GLOSSA
psycholinguistics

eScholarship
University of California

(How) Visual properties affect the perception and description of transitive events

Judith Schlenter, Department of Special Education and Rehabilitation, Faculty of Human Sciences, University of Cologne, Cologne, Germany; Department of Language and Culture at UiT The Arctic University of Norway, Tromsø, Norway, judith.schlenter@gmail.com

Martina Penke, Department of Special Education and Rehabilitation, Faculty of Human Sciences, University of Cologne, Cologne, Germany, martina.penke@uni-koeln.de

In a non-verbal aesthetic judgement task and a pre-registered production task, we tested how the orientation of the patient relative to the agent in a visual scene affects the perception and description of the depicted transitive event. Previous research has shown that a visual property like the position of the patient relative to the agent can affect speakers' verbalization of events. Here, we investigated whether orientation constitutes another factor besides position that affects scene description. While for most of the events speakers of German displayed a preference for scenes in which agent and patient faced each other, these scenes needed more time for sentence planning than the same scenes that showed the patient looking in the same direction and thus away from the agent. Moreover, we elicited more patient-initial sentences for face-to-face scenes than for same-direction scenes. The increase in patient-initial sentences was comparable to the increase in patient-initial sentences for scenes with left-positioned patients as compared to right-positioned patients. Based on our findings, we argue that manipulations of both position and orientation can change the prominence of the patient. The more prominent the patient (facing the agent, being placed to the left of the agent), the more likely speakers are to choose the patient as the sentence-initial subject. Hence, subtle changes of visual properties may affect not only how speakers perceive an event but also how they describe an event. Our findings are of relevance for a range of tasks that use visual materials.

1. The influence of visual properties on scene description

Research shows that visual attention is guided by a multitude of factors and can be user-driven as well as stimulus-driven. In a visual-search task, for example, a stimulus can be perceptually more salient or prominent because of its visual properties (e.g., color, motion, orientation, size) and can thus attract more attention as opposed to other stimuli in the visual scene, facilitating its recognition (for review, see Wolfe & Horowitz, 2017). In the current paper, we focus on effects of visual properties in two-dimensional, semi-realistic scenes as they are often used in language production research to elicit a one-sentence description of a depicted event (e.g., Esaulova et al., 2019; Gleitman et al., 2007; Norcliffe et al., 2015; Pokhoday et al., 2019; Sauppe et al., 2013). More precisely, we seek to investigate how two visual properties, character position and character orientation, affect the description of simple transitive events such as *A pushes B*. Throughout the text, we will use the term 'agent,' to refer to the initiator of the event or action, and 'patient' to refer to the undergoer of the action in line with Dowty (1991). We aim to know more about how character position and character orientation may influence sentence-planning times and which thematic role, agent or patient, is placed in sentence-initial position and thus becomes the starting point of the describing utterance.

The investigation of factors that determine the starting point in language production has a long tradition in psycholinguistic research. According to proponents of the starting point hypothesis, "what is most prominent to the speaker at the moment of utterance becomes the starting point" of the utterance (Bock et al., 2004, p. 249). With the advent of eye-tracking as a technique to study language production, however, it became obvious that the process between scene viewing and verbalization in a speaker's mind/brain is informed by a complex interplay between cognitive functions, such as perception and attention, and language, and that prominent elements are not necessarily the ones uttered first (for review, see Bock et al., 2004). So, what happens if the prominence of the patient in a transitive event scene is increased, does the patient become the starting point? First, we need to keep in mind that starting a sentence with the patient often means to begin a sentence with a structure that is typically more marked and less frequent. In English, for example, only a passive (*B is being pushed by A*) can follow. Moreover, we need to know what determines prominence and in how far prominence-lending factors affect language production. In the next two subsections, we review the literature indicating that agent and patient position and agent and patient orientation are two candidates for visual properties in scenes that not only affect perception but, crucially, might also affect scene description. Thus, position and orientation might be two prominence-lending factors, but as yet we do not know whether both have an effect on language production, meaning that manipulations of both position and orientation will affect sentence-planning times or syntactic choices, nor do we know whether these two factors interact. This is the topic of the current research.

1.1 Agent and patient position

It is common in language production studies to counterbalance the position of the agent relative to the patient character in a scene within and across presentation lists by creating a mirror-imaged version. That is, in most scene description tasks, agents appear equally often to the left than to the right of patients. This is done to account for the often-observed preference to scan pictures from left to right (e.g., Griffin & Bock, 2000), an observation which dates back to Buswell (1935), who was one of the first to report this preference for English speakers. Experimental studies show that literate speakers of European languages place the agent of an event more often to the left of the patient than to the right in drawing tasks (e.g., Chatterjee et al., 1999; Maass & Russo, 2003; Maass et al., 2014; Suitner et al., 2021), are faster in picture matching if the agent is positioned on the left (Chatterjee et al., 1999; Maass & Russo, 2003), preferably map a transitive event description onto a scene with a left-positioned agent (Maass et al., 2014), and indicate a scene with a left-positioned agent as more natural in an aesthetic judgement task than the same scene with a right-positioned agent (Esaulova, Dolscheid, Reuters et al., 2021). Between-language and between-group comparisons suggest that this so-called spatial-agency-bias (SAB) is related to speakers' reading/writing direction (e.g., Dobel, Diesendruck et al., 2007; Esaulova, Dolscheid, Reuters et al., 2021; Maass & Russo, 2003; but see Altmann et al., 2006) and may be further modulated by word order specifics (Maass et al., 2014; Suitner et al., 2021).[1] Literate speakers of Non-European languages such as Hebrew (Dobel, Diesendruck et al., 2007) or Arabic (Esaulova, Dolscheid, Reuters et al., 2021; Maass & Russo, 2003; Maass et al., 2014), who are familiar with a script that is written from right to left, have been found to display the reverse bias. Taken together, the findings from non-verbal tasks reveal that agents are typically associated with a certain position in the scene. For German, the language under investigation in this paper, the position of the agent is typically associated with the position to the left of the patient in transitive events (Esaulova, Dolscheid, Reuters et al., 2021; Suitner et al., 2021) and ditransitive events (Dobel, Diesendruck et al., 2007). German is a language with a left-to-right script that places subjects before objects when following the pragmatically unmarked word order (e.g., Bader & Häussler, 2010). Thus, a left-positioned agent mentally aligns with a preference for agent-initial active sentences (Esaulova, Dolscheid, Reuters et al., 2021).

Although spatial preferences have been acknowledged in experimental designs, character position has rarely been the focus of attention as a factor modulating sentence planning and production. In a scene description task in German, Esaulova, Dolscheid, Reuters et al. (2021)

---

[1] In a comparison between illiterate speakers of Yucatec Maya (VOS word order) and illiterate speakers of Spanish (SVO word order) by Dobel et al. (2014) both groups displayed no SAB, which led the authors to conclude that literacy is the key factor influencing the mental representations of events. Other factors appear to be of secondary importance.

found that speakers needed more time to produce a one-sentence description for transitive event scenes with agents situated to the right of patients than for scenes with agents situated to the left of patients (see also Esaulova et al., 2019, 2020; Esaulova, Dolscheid, & Penke, 2021). Thus, a violation of the SAB resulted in longer sentence-onset latencies. This indicates that a visual scene which does not conform to the mental linearization of events and, therefore, leads to a misalignment between visual event representation and a linguistically preferred structure (subject/agent-verb-object/patient) can affect the planning process and delay the initiation of the describing utterance. Moreover, in studies that made use of visual cueing of scene characters prior to scene onset and manipulated the position of agent and patient in the scene, speakers have been found to produce more patient-initial sentences (passives, patient-initial active sentences) when patients appeared in the left position (Esaulova et al., 2019, 2020; Pokhoday et al., 2019). For example, speakers of Russian produced 13% patient-initial sentences for agent-left scenes but 22% patient-initial sentences for patient-left scenes when presented with a red dot (for 500 ms) at the position the patient was to appear in the visual scene (Pokhoday et al., 2019). Averaged across cue positions (patient and agent), this resulted in an increase of around 5%. Similarly, Esaulova et al. (2020) reported a significant increase of passives for patient-left relative to agent-left scenes of around 4% for German. Altogether, previous studies that took the position of the agent relative to the patient into account demonstrate that the position of agent and patient in transitive event scenes can affect the planning of a verbal description. The preference for producing canonical, agent-initial active sentences in speakers of European languages might be reduced or the production delayed if the patient is placed in the visually prominent left position.

1.2 Agent and patient orientation

Another factor that might affect scene perception and description is the orientation of agent and patient towards each other: whether both face each other or both look in the same direction. Following Hafri et al. (2013), event role features such as head orientation (toward vs. away), body orientation (toward vs. away), extremities (outstretched vs. contracted) and body lean (toward vs. away) are visual indicators of agent- and patient-hood. Based on Dowty's proto-role entailments (volition, sentience, cause, autonomous motion), the authors suspect that how head and body are oriented corresponds to the degree of volition of a character. In addition, whether a character leans toward or away from the other character may influence the degree of volition and direction and degree of movement. Hence, the more a patient exhibits features typically associated with the agent (oriented toward the other character, outstretched extremities), the more prominent it may become, and the more attention is paid to the patient in comparison to a more prototypical patient that exhibits no such agentive features. After brief exposure to scenes (37 ms and 73 ms), Hafri et al. (2013, Exp. 4) indeed found the recognition of event roles to be

hindered if the patient was shown oriented toward the agent and with outstretched extremities.[2] Based on this finding, one might suspect that orientation of agent and patient towards each other also affects scene perception and, probably, description. A potential influence of agent-patient orientation might be enhanced by another entailment mentioned above, motion, or the direction thereof. As also noted by Bock et al. (2004), selective attention might be paid to agents because they are typically associated with the onset of motion. In scenes in which the patient looks in the same direction as the agent, the event direction, leftwards or rightwards, is determined solely by the agent. There is no 're-action' by the patient.

First indication that character orientation plays a role in scene perception is provided by Dobel, Gumnior, et al. (2007). The authors tested how accurately speakers of German identified the actants (humans, animals), objects and actions in line drawings when varying the presentation duration of the scene from 100 ms to 300 ms. Additionally, the participants in this study had to judge whether the scene was meaningful or not. The scenes were not presented in the center of the screen, but in the participants' visual periphery. Amongst other things, Dobel and colleagues found that correct identification of the agent hinged on the coherence of the scene: Agents were identified more accurately when the scene showed the agent and patient/recipient facing each other (coherent scene) than when the scene showed both agent and patient turned away from each other (incoherent scene). Based on their findings, the authors argue that agentivity appears to be related to a meaningful interaction between agent and patient/recipient and point to the potential influence of a character's orientation (see also Dobel et al., 2011). While the study of Dobel and colleagues gives a first indication that the orientation of the agent towards the patient affects role recognition, it is as yet unclear whether and how the orientation of the patient relative to the agent influences the perception and verbalization of events in which an agent acts upon a patient.

Given that orientation of head and body as well as body lean are critical event role features, we expect character orientation to be another visual property – next to spatial position – that affects scene perception and description: Speakers might produce more patient-initial sentences for face-to-face scenes because here the patient is more prominent and less prototypical as compared to scenes where both agent and patient have the same orientation. The effect of orientation might be even stronger if the patient is not only facing the agent but is also positioned to the left of the agent. To investigate how character orientation affects sentence planning and production,

---

[2] Note that in the materials the two conditions that were contrasted, the prototypical patient and the less typical patient condition, did not always differ in terms of the head/body orientation of the patient. That is, for some but not all events the patient was oriented away from the agent in the prototypical condition. Although indicative that orientation is a critical factor, it cannot be ruled out that outstretched extremities was a more important factor influencing role recognition.

and whether it may interact with character position, we conducted a scene description task. This production experiment was preceded by a non-verbal aesthetic judgement task to assess speakers' visual preferences for the simple transitive events that were displayed in our study. The material, data, and scripts for both tasks are available at the Open Science Framework (OSF), see Data Availability/Supplementary Files.

2. Pre-examination: Aesthetic judgement task

First, we aimed to know whether the orientation of agent and patient towards each other affects scene perception. If so, differences in character orientation should be reflected by differences in people's preferences for scenes as previously shown for character position (e.g., Esaulova, Dolscheid, Reuters et al., 2021; Maass et al., 2014). Here we used an aesthetic judgement task that included a subset of the scenes used in the main experiment. The experimental design and procedure resembled the ones in Esaulova, Dolscheid, Reuters et al. (2021), with the only exception that our aesthetic judgement task was not conducted as a paper-and-pencil but as a web-based experiment. An aesthetic judgement task has the advantage that it assesses speakers' perceptual preferences without the need to provide a verbal description of the scene(s) or to perform an additional task, such as drawing the event, which requires additional motor activity. If the orientation of agent and patient towards each other affected scene perception in a way that speakers prefer scenes in which the patient is facing the agent, as could be assumed following Dobel, Gumnior, et al. (2007), these scenes should be indicated as better than scenes in which both look in the same direction.

2.1 Materials and Methods

2.1.1 Participants

Thirty speakers who either learned German or German and one or more other languages between the ages of zero and six participated in the aesthetic judgement task (Mean age: 22 years, range: 18–63; 27 female; 3 lefthanded). Most of them were students at the University of Cologne or University of Erfurt. None of them reported knowledge of a language which deviates from the left-to-right, top-to-bottom writing system common to European languages. Two further speakers participated but were later excluded, because they reported to be familiar with other writing systems.

2.1.2 Materials

In the aesthetic judgement task, the participants were presented with two visual scenes on the screen simultaneously (see **Figure 1**) and had to indicate whether they preferred one scene over the other. In all scenes the agent was displayed to the left of the patient in line with German speakers' SAB. The scenes only differed in the orientation of agent and patient, that is, they displayed the same event between two human characters.
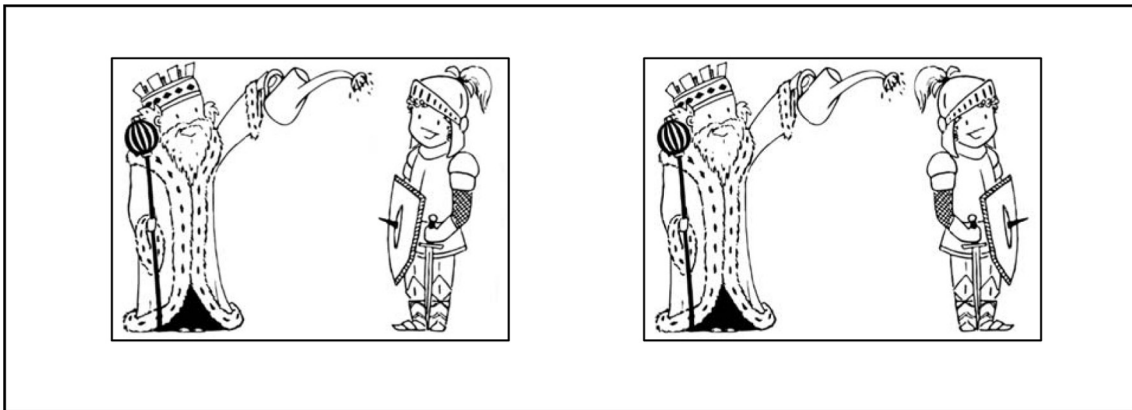
Figure 1: The patient faces the agent versus the patient looks in the same direction away from the agent. In the aesthetic judgement task, participants had to select the scene which they considered to be more natural (left or right) or, alternatively, had to indicate no preference.

There were eight different action events, resulting in eight contrasts: *fotografieren* 'to take a picture,' *filmen* 'to film,' *messen* 'to measure,' *bemalen* 'to paint,' *bürsten* 'to brush,' *gießen* 'to water,' *schlagen* 'to hit,' *schieben* 'to push.' The ordering of the face-to-face and same-direction scene in the contrast, left or right on the screen, varied. No fillers were used. Each participant saw all eight contrasts/items.

2.1.3 Experimental procedure

For the aesthetic judgement task, we used Google Forms. Participants had to indicate their age, gender, handedness, first language(s), and the languages they can read and write in. Information about data use was provided at the beginning of the questionnaire and participants had to agree to data use before they could start the experiment. In the questionnaire, each item (i.e., each contrast) was presented in a randomized order one below the other, and participants had to select one out of three options: (1) I prefer the left picture, (2) I prefer the right picture, (3) I have no preference. Participants were instructed to judge which picture was more typical, natural, or better than the other and to do so spontaneously. They were told that there was no correct answer. The task could be completed within 5–10 minutes.

2.1.4 Data analysis procedure

All responses in which participants indicated a preference for either the left or the right picture were coded as "away" (same-direction scene) or "facing" (face-to-face scene) responses. All other responses were "none" responses as participants indicated no preference. Since we wanted to know whether scenes in which agent and patient face each other are selected more often than the other response options, we analyzed the selection of the face-to-face scene (coded as 1 and 0) in a generalized linear mixed effects model (GLMM) with a logit link function. Since we were

interested in the intercept of the model, the model included subjects and items as random effects, but no further predictors as fixed effects.[3]

2.2 Results

In the aesthetic judgement task, participants had to judge which scene, agent and patient facing each other or agent and patient looking in the same direction (i.e., the patient looks away from the agent), they liked better. The bar chart in **Figure 2** shows the proportions of responses per event. There were 30 responses per event, 240 in total. The chart shows that, overall, participants preferred scenes with agents facing patients.
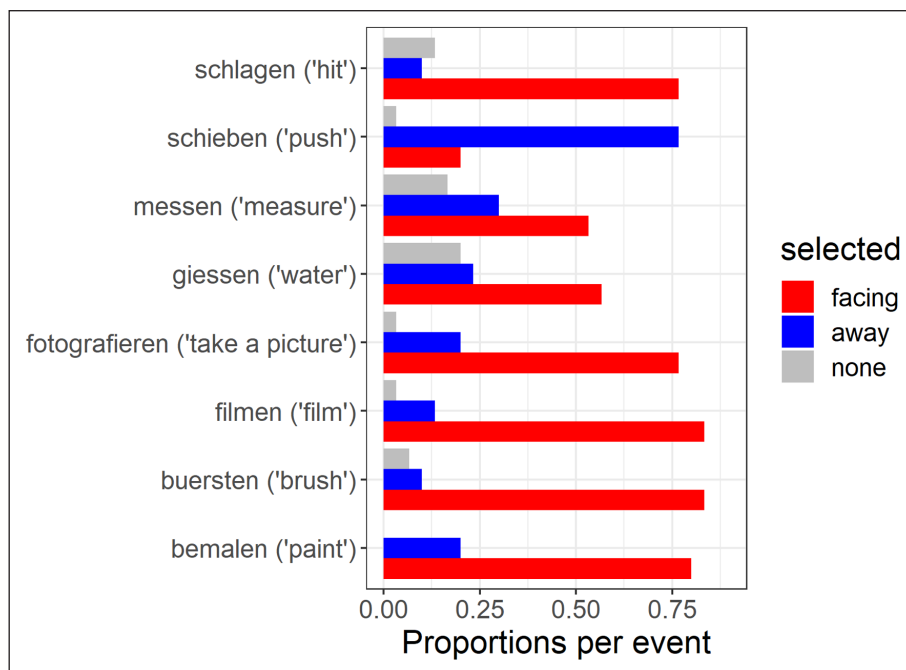

Figure 2: Bar chart showing the proportions of selected responses in the aesthetic judgement task.

The statistical analysis confirmed that participants preferred pictures with agents facing patients. When analyzing the frequency of selected "facing" responses, the GLMM showed a significant intercept in a positive direction (Est. $= 0.76$, SE $= 0.35$, $z = 2.17$, $p = 0.03$), indicating that this response option was selected more often than the two other response options together.

However, as visible in **Figure 2**, an exception to the face-to-face preference was the picture that depicted a 'pushing' event. For this item, participants preferred the scene in which the patient looked in the same direction as the agent, here the direction of the pushing motion. In a

---

[3] A model computed to check for influences of scene ordering showed no effect, so this factor was dropped.

statistical model excluding the 'pushing' event, the intercept value increased to 1.03 (SE = 0.23, $z = 4.43, p = < 0.001$).

2.3 Discussion

The non-verbal aesthetic judgement task demonstrated that the orientation of agent and patient towards each other affected German speakers' scene perception, at least for the eight different actions that were included in this task. When presented with the same two scenes that only differed in the orientation of agent and patient vis-à-vis each other, overall, speakers were found to display a preference for face-to-face scenes rather than indicating no preference or being equally likely to select the face-to-face or same-direction scene. A previous aesthetic judgement task (Esaulova, Dolscheid, Reuters et al., 2021) showed a preference for left-positioned agents over right-positioned agents. Thus, like character position, character orientation turned out to be a visual property that affected German speakers' aesthetic judgements. Yet, it should be pointed out that visual preferences were action specific.

For seven of the events, speakers of German judged scenes in which agents and patients were facing each other as more natural than scenes in which agents and patients looked in the same direction. This suggests that a preference evaluation of an event between two animate characters involves gaze orientation. The event 'pushing' in our materials, however, formed an exception. For this event, speakers preferred the scene in which the patient was looking in the same direction as the agent. Note, however, that in this case gaze direction corresponds to the direction of the motion. For this event, it is obviously more meaningful if both agent and patient look in the same direction. There are other events for which this is possibly also the case (e.g., chasing). Taken together, the current results suggest that agent-patient orientation is a visual property in scenes that affects their perception. The aim of our main experiment was to investigate whether the orientation of agent and patient vis-à-vis each other also affects the verbal description of the events, as has been shown to be the case for the visual factor position of the agent vis-à-vis the patient (see e.g., Esaulova et al., 2019; Pokhoday et al., 2019).

3. Main experiment: Scene description task

Given that agent-patient orientation affected speakers visual preferences, the main experiment, a scene description task, examined whether and how character orientation (facing, away) affected not only the perception but also the verbal description of transitive event scenes. Additionally, this experiment included the factor character position (left, right) and its potential interaction with character orientation.

In German, transitive events involving an agent and a patient, such as a king watering a knight, see example 1, can be described by producing either a canonical active SVO sentence (1a), a non-canonical active OVS sentence (1b), or a non-canonical sentence in passive voice (1c).

(1) a. *Der   König gießt   den   Ritter.*
     The._{NOM} king   waters the._{ACC} knight

  b. *Den   Ritter gießt   der   König.*
     The._{ACC} knight waters the._{NOM} king
     'It is the knight the king waters.'

  c. *Der   Ritter wird vom   König gegossen.*
     The._{NOM} knight is   by the king   watered
     'The knight is (being) watered by the king.'

Although all three structures are viable in German, previous production studies (Esaulova et al., 2019, 2020; Esaulova, Dolscheid, & Penke, 2021; Esaulova, Dolscheid, Reuters et al., 2021; Sauppe, 2017a, 2017b; Schlenter et al., 2022) have consistently shown that German speakers have a strong preference for canonical active SVO sentences when describing transitive event scenes. If they produce a non-canonical structure, German speakers produce passives rather than active OVS sentences (for more discussion, see Schlenter et al., 2022).

The aim of the scene description task was threefold. Previous scene description tasks were conducted in laboratory settings. This, however, was not possible during the corona pandemic. A first aim of our study, thus, was to replicate the effect of character position on language production in a web-based scene description task, for which we analyzed participants' syntactic choices and sentence-onset latencies via speech onset times (SOTs). The replication would not only indicate the suitability of the chosen procedure, but also enhance the credibility of findings related to character orientation collected via a web-based experiment. For the experiment, we employed the same cueing paradigm we used in a previous study (Schlenter et al., 2022, Exp. 2) as it proved to be suitable for the elicitation of patient-initial sentences and thus allowed us to analyze how the manipulated variables affected syntactic choices. If there were an effect of position and the design and paradigm worked as intended, we should find a difference between scenes with left-positioned patients and scenes with right-positioned patients, that is, an effect of character position on syntactic choice and sentence-onset latencies. In accordance with previous findings (Esaulova et al., 2019, 2020), we expected to elicit more patient-initial sentences for scenes with left-positioned patients because the patient occupies the perceptually prominent left position. We also expected to observe longer SOTs for such scenes because of a misalignment of the perceptual SAB in the presented event scene with a linguistically preferred agent-first SVO structure, which was found to delay the initiation of scene descriptions (e.g., Esaulova, Dolscheid, Reuters et al., 2021).

Our second aim was to test whether and how character orientation affects speakers' syntactic choices and sentence-onset latencies. If there were an effect of agent-patient orientation, as the results from the aesthetic judgement task suggest, then we should observe a difference between scenes with agents facing patients and scenes with agents and patients looking in the same

direction for syntactic choice and SOTs. Concretely, we expected to elicit more patient-initial sentences for face-to-face scenes than for same-direction scenes because of a difference in patient prominence. In face-to-face scenes, the patient should be perceived as more volitional. The addition of a feature associated with agentivity should increase the prominence of the patient, making it a suitable starting point for a descriptive utterance. The increase in patient prominence should also affect SOTs, either due to a prolongation of the time necessary to identify a character's thematic role in the event or due to a misalignment between the patient's prominence and the preference to describe the event with a canonical active SVO sentence.

Third, we were interested in a potential interaction between character position and character orientation. That is, we wanted to know whether the difference in the outcome between the two levels of the factor position depends on the level of the factor orientation.

3.1 Materials and Methods

3.1.1 Participants

Forty-four German native speakers participated in the scene description task (Mean age: 25 years, range: 18–59; 41 female; 6 lefthanded). They were either volunteers or students at the University of Cologne who participated for course credit. Eighteen additional speakers participated but were not included into the analyses because they reported to be early bilingual (9) or non-native speakers of German (3), or because of technical problems (2).[4] Four participants had to be excluded later because they did not fulfill our inclusion criterion of providing at least 80% analyzable responses.

3.1.2 Materials

The experimental item set comprised 32 event depictions that showed a human agent acting upon a human patient. The position of the agent and the orientation of agent and patient towards each other were manipulated such that there were four versions of each item, resulting in a total of 128 visual scenes; for an example, see **Figure 3**. The items were distributed equally across four presentation lists using a Latin square design, so that each participant encountered an item only once in one of the four conditions (8 items per condition).

The design of the study was adopted from Schlenter et al. (2022). All visual scenes were preceded by a picture of the patient in the same position and with the same orientation as in the scene. The patient picture served as a perceptual cue and was presented for 700 ms prior to scene onset. Previous scene descriptions tasks observed a strong preference in German-speaking participants to describe transitive event scenes with active SVO sentences (Esaulova et al., 2019, 2020; Esaulova, Dolscheid, & Penke, 2021; Esaulova, Dolscheid, Reuters et al., 2021; Sauppe,

---

[4] One participant only submitted empty audio files. For another participant, there was no record in the results file, probably because s/he closed the browser too early, so only audio files had been transferred.
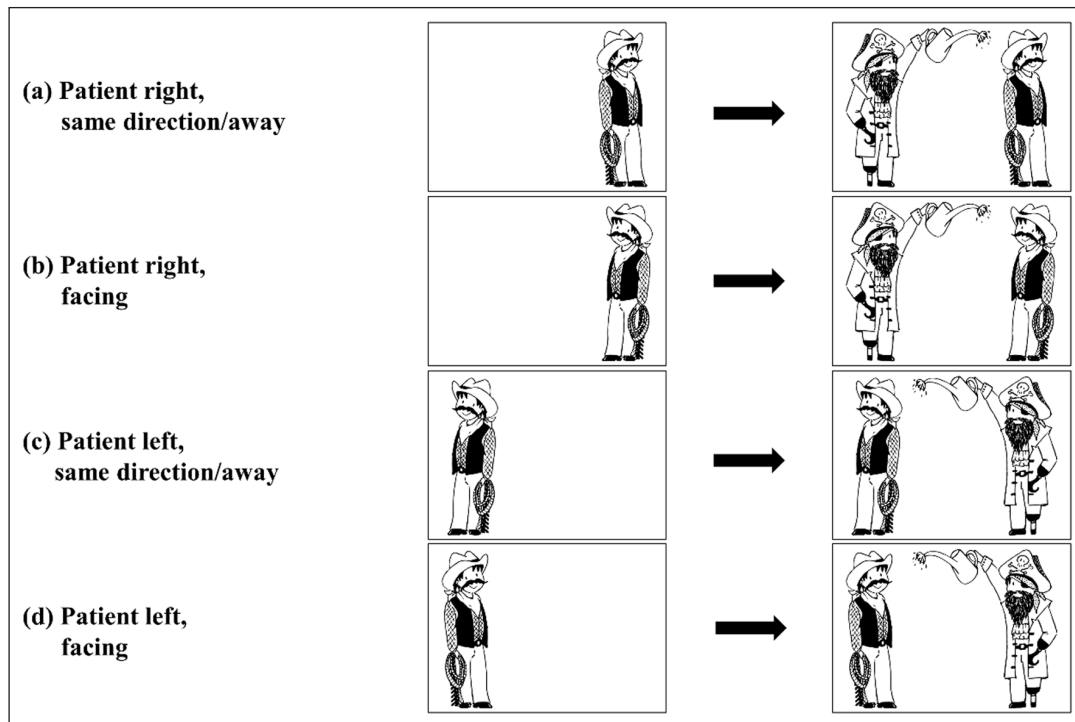
Figure 3: An example for an experimental item in all four conditions in the scene description task. Each scene was preceded by a patient preview.

2017a, 2017b; Schlenter et al., 2022). Since we were interested in evaluating the number of patient-initial utterances in relation to the visual properties position and orientation, we adopted the methodology of referential cueing to avoid floor effects. Referential patient-cueing has been shown to effectively increase the proportion of patient-initial utterances in scene description tasks in comparison to no-cueing or other cueing conditions (Esaulova et al., 2020; Myachykov et al., 2018; Schlenter et al., 2022).

For the scenes, only male characters were used as agent and patient, first, to control for effects of sex/gender and, second, because only masculine nouns in German are unambiguously marked as subject (nominative case) and object (accusative case). All characters, 16 in total, were portrayed with stereotypical features (e.g., lasso, sword) so they could be easily identified as cowboy, knight etc. Each of the characters appeared twice as an agent and twice as a patient in different action events. We used the same eight events as in the aesthetic judgement task, that means that each event was shown four times, but each time with different characters. Agents were always depicted with outstretched extremities, while patients were neutral in this regard.

The experimental items were interspersed among 32 fillers that consisted of 16 intransitive events (e.g., the woman cries, the ice cream melts) and 16 locative scenes showing two objects in different spatial arrangements (e.g., a ball underneath a table). Each locative scene appeared twice, once with a preview of the first object (e.g., ball) and once with a preview of the second

object (e.g., table). The intransitive event scenes were preceded by a blank screen. The fillers were added to prevent participants from developing a response strategy.

3.1.3 Experimental procedure

The scene description task was built with PennController 1.9 and run on PCIbex Farm (Zehr & Schwarz, 2018). Participants received a link to the experiment and took part from home. They were automatically assigned to a list using PCIbex' internal counter, which incremented when a participant opened the link. After giving informed consent and answering the participant questions (age, gender, handedness, only German as a language between 0 and 6 years: yes/no), participants were provided with an example scene (a diver hitting a sack) and three different options of how to describe the scene in one sentence (active SVO, active OVS, passive). They were told to describe the picture presented to them as soon as a respective message appeared. They were reminded to respond spontaneously. Before the actual experiment started, participants had to grant access to their microphone. Each trial started with a fixation cross in the center of the screen presented for a duration of 500 ms, followed by a 700 ms preview cue (blank screen for the intransitive fillers). A 500 ms blank screen was inserted between preview and scene to avoid an animation effect. The visual scene was shown on the screen for a duration of 6000 ms together with the message 'Describe the picture.' Afterwards, a button appeared at the top of the screen, saying 'Press here to continue.' Thus, participants could decide when to continue to the next trial. A progress bar was shown at the top of the screen. **Figure 4** illustrates the procedure for an experimental trial. The order of trials was fully randomized on a by-participant basis by the experimental software.
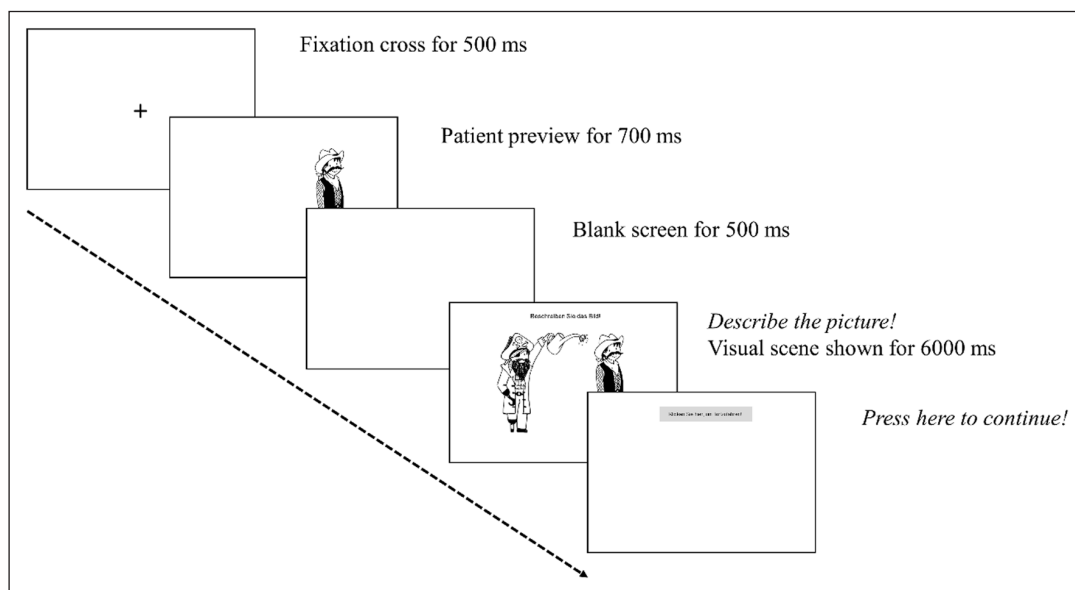

Figure 4: Illustration of the procedure for an experimental trial in the scene description task.

Throughout the experiment, participants were always able to see when the system recorded their voice. All audio files recorded during the experiment were transferred to a server at the University of Cologne. The whole experiment could be completed within less than 15 minutes. Prior to data collection, this experiment was pre-registered on OSF.

3.1.4 Data analysis procedure

The responses in the scene description task were coded as agent-initial and patient-initial. Utterances in which participants did not start with the agent or the patient or utterances which did not describe a transitive event between two animate characters (e.g., *Ein König und ein Vampir stehen sich gegenüber* 'A king and a vampire are standing opposite to each other') were excluded from the analyses. We also excluded utterances with sentence-initial corrections if these altered the syntactic structure of the sentence as well as sentences that were incomplete and where relevant information for coding was missing. We did not exclude utterances in which participants incorrectly identified a character as female and thus used a feminine noun because previous research did not report any influence of ambiguity in case marking (Schlenter et al., 2022). Since the other character was referred to with a masculine noun and thus unambiguously marked as subject (nominative) or object (accusative), we were still able to code these responses. We excluded sentences in which participants falsely used an inanimate noun (e.g., the statue of the king) since animacy has been identified as a factor that influences speakers' syntactic choices (e.g., Esaulova et al., 2019). Together with null responses this amounted to 190 observations that had to be removed, leaving 1218 analyzable responses. The utterances were coded by a research assistant and the first author and cases of doubt were discussed among the whole research team. To determine sentence-onset latencies for the analyzable responses, we used the software Praat (Boersma, 2001).[5]

For statistical analyses, we computed mixed-effects models in R (R Core Team, 2021, Version 4.1.2), using the *lme4* package (Bates et al., 2015, Version 1.1.27.1). To obtain p-values, we used the *lmerTest* package (Kuznetsova et al., 2017, Version 3.1.3). We started with the maximally specified model (Barr et al., 2013) and successively removed random slopes if the maximal model failed to converge. The best-fitting converging model was selected based on the lowest AIC value (Matuschek et al., 2017). For the syntactic choice data that had a binary-coded dependent variable, a GLMM with a logit link function was computed. Pairwise comparisons were conducted with the help of the *emmeans* package (Lenth, 2022, Version 1.7.2), which adjusts p-values using the Tukey method for multiple comparisons.

---

[5] Note that we analyzed compressed audio files (mp3) that were recorded by participants' own hardware devices. Since we used a within-subjects design, any variability due to differences in hardware should not affect the overall pattern of results. In fact, the SOTs measured in the current study fall within the range of SOTs reported in previous studies that employed this experimental methodology in the lab (Esaulova et al., 2020; Schlenter et al., 2022).

3.2 Results

3.2.1 Syntactic choice

Overall, cueing the patient by means of a preview led to 27% patient-initial utterances, all of which were passives. Descriptively, most passives were elicited when the patient appeared in the prominent left position and when the patient was facing the agent (32.1% for patient left, facing). Conversely, the least number of passives was elicited when the patient appeared in the right position and had turned its back toward the agent (21.2% for patient right, same direction/away). We elicited 27.8% passives for the patient-left/away condition and 27.7% passives for the patient-right/facing condition.

For statistical analyses, deviation coding (0.5, –0.5) was used for the factors position and orientation. The best-fitting model that converged, the model with random intercepts for subjects and items, revealed a main effect of position (Est. = –0.59, SE = 0.19, z = –3.13, p = 0.002) and a main effect of orientation (Est. = –0.56, SE = 0.19, z = –2.94, p = 0.003) on syntactic choice, but no significant interaction (Est. = –0.36, SE = 0.38, z = –0.96, p = 0.34). For better illustration, the probabilities predicted by the model are shown in **Figure 5**. Pairwise
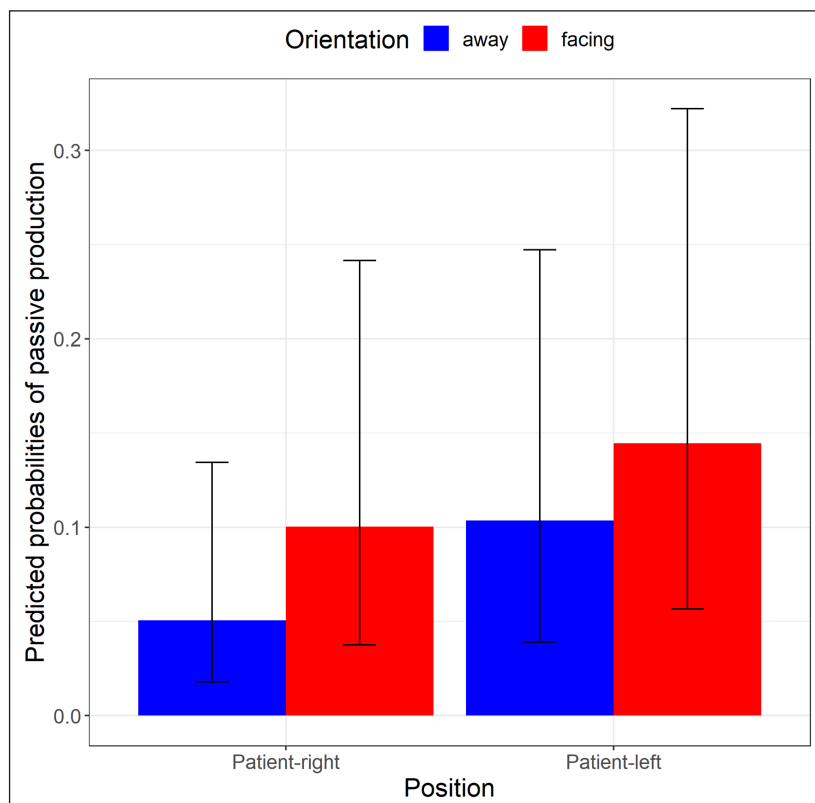

Figure 5: Predicted probabilities of passive production and confidence intervals for right- and left-positioned patients depending on orientation. Blue bars correspond to the conditions with patients oriented away from the agent, red bars to conditions with patients facing agents.

comparisons showed that the number of patient-initial responses for the patient-right/away condition differed significantly from all other conditions, which did not differ from each other. Taken together, in our frequentist model we found that both visual properties – position and orientation – affected German speakers' syntactic choice but did so independently. Placing the patient left (vs. to the right of the agent) as well as displaying the patient oriented towards the agent (vs. looking in the same direction away from the agent) increased the production of passives by 5%.

To address a reviewer's concern, who notes that a more maximal model should be preferred, we also calculated a Bayesian GLMM for our syntactic choice data, using the R package *brms* (Bürkner, 2018). For frequentist models, the maximal random effects structure often leads to non-convergence, a problem that does not apply to Bayesian models. We used weakly informative priors to fit the maximal model. The full code is available on OSF. This model showed that the posterior probability of the position effect (Est. = –0.78) being below zero was 0.99, with a 95% credible interval that did not include zero [–1.46, –0.23]. Similarly, a one-sided test showed that the posterior probability of the orientation effect (Est. = –0.61) being below zero was 0.96, however, here the 95% credible interval included zero [–1.32, 0.07]. Thus, although the effects are small, there is evidence for the existence of these effects in the reported direction. Again, we have no compelling evidence for the existence of an interaction effect, with the coefficient of –0.38 having a wide credible interval including zero [–1.44, 0.74].

In an exploratory analysis of the syntactic-choice data, we focused on the items including a 'pushing' event to see whether they fit into the overall pattern or deviate from it, as may be expected from the aesthetic judgement task. For 'push' items (164 observations), participants produced 23.4% passives for the patient-right/away condition, 39% passives for patient-right/facing, 36.6% passives for patient-left/away and 37.1% passives for patient-left/facing. Thus, descriptively, the data do not point to strong differences between 'push' items and the overall pattern of results.

3.2.2 Speech onset times

Next, we analyzed speakers' SOTs that indicate how the factors position and orientation modulated sentence planning, by slowing down or facilitating the production of active or passive voice. Descriptively, the longest mean SOTs were measured for active utterances when agent and patient were facing each other, with no visible difference between left-positioned patients and right-positioned patients. Overall, speakers were faster in producing passives (Mean: 1535 ms, SD: 557 ms) than actives (Mean: 1779 ms, SD: 573 ms).[6]

As for syntactic choices, deviation coding (0.5, –0.5) was used for the factors position and orientation in a linear mixed-effects model. Moreover, we added the factor voice choice (active:

---

[6] For an overview of all means and standard deviations per condition, see the Supplementary File on OSF.

0.5, passive: –0.5) and its interaction with position and orientation to the model analyzing SOTs. SOTs were log-transformed to approach a normal distribution. The best-fitting converging model included voice as by-subject and by-item random slope. The output of this model is given in **Table 1** and shows significant effects of orientation and voice, as further illustrated in **Figure 6**, which shows the back-transformed SOTs predicted by the model.

	Estimate	**Std. Error**	**t-value**	**p-value**
(Intercept)	7.399	0.030	245.53	<0.001
Position	–0.005	0.017	–0.314	0.753
Orientation	–0.054	0.017	–3.193	0.001**
Voice	0.098	0.042	2.327	0.027*
Position:Orientation	0.000	0.033	0.002	0.998
Position:Voice	0.014	0.033	0.427	0.669
Orientation:Voice	0.052	0.034	1.554	0.120
Position:Orientation:Voice	–0.000	0.066	–0.003	0.998

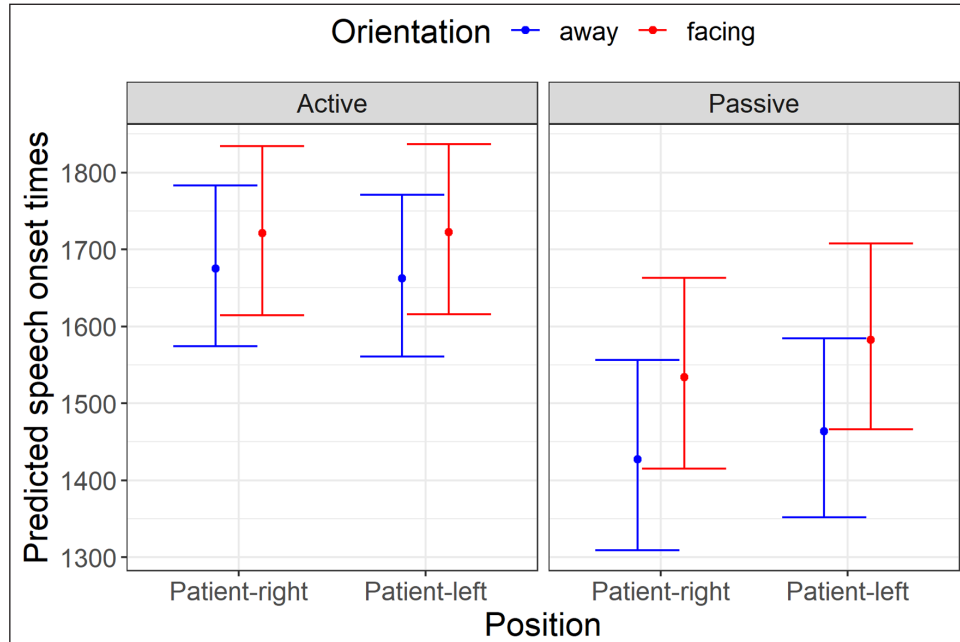**Table 1:** Output of the model analyzing (log-transformed) SOTs.

Figure 6: Predicted speech onset times and confidence intervals for active (left) and passive voice utterances (right) for right- and left-positioned patients depending on orientation. Blue lines correspond to the conditions with patients oriented away from the agent, red lines to conditions with patients facing agents.

A Bayesian model with the maximal random effects structure computed with the SOT data shows similar estimates for the coefficient of orientation (Est. = –0.05) and voice (Est. = 0.10). For orientation, the 95% credible interval does not include zero [–0.09, –0.01], the estimated posterior probability of the effect being below zero is 0.99. For voice, the 95% credible interval does not include zero [0.01, 0.19], the estimated posterior probability of the effect being above zero is 0.99. There is no compelling evidence for an effect of position on SOTs (Est. = –0.01) with a credible interval that includes zero [–0.05, 0.03] and an estimated posterior probability of the effect being below zero that is 0.64. Neither is there compelling evidence for any interactions between factors.

To sum up, analyses of SOTs showed that, overall, our speakers needed less time to prepare a patient-initial passive sentence than an agent-initial active sentence. Moreover, sentence-onset latencies differed between agents and patients facing each other and agents and patients looking in the same direction. Speakers were faster for scenes in which both agent and patient looked in the same direction than for scenes in which agent and patient faced each other.

3.3 Discussion

In a production experiment, we tested how the orientation of agent and patient together with the position of agent and patient affected scene description. For this, we presented speakers of German with an explicit visual cue, a preview of the patient, prior to the scene they had to describe. The patient referent was shown at the same position and with the same orientation as in the upcoming scene. Thus, the referential cue was not only lexically informative but also ultimately informative as regards visuo-spatial properties that may affect scene perception and forthcoming scene description. The scene description task with prior patient-cueing yielded evidence for a modulating role of both position and orientation on syntactic choice: When the patient was placed in the prominent left position, more passives were produced. Similarly, more passives were produced in the current experimental setting when the patient in the scene was facing the agent as compared to scenes in which agent and patient looked in the same direction. However, we found no significant interaction between both factors. Although for seven of the events displayed in the experiment speakers had shown a preference for scenes in which agent and patient were facing each other (aesthetic judgement task), face-to-face scenes in comparison to same-direction scenes were the ones that needed the longest planning times. The position of agent and patient did not affect sentence-planning times, at least there was no significant effect.

The effect of position on syntactic choice is well in line with previous studies showing that German speakers' SAB modulates their scene descriptions (e.g., Esaulova et al., 2019, 2020). When instructed to describe a scene in which the patient appears at the perceptually more prominent position left of the agent, German speakers were more likely to produce a non-canonical passive sentence. The 5% increase of patient-initial sentences was similar to the increase reported for

previous laboratory experiments that made use of attentional cueing (Esaulova et al., 2020; Pokhoday et al., 2019), indicating that the prominence-lending factor position also affects language production outside of the lab.

The novelty in our study was the manipulation of character orientation. Previous studies on scene apprehension had shown that speakers had more difficulty to identify event roles if those were not prototypical (Dobel, Gumnior, et al., 2007; Hafri et al., 2013). According to Hafri et al. (2013) the event role features head orientation, body orientation and body lean may modulate the degree of volition and therefore agentivity. Hence, we had assumed that if the patient looks toward the agent, the patient may be perceived as more prominent (since less prototypical) as if the patient has turned its back toward the agent. Additionally, we had suspected that if both characters look in the same direction, the agent may be more likely to become the starting point of the action and thus the sentence-initial element. This is exactly what we found. There were more patient-initial utterances for face-to-face scenes than for same-direction scenes. In fact, the effect of agent and patient orientation on syntactic choice was comparable to the – by now well established – effect of agent and patient position. We take our findings as evidence that agent and patient orientation modulates how speakers attend to the agent and patient in a scene, and which one is selected as the sentence-initial subject. Moreover, SOT analyses showed that speakers of German needed more time to describe scenes in which agents and patients were facing each other than scenes in which agent and patient looked in the same direction. It might be the case that participants needed more time to identify who of the two referents was the agent and who the patient in the depicted event. In addition, we may speculate that the less prototypical and the more prominent the patient, the more competition exists between alternative sentence structures. Thus, longer sentence-onset latencies may reflect not only difficulties in role identification, but also the decision process between a patient-initial passive sentence that would align the perceptually prominent patient with the functionally prominent subject-position and a canonical active sentence. We thus conclude that the orientation of agent and patient not only modulates how speakers perceive a transitive event scene, but orientation can also modulate sentence planning and syntactic choices.

Across experimental conditions, referential cueing led to 27% passives, indicating that the cueing manipulation in our web-based experiment was equally effective in guiding attention toward the patient referent as in a previous lab experiment (Schlenter et al., 2022). Moreover, like in previous lab experiments (Esaulova et al., 2020; Schlenter et al., 2022), referential patient-cueing facilitated the production of passives also in terms of sentence-planning time. Previous studies found that after participants were presented with a preview of the patient, either in the center of the screen (Esaulova et al., 2020) or in the position the patient was to appear in the subsequently presented scene (Schlenter et al., 2022), participants were faster in producing passives than actives. At first sight, this finding seems to be in contrast with the often-expressed assumption that passives should be associated with more processing costs and

thus take longer to be produced due to their deviance from canonical linking and to their lower frequency (e.g., Sauppe 2017a for German). However, referential patient-cueing can ease the production of passives as compared to actives. As mentioned in the beginning of this section, a referential cue is lexically informative. The duration of cue presentation (700 ms) gave speakers enough time to select a lemma for the patient referent. German nouns have a grammatical gender (masculine, feminine or neuter) that is expressed on the determiner preceding the noun. According to current psycholinguistic models, activation of a noun's lemma activates this gender information in the mental lexicon and a determiner is selected accordingly (Schiller & Caramazza, 2003; Schriefers et al., 2002). As German determiners fuse gender, number and case information, the activated determiner always expresses case information. In the default case, this is nominative case (Emonds, 1985). Thus, referential cueing of the patient results in the activation of a noun phrase marked for default nominative case. Given the activation of this noun phrase, the easiest way to proceed with the utterance is to go on with a passive clause where the patient is realized as the nominative case-marked subject (i.e., *Der Cowboy*$_{[NOM]}$ *wird vom Piraten gegossen* 'The cowboy is being watered by the pirate'). The proposed strategy is in line with the 'Principle of Immediate Mention' by Ferreira & Dell (2000, p. 299) according to which "production proceeds more efficiently if syntactic structures are used that permit quickly selected lemmas to be mentioned as soon as possible." The suggested process also conforms to the proposal of Gleitman and colleagues (2007) that linguistic encoding of a message proceeds in a linear incremental manner, building-up from the first lemma retrieved in the mental lexicon that then constrains the structure of the sentence to be produced. In contrast, the longer SOTs for actives after patient-cueing possibly reflect the time needed to inhibit the activated patient lemma and select a lemma for the agent to initiate an active SVO sentence.

4. General Discussion

This study, comprising a non-verbal aesthetic judgement task and a scene description task, showed that the orientation of the agent and patient towards each other in a transitive event scene affects how speakers of German perceive and describe such an event. Based on the results from the scene description task (main experiment), we concluded that orientation, next to the position of agent and patient in a scene, is another factor that modulates the prominence of the patient and thus speakers' sentence planning. Although previous research by Dobel, Gumnior, et al. (2007) has pointed towards a role of agent and patient orientation, here on role recognition after brief exposure, it has not been systematically investigated so far. Moreover, we replicated the findings from previous studies indicating that referential patient-cueing facilitates the production of passives in German as measured by speech onset times (Esaulova et al., 2020; Schlenter et al., 2022); a detailed explanation is provided in the discussion section of the main experiment. In the following, we will pinpoint how a manipulation of orientation can help in overriding the strong preference

for producing active SVO sentences in picture-description tasks in German (4.1). We will further discuss the possible discrepancies between perception and description in our study (4.2) and point to at least one other visual property that waits for an experimental investigation (4.3). Finally, we will discuss how the visual properties which we found to play a role in sentence production may affect other experimental designs (4.4) and discuss the limitations of our study (4.5).

4.1 Agent and patient orientation affects scene perception and description

In the scene description task, the prominence of the patient was promoted in three ways: by patient-cueing and by the position and orientation of the patient relative to the agent. In line with previous research, we found that patient-cueing facilitated the production of patient-initial utterances, as indicated by reduced SOTs for passives as compared to actives. Moreover, we replicated an effect of position on syntactic choice. We elicited more passives for scenes with left-positioned patients than for scenes with right-positioned patients. The novelty in our study was the manipulation of the orientation of the patient relative to the agent. We found that speakers of German produced more patient-initial sentences for scenes in which the patient was facing the agent as compared to scenes in which both looked in the same direction. Following Hafri et al. (2013), we argue that even for semi-realistic drawings the orientation of the patient relative to the agent can change the prominence of the patient: If a patient appears more volitional and thus more agentive, the more likely it is that the patient becomes the sentence-initial subject. Hence, similar to position, orientation is another visual prominence-lending factor that affects scene perception and forthcoming scene description. Moreover, the longer SOTs for face-to-face scenes as compared to same-direction scenes indicate that the increase in patient-prominence affected sentence-planning times. This increase might be due to delays in identifying the agent and patient of the depicted action (i.e., thematic role identification). The difference in timing might also result from competition between alternative sentence structures (active SVO, passive) used by speakers of German to express differences in the prominence of agent and patient linguistically. Summarizing, our data show that character position and character orientation are two visual properties in scenes that affect the prominence of characters in the scene, the identification of thematic roles and grammatical role assignment.

4.2 Preferred scenes are not necessarily easier to process

At first sight, it might seem puzzling that a one-sentence description for face-to-face scenes needed longer planning than a description for same-direction scenes, although face-to-face scenes with only one exception were preferred in a non-verbal judgement task. Moreover, although this preference in perception was flipped for the 'pushing' event, speakers were similarly affected by the manipulation of orientation and produced more passives for face-to-face scenes than for same-direction scenes. We attribute this effect of orientation to the increase in patient prominence (see 4.1). The results from both tasks indicate that aesthetic preferences do not necessarily affect

the verbalization of events. Let us take as an example from the current study a scene that shows the agent filming the patient from behind: Speakers of German judged this scene as less natural than the same scene that showed the agent filming the patient while the patient looked at the agent. However, when instructed to describe the scenes, speakers of German needed less time to produce a description for such dispreferred scenes than for the preferred scenes. In other words, at least in the current scene description task with prior patient-cueing, 'better' scenes were not necessarily easier to verbalize. Future research might want to investigate the effect of orientation in a lab experiment with concurrent eye-tracking to see whether the difference between face-to-face and same-direction scenes also shows up in saccades to or fixations on agent and patient, as shown for example by Esaulova et al. (2019) for left-positioned patients in comparison to right-positioned patients. So far, we can tell from our syntactic choice data and from speech onset times that the orientation of agent and patient makes a difference.

4.3 Further visual properties that may affect sentence production: affectedness

By investigating how the position and the orientation of agent and patient affect the description of transitive event scenes, we focused on two visual-spatial properties. There are other visual properties that may influence how speakers perceive and describe scenes. One such factor could be *affectedness* of the patient. Kretzschmar et al. (2019, pp. 117–119) mention *affectedness* as a prominent proto-patient feature, which likely facilitates certain types of passivation in German. If we assume that affected patients stand out, as Kretzschmar and colleagues suggest, there should be more patient-initial sentence descriptions for scenes with affected patients as compared to scenes with neutral patients all other things being equal. Visually, this could be realized by showing the patient of an action in the affectedness-condition in a different way, for example, for an action like 'painting' the patient could be shown with color stains. Note that the increase of patient-initial utterances for face-to-face scenes in the current study was probably the result of two steps: the orientation of the patient toward the agent let the patient appear more agentive, because orientation toward another character comes along with a higher degree of volition, and as a result the patient became more prominent. Similarly, for position: if the patient appeared to the left of the agent and thus occupied the position typically associated with the agent, the patient became more prominent. In both cases, prominence is related to agentivity. An investigation of *affectedness* could help to clarify whether a visual property not associated with agentivity but with the patient can also modulate scene perception and description.

4.4 Implications for other experimental designs

In the production task with prior patient-cueing that we described here, we investigated how agent and patient position and agent and patient orientation in a visual scene affect sentence production. Although effects might be more subtle in other experimental settings, we argue that the findings from this study have implications for psycholinguistic research in more general and are relevant

for a range of tasks that use visual materials. Since the SAB, although to varying degrees (Suitner et al., 2021), appears to be quite persistent in literate speakers of European languages, we need to assume that experiments with visual materials run at risk when not controlling for agent and patient position and orientation in their design. On the other hand, the purposeful application of the prominence-lending factors position and orientation in experimental designs may also serve to facilitate, for example, sentence-picture matching. Imagine you are being instructed to find a match for the sentence "The prince is being hit by the vampire" while looking at two scenes displaying a competitor and a target picture; see **Figure 7**. Matching might be easier, faster, and more accurate if the target picture (b) is presented as compared to (a). The alignment of a visually depicted event scene with the linearization of subject and object in the sentence might especially foster the comprehension of language-impaired persons. It could also facilitate the production of specific syntactic structures in therapeutical settings. The elicitation of non-canonical sentences might be more easily achieved by increasing the visual prominence of the patient by placing it in a visually more prominent position or by changing its orientation vis-à-vis the agent.
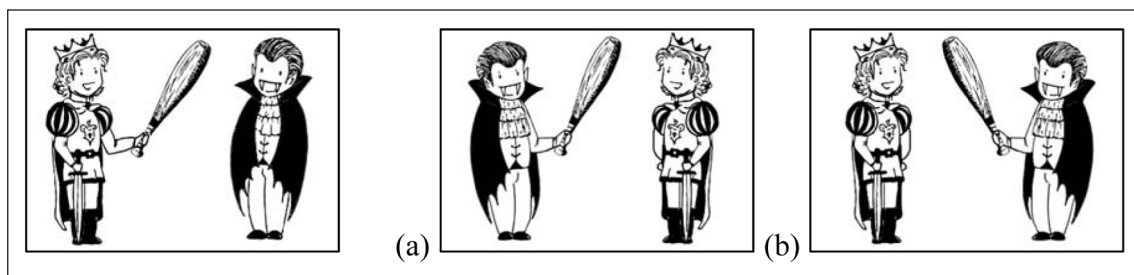

Figure 7: Competitor (left) and possible target pictures (mid, right) in a sentence-picture matching task. Target picture (b) is in alignment with the linearization of subject and object in a passive clause if the prince corresponds to the patient of the action and the vampire to the agent.

The factors position and orientation might also be relevant for syntactic priming studies that use prime pictures together with a verbal description, for example a passive clause (e.g., Messenger et al., 2012). Participants may be more likely to repeat the syntactic structure if it aligns conceptually with prime and target picture.

To conclude, we would like to stress that subtle changes of visual properties may easily turn to confounds in an experiment. Hence, visual properties should be controlled for as best as possible. More research on stimulus-driven attention orientation is needed to shed more light on its role in language production and comprehension.

4.5 Limitations

This study comes with several limitations. First, our study only included eight different action events. Further research is needed to see how our findings generalize to other events. We cannot rule out that other action events have other orientation preferences. Note, however,

that, as discussed in 4.2, aesthetic preferences and ease of verbalization do not necessarily go hand in hand. Moreover, our materials were restricted to black-and-white line drawings of fictional characters. These characters were shown with the same distance between each other as some of these drawings were previously used in eye-tracking experiments (same-direction scenes only). Using the same or similar scenes allowed us to compare our results to previous findings. However, we cannot say whether results might change when, for example, there is contact between characters or when non-fictional characters or photographs are shown. Last, but not least, we used an experimental procedure that directed speakers' attention towards the patient. As mentioned previously, effects of position and orientation might be more subtle in other experimental designs. Yet, the finding that speakers of German display a preference for certain scenes when asked to judge their naturalness and that they show differences in speech onset times and syntactic choices for a larger set of the same scenes, to us suggests that character orientation is an interesting and as yet neglected visual property in scenes.

5. Conclusion

This study, for the first time, showed that the orientation of the patient relative to the agent in a transitive event scene (face-to-face vs. same direction/away) influenced scene perception and description. We elicited more patient-initial (passive) sentences in German for face-to-face scenes than for same-direction scenes. In addition, speakers needed more time to plan a one-sentence description for face-to-face scenes. The current study also replicated an effect of character position on speakers' syntactic choice in a web-based experiment. Previous research has shown that for literate speakers of European languages, the left position in a scene typically corresponds to the position of the agent, in alignment with a preference for agent-initial active sentences. In a web-based scene-description task with prior patient-cueing, we elicited more patient-initial sentences for left-positioned patients than for right-positioned patients. Both the effect of orientation and position can be explained by a change in visual prominence. We argue that the increase in prominence for patients facing agents results from an association with volition and thus agentivity. Similarly, left-positioned patients become more prominent because they occupy the position typically associated with the agent. To conclude, patient orientation, like the position of the patient relative to the agent, can change the prominence of the patient in a transitive event and, as a result, how the event is perceived and described.

Data accessibility statement

The materials, data and scripts for this study are available at OSF under https://osf.io/vqc64/ (DOI 10.17605/OSF.IO/VQC64). The project named 'VisPro' comprises the following three components:

(1) Materials

AllScenes.pdf

This file provides on overview of all the experimental items that were used in the study.

(2) Data

Data_AestheticJudgementTask.csv

Data_SceneDescriptionTask.csv

This component contains two csv files. The file 'Data_AestheticJudgementTask.csv' includes the coding of participants' choices in the aesthetic judgement task. The file 'Data_SceneDescriptionTask.csv' includes participants' verbal responses and the coding of utterances (agent-initial, patient-initial) in the scene description task as well as the speech onset times in milliseconds.

(3) Analysis scripts

AestheticJudgement.Rmd, AestheticJudgement.pdf

SceneDescription_Bayesian_analyses.Rmd, SceneDescription_Bayesian_analyses.pdf,

SceneDescription_SpeechOnsetTimes.Rmd, SceneDescription_SpeechOnsetTimes.pdf

SceneDescription_SyntacticChoice.Rmd, SceneDescription_SyntacticChoice.pdf

This component contains all scripts (R Markdown and pdf output) used to analyze the data from both tasks.

The scene description task was pre-registered on the 26th of May 2021. The pre-registration is available under https://osf.io/5zsdj.

Ethics and consent

All procedures were in accordance with the Declaration of Helsinki. Prior to the experiments, informed consent was obtained from the participants.

Acknowledgements

First of all, we thank Barbara Zeyer, who set up the aesthetic judgement task, pre-processed the data for both the aesthetic judgement and the scene description task and helped with the coding

of participants' responses. Moreover, we like to thank Cecilia Puebla Antunes for drawing the picture stimuli for us and Maximilian Hörl for supporting us with the statistical analyses. Finally, we like to thank the reviewers for their helpful comments. This study was funded by the *Deutsche Forschungsgemeinschaft* (DFG, German Research Foundation) – Project-ID 281511265 – CRC 1252 Prominence in Language.

Competing interests

The authors have no competing interests to declare.

Author contributions

JS was responsible for the conceptualization and implementation of the study, as well as the data analysis and data curation. JS wrote the manuscript. MP supervised the research and reviewed and edited the manuscript. Both authors contributed to the article and approved the submitted version.

References

Altmann, L. J. P., Saleem, A., Kendall, D., Heilman, K. M., & Rothi, L. J. G. (2006). Orthographic directionality and thematic role illustration in English and Arabic. *Brain and Language, 97*(3), 306–316. DOI: https://doi.org/10.1016/j.bandl.2005.12.003

Bader, M., & Häussler, J. (2010). Word order in German: A corpus study. *Lingua, 120*(3), 717–762. DOI: https://doi.org/10.1016/j.lingua.2009.05.007

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. DOI: https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Bock, K., Irwin, D., & Davidson, D. J. (2004). Putting first things first. In J. M. Henderson & F. Ferreira (Eds.), *The Interface of language, vison, and action* (pp. 249–278). Psychology Press.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5*(9/10), 341–345.

Bürkner, P. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal, 10*(1), 395–411. DOI: https://doi.org/10.32614/RJ-2018-017

Buswell, G. T. (1935). *How people look at pictures*. The University of Chicago Press.

Chatterjee, A., Southwood, M., & Basilico, D. (1999). Verbs, events and spatial representations. *Neuropsychologia, 37*(4), 395–402. DOI: https://doi.org/10.1016/S0028-3932(98)00108-0

Dobel, C., Diesendruck, G., & Bölte, J. (2007). How writing systems and age influence spatial representations of action: a developmental, cross-linguistic study. *Psychological Science, 18*(6), 487–491. DOI: https://doi.org/10.1111/j.1467-9280.2007.01926.x

Dobel, C., Enriquez-Geppert, S., Zwitserlood, P., & Bölte, J. (2014). Literacy shapes thought: the case of event representation in different cultures. *Frontiers in Psychology*, *5*, 290. DOI: https://doi.org/10.3389/fpsyg.2014.00290

Dobel, C., Glanemann, R., Kreysa, H., Zwitserlood, P., & Eisenbeiß, S. (2011). Visual encoding of meaningful and meaningless scenes. In J. Bohnemeyer & E. Pederson (Eds.), *Event representation in language and cognition* (pp. 189–215). Cambridge University Press.

Dobel, C., Gumnior, H., Bölte, J., & Zwitserlood, P. (2007). Describing scenes hardly seen. *Acta Psychologica*, *125*(2), 129–143. DOI: https://doi.org/10.1016/j.actpsy.2006.07.004

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*(3), 547–619. DOI: https://doi.org/10.1353/lan.1991.0021

Emonds, J. E. (1985). *A unified theory of syntactic categories*. Foris Publications. DOI: https://doi.org/10.1515/9783110808513

Esaulova, Y., Dolscheid, S., & Penke, M. (2021). All it takes to produce passives in German. In V. Torrens (Ed.), *Syntax processing* (pp. 75–107). Cambridge Scholars Publishing.

Esaulova, Y., Dolscheid, S., Reuters, S., & Penke, M. (2021). The alignment of agent-first preferences with visual event representations: Contrasting German and Arabic. *Journal of Psycholinguistic Research*. DOI: https://doi.org/10.1007/s10936-020-09750-3

Esaulova, Y., Penke, M., & Dolscheid, S. (2019). Describing events: Changes in eye movements and language production due to visual and conceptual properties of scenes. *Frontiers in Psychology*, *10*, 835. DOI: https://doi.org/10.3389/fpsyg.2019.00835

Esaulova, Y., Penke, M., & Dolscheid, S. (2020). Referent cueing, position, and animacy as accessibility factors in visually situated sentence production. *Frontiers in Psychology*, *11*, 2111. DOI: https://doi.org/10.3389/fpsyg.2020.02111

Ferreira, V. S., & Dell, G. S. (2000). Effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology*, *40*(4), 296–340. DOI: https://doi.org/10.1006/cogp.1999.0730

Gleitman, L. R., January, D., Nappa, R., & Trueswell, J. C. (2007). On the give and take between event apprehension and utterance formulation. *Journal of Memory and Language*, *57*(4), 544–569. DOI: https://doi.org/10.1016/j.jml.2007.01.007

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science*, *11*(4), 274–279. DOI: https://doi.org/10.1111/1467-9280.00255

Hafri, A., Papafragou, A., & Trueswell, J. C. (2013). Getting the gist of events: Recognition of two-participant actions from brief displays. *Journal of Experimental Psychology. General*, *142*(3), 880–905. DOI: https://doi.org/10.1037/a0030045

Kretzschmar, F., Graf, T., Phillip, M., & Primus, B. (2019). An experimental investigation of agent prototypicality and agent prominence in German. In A. Gattnar, R. Hörnig, M. Störzer & S. Featherston (Eds.), *Proceedings of Linguistic Evidence 2018: Experimental data drives linguistic theory* (pp. 101–123). University of Tübingen.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. DOI: https://doi.org/10.18637/jss.v082.i13

Lenth, R. V. (2022, Version 1.7.2). emmeans: estimated marginal means, aka least-squares means. https://CRAN.R-project.org/package=emmeans

Maass, A., & Russo, A. (2003). Directional bias in the mental representation of spatial events: Nature or culture? *Psychological Science*, *14*(4), 296–301. DOI: https://doi.org/10.5194/gi-2016-11-RC2

Maass, A., Suitner, C., & Nadhmi, F. (2014). What drives the spatial agency bias? An Italian-Malagasy-Arabic comparison study. *Journal of Experimental Psychology. General*, *143*(3), 991–996. DOI: https://doi.org/10.1037/a0034989

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, *94*, 305–315. DOI: https://doi.org/10.1016/j.jml.2017.01.001

Messenger, K., Branigan, H. P., McLean, J. F., & Sorace, A. (2012). Is young children's passive syntax semantically constrained? Evidence from syntactic priming. *Journal of Memory and Language*, *66*(4), 568–587. DOI: https://doi.org/10.1016/j.jml.2012.03.008

Myachykov, A., Garrod, S., & Scheepers, C. (2018). Attention and memory play different roles in syntactic choice during sentence production. *Discourse Processes*, *55*(2), 218–229. DOI: https://doi.org/10.1080/0163853X.2017.1330044

Norcliffe, E., Konopka, A. E., Brown, P., & Levinson, S. C. (2015). Word order affects the time course of sentence formulation in Tzeltal. *Language, Cognition and Neuroscience*, *30*(9), 1187–1208. DOI: https://doi.org/10.1080/23273798.2015.1006238

Pokhoday, M., Shtyrov, Y., & Myachykov, A. (2019). Effects of visual priming and event orientation on word order choice in Russian sentence production. *Frontiers in Psychology*, *10*, 1661. DOI: https://doi.org/10.3389/fpsyg.2019.01661

R Core Team (2021, Version 4.1.2). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Sauppe, S. (2017a). Symmetrical and asymmetrical voice systems and processing load: Pupillometric evidence from sentence production in Tagalog and German. *Language*, *93*(2), 288–313. DOI: https://doi.org/10.1353/lan.2017.0015

Sauppe, S. (2017b). Word order and voice influence the timing of verb planning in German sentence production. *Frontiers in Psychology*, *8*, 1648. DOI: https://doi.org/10.3389/fpsyg.2017.01648

Sauppe, S., Norcliffe, E., Konopka, A. E., van Valin, R. D., & Levinson, S. C. (2013). Dependencies first: Eye tracking evidence from sentence production in Tagalog. *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, 1265–1270. https://mindmodeling.org/cogsci2013/papers/0242/

Schiller, N. O., & Caramazza, A. (2003). Grammatical feature selection in noun phrase production: Evidence from German and Dutch. *Journal of Memory and Language*, *48*(1), 169–194. DOI: https://doi.org/10.1016/S0749-596X(02)00508-9

Schlenter, J., Esaulova, Y., Dolscheid, S., & Penke, M. (2022). Ambiguity in case marking does not affect the description of transitive events in German: Evidence from sentence production and eye-tracking. *Language, Cognition and Neuroscience*. Advance online publication. DOI: https://doi.org/10.1080/23273798.2022.2026419

Schriefers, H., Hantsch, A., & Jescheniak, J. D. (2002). Determiner selection in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(5), 941–950. DOI: https://doi.org/10.1037/0278-7393.28.5.941

Suitner, C., Maass, A., Navarrete, E., Formanowicz, M., Bratanova, B., Cervone, C., Hakoköngäs, J. E., Kuppens, T., Lipourli, E., Rakić, T., Scatolon, A., Teixeira, C. P., Wang, Z., Sobral, M. P., & Carrier, A. (2021). Spatial agency bias and word order flexibility: A comparison of 14 European languages. *Applied Psycholinguistics, 42*(3), 657–671. DOI: https://doi.org/10.1017/S0142716420000831

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour, 1*(3), 1–8. DOI: https://doi.org/10.1038/s41562-017-0058

Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX)*. DOI: https://doi.org/10.17605/OSF.IO/MD832