates the spread of visual attention across a multisensory object. *Neuroimage* 52, 606–616. doi: 10.1016/j.neuroimage.2010.04.245

## 5 Abkürzungsverzeichnis

A	auditiver Ton als Stimuluskomponente
ACC	anteriorer cingulärer Kortex
B	visueller Hintergrund als Stimuluskomponente
CCC	Bedingung mit Kongruenz von V, A und B
CCI	Bedingung mit Kongruenz von V und A sowie Inkongruenz von B
CIC	Bedingung mit Kongruenz von V und B sowie Inkongruenz von A
DIC	deviance information criterion
(H)DDM	(hierarchical) drift diffusion model
ICC	Bedingung mit Kongruenz von A und B sowie Inkongruenz von V
III	Bedingung mit Inkongruenz von V, A und B
fMRT	funktionelle Magnetresonanztomographie
PPC	posterior predictive check
UKE	Universitätsklinikum Hamburg-Eppendorf
RT	response time
SD	Standardabweichung
V	visueller Vordergrund als Stimuluskomponente

## 6 Abbildungs- und Tabellenverzeichnis

### Abbildungsverzeichnis

Abbildung 1: Fehlerrate und Antwortzeiten Pilotstudie ohne gerichtete Aufmerksamkeit.....	15
Abbildung 2: Stimulusbeispiel .....	18
Abbildung 3: Drift Diffusion Model aus Ratcliff, 2008 .....	20
Abbildung 4: Fehlerrate und Antwortzeiten in den drei Gruppen mit unterschiedlichem Aufmerksamkeitsfokus jeweils in den fünf Bedingungen.....	21
Abbildung 5: Antworthäufigkeiten zur Untersuchung einer möglichen Verzerrung...22	
Abbildung 6: HDD-Modell-Vergleich mittels DIC-Bewertung. Modellparameter abhängig von der Bedingung (Mitte) bzw. Fokus (Unten) .....	23
Abbildung 7: Parameterschätzung aus Modell 1 .....	24
Abbildung 8: PPC mit generierten Daten aus Modell 4 im Vergleich mit erhobenen Daten .....	24

### Tabellenverzeichnis:

Tabelle 1: Bedingungen und Beispiele für Stimuluskombinationen .....	18
--	----

## 7. Erklärung des Eigenanteils an der Promotion

Hiermit versichere ich, Julia Spilcke-Liss, dass ich folgende Anteile für die Erstellung der Publikationspromotion selbständig erarbeite habe:

- Literaturrecherche und –auswertung
- Findung und Formulierung der Fragestellung zusammen mit Dr. Jan Gläscher
- Promotionsskizze
- Programmierung und Durchführung der Verhaltensexperimente
- Datenerhebung und -aufarbeitung zusammen mit Dr. Jan Gläscher
- Ausarbeitung der Drift-Diffusion-Modelle
- Erstentwurf und Bearbeitung des Manuskripts zusammen mit Dr. Jan Gläscher

Die statistische Datenauswertung erfolgte in Zusammenarbeit mit Dr. Jan Gläscher.

Die finale Manuskriptbearbeitung der Publikation erfolgte in Kooperation mit J. Zhu, S. Gluth, M. Spezio und J. Gläscher.

## **8. Danksagung**

Ganz besonders möchte ich mich bei meinem Doktorarbeitsbetreuer und erstem Ansprechpartner in allen wissenschaftlichen und methodischen Fragen Herrn Dr. Jan Gläscher bedanken. Ohne seine Unterstützung und Geduld wäre diese Promotion nicht zu Stande gekommen.

Außerdem gilt mein Dank Prof. Christian Büchel, der mir als Doktorvater immer unterstützend zur Seite stand.

Zudem möchte ich mich beim gesamten Team der Arbeitsgruppe „Valuation and Social Decision Making“ und den Mitarbeitern am Institut für systemische Neurowissenschaften bedanken, die alle jederzeit ein offenes Ohr für mich hatten und mich im Institut willkommen geheißen haben.

Schließlich möchte ich noch meiner Familie und meinen Freunden für die unermüdliche Stärkung und Motivierung von ganzem Herzen danken. Ohne den vorgeführten Perspektivenwechsel und die hartnäckige Bekräftigung von Elias von Salomon und Dr. Beatrice von Jeinsen hätte ich die Arbeit nicht in dieser Form vollendet.

## **9. Lebenslauf**

Lebenslauf aus datenschutzrechtlichen Gründen nicht enthalten

Lebenslauf aus datenschutzrechtlichen Gründen nicht enthalten

## **10. Eidesstattliche Versicherung**

Ich versichere ausdrücklich, dass ich die Arbeit selbständig und ohne fremde Hilfe verfasst, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt und die aus den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen einzeln nach Ausgabe (Auflage und Jahr des Erscheinens), Band und Seite des benutzten Werkes kenntlich gemacht habe.

Ferner versichere ich, dass ich die Dissertation bisher nicht einem Fachvertreter an einer anderen Hochschule zur Überprüfung vorgelegt oder mich anderweitig um Zulassung zur Promotion beworben habe.

Ich erkläre mich einverstanden, dass meine Dissertation vom Dekanat der Medizinischen Fakultät mit einer gängigen Software zur Erkennung von Plagiaten überprüft werden kann.

Unterschrift: .....