Cognitive and Affective Aspects of Verbal Humor: A Visual-World Eye-Tracking Study

Laura Israel, Lars Konieczny and Evelyn C. Ferstl *

Center for Cognitive Science, Institute of Psychology, University of Freiburg, Freiburg, Germany

Many theories of verbal humour postulate that the funniness of jokes is caused by an incongruency in the punchline whose resolution yields a feeling of mirth. While there are studies testing the prediction that this situation model updating leads to increases in processing costs, there are few studies directly assessing the time course of when the alternative situation models are entertained. In a visual world paradigm, stories were presented auditorily and displays were presented illustrating either the situation implied by the context or the final interpretation after the punchline. Eye movement data confirmed the switch from the initial to the final interpretation for jokes as well as for non-funny control stories that also required a situation model revision. In addition to these effects of the cognitive revision requirements, the pupil dilations were sensitive to the affective component of joke comprehension. These results are discussed in light of incongruency theories of verbal humour.

Keywords: verbal humor, joke comprehension, eye-tracking, visual world, pupil dilation, language, emotion, incongruity

1 INTRODUCTION

Humour is ubiquitous in our social interactions. It is considered relevant for the choice of partners and friends (Bressler and Balshine, 2006), it helps to strengthen in-group cohesion, and it can be an effective tool for relieving tension and reducing stress (Kimata, 2004; Buchowksi et al., 2007).

Psycholinguistic humour research is concerned with the interaction between two aspects of verbal humour: an affective reaction (i.e., the feeling of mirth or enjoyment, possibly shown in smiling or laughing), and a cognitive element, reflecting the correct linguistic interpretation of the intended meaning of the humourous utterances. Goel and Dolan (2001) referred to this distinction as the hot and cold aspects of joke comprehension. In the following we refer to these processes as joke appreciation and joke comprehension, respectively.

The goal of the present study is to use a visual world paradigm to study both aspects. The viewing patterns for pictures illustrating the gist of jokes and appropriate control texts are expected to provide information about the interpretations entertained during listening to the stories. In addition, pupil size measures will be shown to be sensitive to the affective aspects of joke comprehension. Before describing the study in more detail, we will summarize previous psycholinguistic research on verbal humour.

Psycholinguistic research often studies the comprehension of verbal jokes. A prototypical joke, also referred to as canned joke (Martin, 2007) or garden-path joke (Mayerhofer, 2015), is a short, funny story that offers different possibilities of interpretation. After a setting activating a certain

situational context, the punchline at the end of the story contains a surprising incongruity whose resolution elicits a feeling of mirth (Dynel, 2009). Incongruity as a crucial prerequisite for humour processing was first described by Kant (1790/1977) and later adapted in incongruity theories of humour (Suls, 1972; Raskin, 1979; Attardo and Raskin, 1991; Giora, 1991; Attardo, 1997; Coulson, 2006). The models assume a two-stage process that can be demonstrated with the following example by Raskin (1979, p. 332):

> "Is the doctor at home?" the patient asked in his bronchial whisper. "No" the doctor's young and pretty wife whispered in reply. "Come on right in".

In the context sentence of this joke the lexical items doctor, patient and bronchial indicate the situation to be a doctor's visit and hence activate a matching narrative scheme or script (also: situation model, van Dijk and Kintsch, 1983; Ferstl and Kintsch, 1999; or frame, Coulson, 2006). This representation of experience—based world knowledge is used to make assumptions about the text continuation. However, the punchline is not consistent with a doctor's visit; the wife's invitation, a surprising violation of the expectations, induces a switch to the correct narrative script—a love affair. Incongruity theories postulate that the funniness of the joke is elicited by the overlap of the two conflicting scripts as well as by the strength of their semantic opposition.

Similar to more general theories of figurative language processing (e.g., Standard Pragmatic View, Grice, 1975; Searle, 1979; Graded Salience Hypothesis, Giora, 2002) incongruity models assume serial processing for the detection and resolution of incongruities. Therefore, they predict that the resolution of an incongruity is more time—consuming and leads to higher cognitive costs than the processing of texts which do not induce situation model updating or frame shifting. In line with this prediction, several studies show a processing disadvantage for jokes over non-funny texts. In a self-paced reading study Coulson and Kutas (1998) presented jokes and matching control texts that differed only in regard to their last word, e.g.,

> I asked the women at the party if she remembered me from last year and she said she never forgets a dress/ face/name.

The last word *dress* in this joke triggered an incongruity and a revision, compared to a more expected word, such as *face*. Control texts ended in a similarly unexpected but contextually compatible word, such as *name*. Reading times for the sentence endings were longer in jokes than in the control texts. These results were replicated using eye tracking (Coulson, 2006). In neuroscientific studies, an enhanced N400 component, an indicator for contextual integration difficulties, has been reported for funny texts compared to control texts (Coulson and Kutas, 2001; Coulson and Lovett, 2004; Coulson and Williams, 2005; Mayerhofer and Schacht, 2015). In addition, late integration processes, as indicated by components such as

the P600 or the LPC, follow the incongruency detection, but are completed within about 1,500 ms (Canal et al., 2019). Interestingly, there is also evidence that humour facilitates processing. Memory for funny contents or jokes is better than for non-funny texts—a finding that is labelled humour effect (Schmidt, 2002; Strick et al., 2010). In a study comparing jokes to non-funny proverbs, Mitchell et al., 2010 confirmed this memory advantage and reported shorter reading times for the joke punchlines compared to the non-funny proverb endings.

The divergent results of these studies might in part be due to confounds of the funniness of the materials with their revision requirements. In the Coulson and Kutas (1998) study, only the joke ending *dress* requires a frame shift. Both the expected ending *face* and the control ending *name* are aspects of the identity of the person. In the Mitchell et al. (2010) study, on the other hand, inspection of the materials confirms that the proverbs also require processing of non-literal language, and are more difficult than straightforward control texts.

Based on these observations, Siebörger (2006, see also Hunger et al., 2009; Volkmann et al., 2008) argued that the relative contribution of incongruity resolution to joke appreciation requires disentangling the need for revision from the funniness of the stories. Thus, two appropriate control conditions are warranted: First, it is necessary to use identical punchlines to minimize lexical and sentence level differences. Second, and more importantly, the linguistic demands of incongruity resolution need to be compared in funny and non-funny stories. The hypothesis was that incongruity resolution alone is necessary, but not sufficient for making a joke funny, and that comprehension difficulty increases with the cognitive requirements, but does not depend on the emotional content.

To test this hypothesis, jokes and corresponding revision stories were written (Siebörger, 2006). Example materials are shown in **Table 1**. Each text consists of one or two context sentences and a target sentence (punchline). The stories are short dialogues of two protagonists. The punchlines of the jokes and revision texts both contain a sudden twist that causes an incongruity and calls for a revision of the story. The incongruity may depend on a homonym or on a situational shift. That jokes and revision stories differed with respect to their funniness was confirmed in several rating studies (see Siebörger, 2006). Control texts were derived from each joke and each revision story by writing a context sentence that already activated the situation model required for correctly interpreting the punchline. Thus, a straightforward, coherent condition was created with the identical punchline sentences as the experimental stories (jokes or revision texts). This design allowed us to evaluate the effects of linguistic revision independent of word- and sentence-level features of the punchlines. The results of a self-paced reading study included better question answering performance for jokes compared to revision stories and shorter reading times for the punchline of jokes.

In a recent a study using eye tracking during reading we replicated and extended these findings (Ferstl et al., 2017). The more fine-grained temporal analysis confirmed the general

TABLE 1 | Example materials for the joke and revision texts and their matching control texts used in the present study as well as incoherent control texts used by Siebörger (2006) and Ferstl et al. (2017). The punchlines are identical for the experimental stories and their matched control texts. Asterisks denote the revision points.

	Joke	Joke control text
Context	Herbie to his pal Mark: "Why have all these women stopped chasing after you lately?"	Herbie to his pal Mark: "Great that you've managed to stay within the law lately!"
Punchline	Mark: "Well, life as a pickpocket * turned out to be too risky in the long run."	
	Revision Text	**Revision Control Text**
Context	The chef to his wife: "Oh, too bad—the soup is too salty!"	Mr. Miller to his wife: "You oversalted our dinner again!"
Punchline	The wife: "I'm sorry *, I'll try to do better next time!"	

facilitation for jokes, but showed the effects to be due to reading strategies. While first-pass reading times were similar across conditions, total reading times were longer and more regressions were made for non-funny texts compared to jokes. These findings were more pronounced in men than in women, and when the instruction required a meta-cognitive evaluation of the revision requirement, compared to an affective funniness rating.

To explain these effects, we proposed that the appreciation of funniness works as an instantaneous, self-generated feedback about the correctness of the comprehension process (Ferstl et al., 2017). When a the reader laughs at the punchline of a joke it becomes directly apparent that they comprehended the text correctly. The lack of such an intrinsic feedback in non-funny texts might induce rereading strategies, in particular in an experimental setting that promotes a focus on accurate task performance.

While the study was intended to disentangle affective and cognitive components of joke comprehension, both task instructions and the reading format might have induced more deliberate strategies than expected. In addition, although eye tracking during reading is a very sensitive method that provides information about the time course of processing, it gives only indirect evidence about the content of the interpretations entertained at any given moment. Usually, increases in reading times or the occurrence of regressions are taken as evidence for increased processing costs, which in turn are interpreted as reflecting situation model updating.

One important paradigm to assess the contents of the situation model directly is sentence-picture matching (e.g., Zwaan et al., 2002). After reading or listening to a short text, one or several pictures are presented. The comprehender's task is to evaluate whether the picture matches the meaning of the presented text or to select the best illustration among several pictures. Errors and reaction times show whether one interpretation is more easily or more quickly accessible than another. Recently, similar tasks have been adopted for visual world paradigms (Huettig et al., 2011; Berends et al., 2015; Salverda and Tanenhaus, 2018). In this method participant's eye movements are recorded while they look at a visual scene and simultaneously hear a spoken utterance. The spoken expression is related to one or more objects in the visual display: To study ambiguity resolution, the visual scene contains pictures that are compatible with either alternative interpretation (e.g., Tanenhaus et al., 1995). The question of interest is when and for how long the listeners view the respective images. While this paradigm has been successfully applied to study lexical access or

syntactic ambiguity resolution (see Salverda and Tanenhaus, 2018, for review), few studies have targeted aspects of higher level comprehension (e.g., Pyykkönen et al., 2010; Pexman et al., 2011) and none, to our knowledge, the comprehension of verbal humour.

The interrelation between visual attention and language processing has been proven in several studies. Cooper (1974) was the first to show that eye movements focus on objects of a visual scene that directly or semantically refer to a simultaneously heard verbal utterance. When hearing the word *boat* subjects were more likely to fixate the picture of a boat (resp., lake) than that of other objects. The number of fixations increases in proportion to the semantic similarity between the visual object and the heard utterance (Huettig and Altmann, 2005). Furthermore, eye movements can even be modulated by words that are only anticipated by the listener and not explicitly stated (Altmann and Kamide, 2007). Because of its very good temporal resolution, depicting language-related effects on oculomotor control within 80–100 ms (Altmann, 2011), and the facilitation of measuring unconscious processes, the visual world paradigm is well suited to gain insight into the temporal characteristics of language processing.

The goal of the present study was to take advantage of this paradigm to describe the viewing patterns for pictures illustrating the alternative situation models elicited by the context and the punchline, respectively. Jokes and revision texts as well as their matching coherent control stories from the material of Siebörger (2006; Ferstl et al., 2017) were used. For each joke and each revision story three scenes were constructed: A target depicting the intended, correct meaning of the story, a competitor that showed the situation implied by the context, and an unrelated distractor. Examples are presented in **Figure 1**.

The images are plausible illustrations of the situations implied by the texts, although other possible scenes might also be associated with the text materials (e.g., a car thief, rather than a pickpocket, as an illustration of a criminal in the example in **Figure 1**).

After the auditory stimulus presentation, participants performed a picture matching task to assess comprehension accuracy and they rated the funniness of the stories. The funniness ratings are expected to be higher for jokes than for all other conditions—which serves as a manipulation check for the stimulus material.

The eye tracking data provided two sources of information. First, the durations of the viewing times (dwell times) of the three pictures were analysed during listening. The viewing times provide evidence for whether and when the depicted situation models are entertained. Even though the
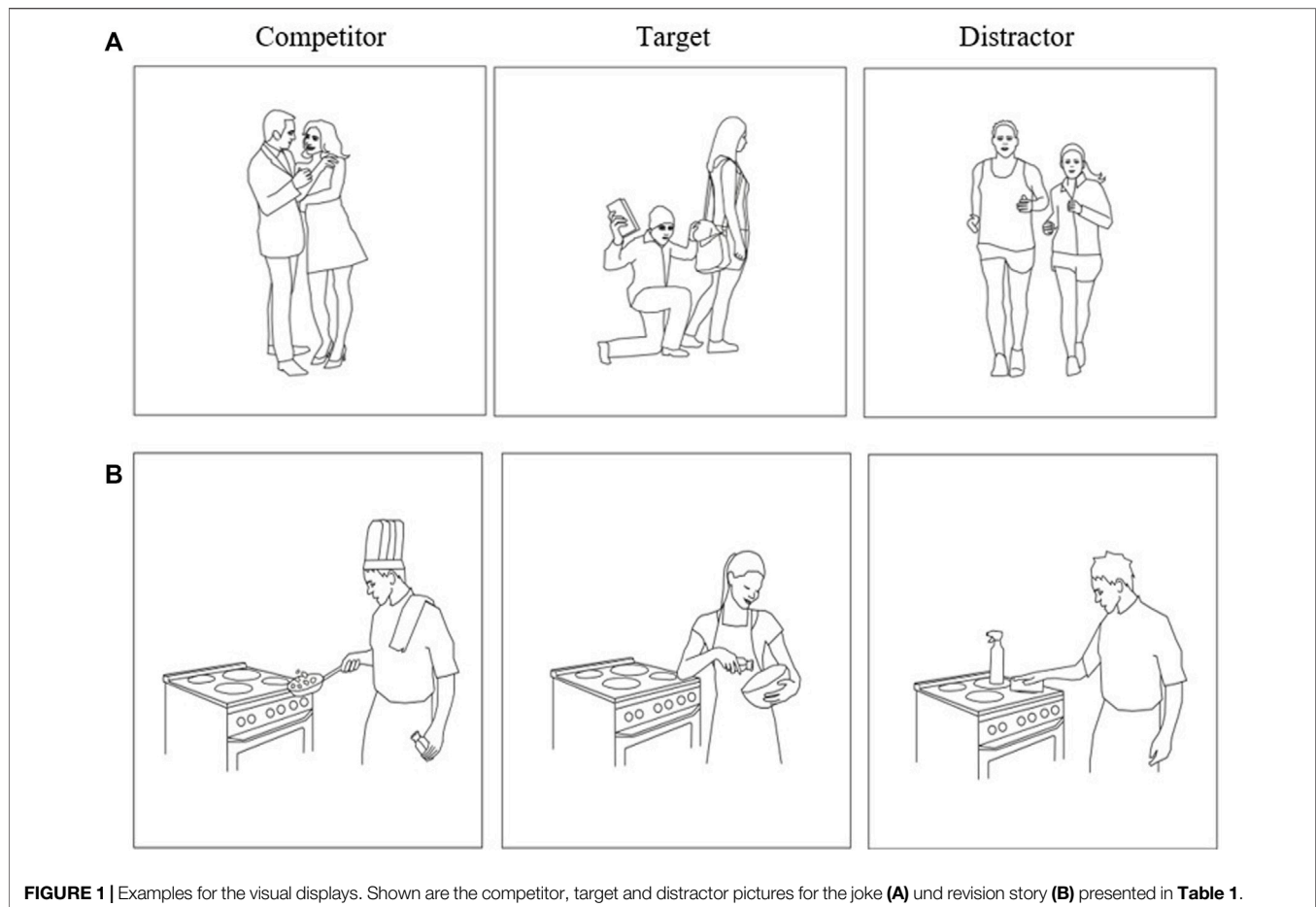
FIGURE 1 | Examples for the visual displays. Shown are the competitor, target and distractor pictures for the joke **(A)** und revision story **(B)** presented in **Table 1**.

present study analyses eye-tracking patterns during joke comprehension in a rather exploratory way, some assumptions were derived from theory and the experimental design: The viewing times were predicted to reflect the current interpretation induced by the stories. Since in the control stories the context sentences already introduced the correct situation model, participants were expected to look more at the target than at the competitor throughout the entire trial. In contrast, in both joke and revision stories, the competitor pictures were expected to be viewed more during the context sentences, followed by a switch to the target only after the incongruency becomes apparent in the punchline. If the previously reported facilitation effect for jokes over the structurally identical revision stories (Ferstl et al., 2017) had been due to meta-cognitive reading strategies only, no differences between jokes and revision stories are predicted. If, on the other hand, the facilitation was due to immediate joke appreciation, the switch to the intended target picture should be faster or more pronounced for jokes than for revision stories.

The second source of information was the pupil size as a measure of physiological arousal (Hess and Polt, 1960; see Sirois and Bisson, 2014, for review). The pupil diameter varies with cognitive load and processing costs (Hyönä et al., 2007; Engelhardt et al., 2010; Wong et al., 2016). More importantly, it also increases with exposure to

emotional stimuli (e.g., facial expressions, or sounds of crying or laughing children; Partala and Surakka, 2003), and it has been a useful measure for emotional reactions in studies on social-affective processes (e.g., Prehn et al., 2013). In a study on joke comprehension, Mayerhofer and Schacht (2015) found increases in pupil sizes that varied with the perceived funniness of their text materials. The influence of emotions on pupil size has been explained by the association of pupil dilation with cardiac acceleration (Bradley et al., 2001) and the sympathetic nervous system (Bradley et al., 2008).

As the pupil dilation varies with cognitive load, we assumed that larger pupil sizes would occur in the ambiguous joke and revision conditions compared to the coherent control stories. The crucial question was whether the emotional arousal elicited by jokes, compared to non-funny stories, would also be reflected in pupil dilation. In that case, joke appreciation was expected to lead to an additional increase in pupil size for jokes compared to non-funny revision stories.

2 MATERIALS AND METHODS

Reproducible scripts, open data and open materials (including all auditory as well as all visual stimuli) are provided via our OSF repository at https://osf.io/qk95j/.
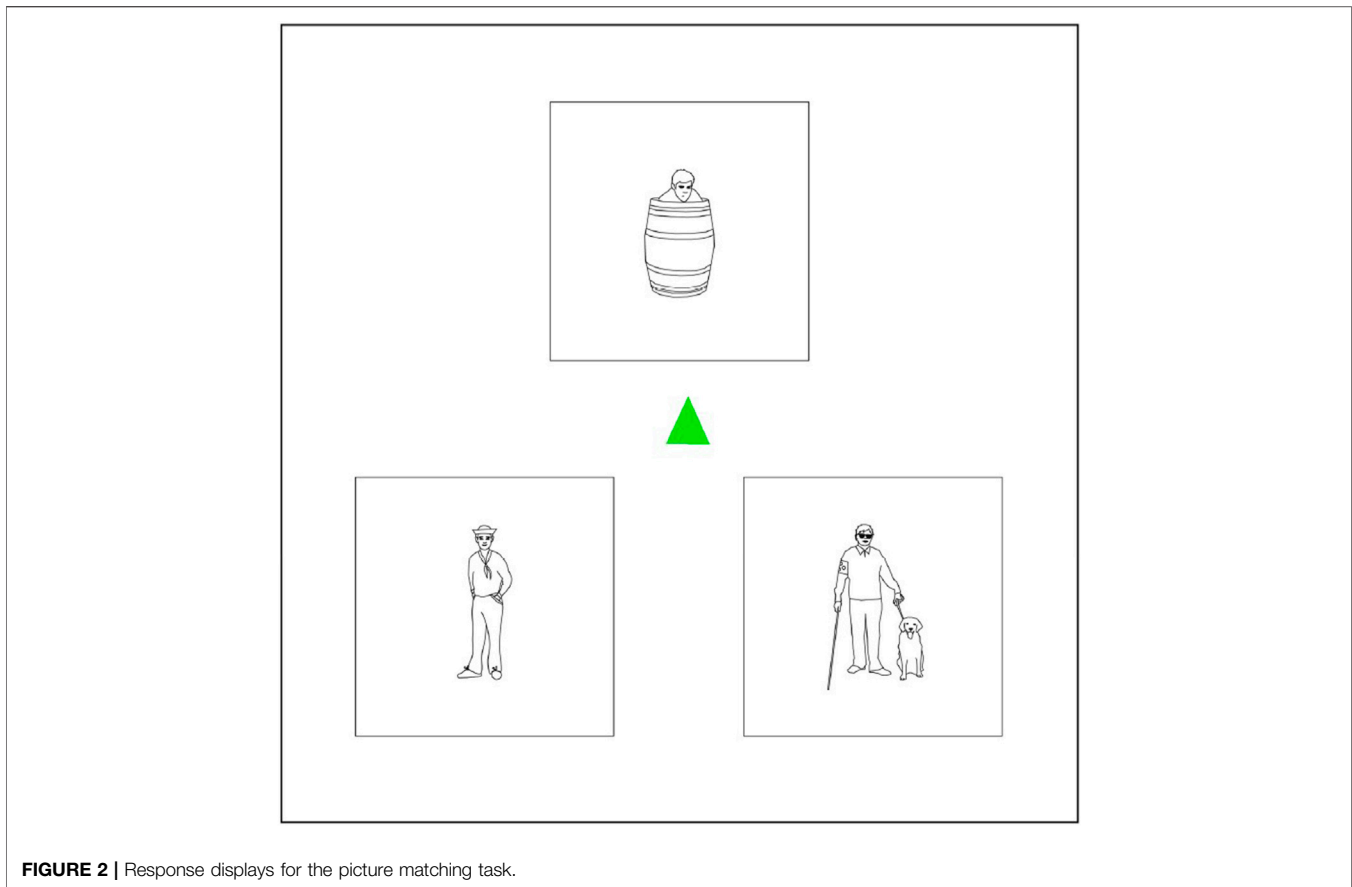
FIGURE 2 | Response displays for the picture matching task.

2.1 Participants

Thirty-one individuals participated in the study. All of the participants were native German speakers and reported to have normal or corrected-to-normal vision. Most of them were students at the University of Freiburg. The data of five participants were excluded. Two subjects had to be eliminated from the sample due to technical problems. Three participants were excluded because of their poor performance in one of the behavioural tasks of the experiment. The final sample used in the analysis of the eye-tracking data consisted of 26 participants (13 women, 13 men). Their mean age was 22.8 years (range: 19-29). Due to the repeated-measures design of the study, where each participant viewed 32 stories, 832 observations were obtained for the evaluation of the behavioural data (funniness ratings and picture matching accuracy).

2.2 Materials

For the experiment 16 jokes and 16 revision stories were selected from the material of Siebörger (2006). The stories were short German dialogues between two characters. All 64 stories had an identical structure, consisting of a context and a target sentence. In the jokes the last sentence (in the following referred to as "punchline") contained information that required a revision of the text and was funny. The revision stories also required a revision, but were not funny. Sieörger (2006) conducted three extensive pretests to select jokes and revision stories for further

experiments. Different groups of approximately 32 student participants rated the texts according to their familiarity, distinguished revision and joke stories in an explicit choice, and evaluated whether the text required a situation model revision or not, and how funny they found the texts. 64 jokes and revision texts were selected so that they differed in funniness, could be assigned to one or the other text category, and were unfamiliar to more than 75% of the participants. For the present experiment, a subset of 16 jokes and 16 revision stories was selected for which straightforward illustrations were possible. The two sets were matched for length.

For each experimental joke and revision text a straightforward, coherent control story had been constructed by changing the context sentence and keeping the punchline identical.

For the implementation of the visual world experiment all texts were recorded with the text-to-speech function of the Mac operating system (Version 10.10.3) and adjusted to a natural speaking rate. The automatic voice was chosen to prevent prosodic cues about the story type. All sentences and texts were recorded using the same voice. Audio sequences had an average length of 10,091 ms (range, 7,861–1,174 ms; SD = 999.3 ms). There were no significant differences in length between the text categories. The context sentences and punchlines were recorded separately and connected within the experimental software.

For an analysis of the time course of processing, the individual revision point of each story was defined as the moment at which the incongruity can be perceived and subsequently, the situation model needs to be revised. To determine the revision point for each experimental text, two independent raters were instructed about the structure of the joke and revision stories and were explicitly asked to mark the point where they discovered the incongruity. The revision point usually occurred towards the end of the stories, at about 760 ms before the ending (range 0–4,981 ms). There was no difference between jokes and revision stories with respect to the temporal position of the revision point. The revision point of each experimental story was carried over to its matching control text. The revision points are marked in the example texts in **Table 1**.

For each joke and revision story three black-and-white line drawings were constructed with the Adobe Illustrator (version CS4; Adobe Inc., 2019). The graphics depicted the salient meaning implied by the context (competitor picture), the correct interpretation of the story obtained after the revision process (target picture), as well as an unrelated scene (distractor picture). Examples of the picture sets for joke and revision texts are presented in **Figure 1**. To avoid obvious visual differences between the three pictures, the number of individuals and objects as well as their spatial configuration in the scene were kept similar throughout each set of pictures. The straightforward control stories were presented with the same picture sets as their experimental counterparts. Each scene was presented in a 13 × 13 cm thin black frame. On the visual display the three frames were arranged in a triangular order (**Figure 2**). For the analysis of the eye movements the display was divided into three even interest areas (IAs)—one for each of the presented frames.

To counterbalance the materials, six lists of 32 stimuli were constructed. Each list contained eight experimental jokes, eight experimental revision texts, eight joke control texts, and eight revision control texts. As a result, each punchline sentence was included exactly once in each list. The positions of the pictures (target, competitor, distractor) on the visual display were balanced over the six experimental lists using a Latin square design. The order of the trials within each experimental list was pseudo-randomised with the constraint of not more than two successive trials of the same condition.

2.3 Apparatus and Calibration
The implementation of the experiment and the response recording were conducted with the software package Experiment Builder [version 1.10.165; Experiment Builder (Computer software), 2014]. Eye movements of the dominant eye, assessed in a short test after arrival of the subject, were recorded using the Eye-Link 1,000 eye-tracking system (www.sr-research.com). During the experiment, the participant's head was stabilized with a headrest, about 60 cm away from the computer screen. The eye-tracker was calibrated at the beginning of each block of 16 trials using a nine-point grid. Drift-correction was performed prior to each trial: a black dot was presented in the centre of the screen and stimulus presentation started only after fixation of the dot.

2.4 Procedure
After their arrival participants were randomly assigned to one of the six experimental lists. Written instructions were then presented on the computer screen to explain the task and the calibration procedure. A short training session with four trials followed. For this session pictures from the material of Volkmann, Siebörger and Ferstl (2008) were used. These drawings were more detailed and less balanced than the pictures used in the current study, but were similar in form and content to the experimental materials.

Each trial consisted of an eye-tracking phase followed by two behavioural tasks. In the eye-tracking phase participants first viewed the set of three pictures for 6 seconds. Making the visual displays available before the onset of the auditory presentations ensures that the eye movements reflect the integration of the subsequent linguistic input with the pictures, rather than processes related to visual object recognition and scene interpretation (see Hüttig et al., 2011). The story was then presented over headphones while participants continuously viewed the picture sets. Subsequently, subjects looked at the visual stimuli for another 4 seconds without auditory input. The end of the eye-tracking phase was indicated by a short acoustic signal. At the same time, the blue triangular cursor symbol in the centre of the screen changed its colour to green. Subjects could now move the cursor symbol with their mouse and click on the picture they considered the best match for the meaning of the story. After the picture selection participants were asked to rate the funniness of the story on a nine-point scale ranging from not funny to very funny (cf. Ferstl et al., 2017). The instructions stressed that the participants should use their own intuitive, personal criteria for the ratings, and that they should try to use the full range of values on the scale.

Participants were allowed to take a short break after half of the experiment (16 trials) and they completed the session in about 35–40 min.

2.5 Design and Data Analysis
Statistical analysis was performed using the software R (version 4.1.1; R Core Team, 2017). Mixed-effects regression models were calculated using the lmer und glmer function from the lme4 package (version 1.1-27.1, Bates et al., 2014). A problem in the statistical evaluation of visual world data is that not all observations are independent, because of the multilevel sampling scheme of the paradigm as well as the properties of eye-movements in general (Barr, 2008). To account for these non-independencies, as proposed by Barr (2008), mixed-effects models with random-effects corresponding to the different clusters in the sampling design were used. To increase the interpretability of the very complex models, main effects and interactions were determined using the Anova function from the car package (Fox and Weisberg, 2011). The significance level α was set to $p = 0.05$ (95%). Reported model estimates were calculated using the emmeans function from the lsmeans package (version 2.30-0, Lenth, 2014). For the analysis of the

eye-movement data, only correctly answered trials were used to prevent bias due to the misunderstanding of the stories.

To analyse the temporal evolution of the eye movements two interest periods (IPs) were defined. Each time bin had a length of 1,500 ms. The IPs were time-locked at the story's individual revision point. They covered 3,500 ms, starting 1,500 ms before and ending 2000 ms after the revision point (see Materials). The first IP (before) covered the immediate 1,500 milliseconds before and up to the revision point, while the second IP (after) started 500 ms after the revision point and extended to 2000 ms after the revision point. The first IP was used to establish the baseline, while the second IP was used to detect any effects of the design manipulations. The window duration of 1,500 ms was chosen based on results from Canal et al. (2019) who reported a similar time frame for the completion of joke processing. We excluded the first 500 ms after the revision point to allow for comprehension and integration of the crucial information into the text representation, and to make sure the interest period were clearly separated. The dependent variables included the subjects' performance in the picture matching task (accuracy in percent), and the funniness ratings (on a scale from one to 9). For the analysis of the eye movements, the dwell times (in ms) for each of the three interest areas (IAs) were collected for each IP. In addition, the mean pupil size (measured in arbitrary units; Eyelink Data viewer User's Manual, 2011) was assessed for each IA and each IP. As we expected large interindividual differences in pupil size (e.g. Winn et al., 1994), we z-standardized the pupil size on subject level to remove this random variance from the variable.

For the picture matching task and the funniness rating, the independent variables text category (joke story/revision story) and item type (experimental item/control item) were varied within participants. For the statistical evaluation of the eye tracking data (dwell times and pupil size), IA (target/competitor/distractor) and IP (before vs. after) were added as further within-subject variables. Main and interaction effects of all these variables were added to the models.

In addition, trial (serial position) was added as a covariate to all models.

For the modelling of dwell times and pupil sizes, two funniness covariates were added. First, the group-mean centred (on subject level) funniness ratings were added to the model, coding how funny each story was perceived by each participant in relation to their own average funniness rating. In the following, we will refer to this variable simply as funniness. In addition, we added the mean funniness rating of each participant over all stories as covariate to the models. The variable hence encodes how reactive the respective participant was to the stories in general and how strongly the participant perceived all stories to be humorous. We therefore will refer to this variable as sense of humour. Again, we assume that sense of humour could influence eye-related measures like the dwell times. Besides the two fixed covariates for funniness and sense of humour, a random slope for the funniness variable was introduced for each subject, controlling for possible individual difference in participants' reactions to the stories.

To evaluate whether the performance in the picture-matching task had an effect on eye-tracking variables, the subject's mean performance (percent correct) was introduced as an additional covariate. For the analysis of the pupil sizes, the same fixed effects predictors were used.

The random effects for both models included terms for subject and item. To account for the variability between the stories, a variable that encodes the item and the respective experimental list was added as random factor. Item was coded to match every experimental text (revision or joke) to its associated control text. The experimental list, furthermore, determines the configuration of interest areas on the screen.

The random slopes differed for the analyses of dwell times and pupil sizes. For the dwell times, only IA was used. Because we included fixations on all three relevant interest areas in our models, summed dwell times in each interest period would very often max out to the length of that IP. In addition to adding IA as a fixed effect predictor to our models, we therefore also added the random slope of IA to the dwell time model, to account for individual differences (random effect of subject) or item-baseddifferences (random effect item_and_list) in the proportions of looking to the target or any other IAs. Also, no random intercepts were necessary, as all dwell time intercepts would be close to the length of the IP.

For the pupil size model, random slopes for trial, text category, item type and funniness were added for the subject random factor. No random intercept for subject was implemented as we already z-standardized the pupil sizes beforehand.

In the provided R-Scripts on our OSF repository, the complete model specifications of the implemented models can be found.

The combined fixation reports from all participants contained 52,544 fixations. These were used to generate combined area of interest (IA) and period of interest (IP) reports, with each trial consisting of the summed dwell times for each of the three IAs and each of the two IPs, yielding 3×2 observations per trial and a total of $26 \times 32 \times 3 \times 2 = 4,992$ observations. Eighty trials (9.6%) with incorrect picture mapping responses were excluded, leaving a total of 4,512 (4,992–80 × 6) observations. This data set was used for the analysis of dwell time. For the pupil dilation data, 1803 observations were excluded if the dwell time for an area of interest was zero, leaving a total of 2,709 observations for pupil data analysis.

3 RESULTS

3.1 Behavioral Data
Estimated means and standard errors for the behavioural measures are shown in **Table 2**.

3.1.1 Funniness Ratings
The average funniness rating across all trials was 2.95 (SD = 2.36). The variance explained by the fixed effects in the model was 33.4%, including random effects raised this to 63.0%. The mixed-effects regression model showed a significant main effect for text category [$\chi2$ (1) = 92.80, p < 0.0001] and item type [$\chi2$ (1) = 177.44, p < 0.0001] as well as a significant interaction of these two

TABLE 2 | Estimated means and standard errors for the funniness ratings and the picture-matching accuracy for the four text categories.

	Joke		Revision	
	Experimental	Control	Experimental	Control
	M (SE)	M (SE)	M (SE)	M (SE)
Funniness rating (Scale 1-9)	5.22 (0.25)	2.19 (0.25)	2.76 (0.29)	1.64 (0.29)
Accuracy (%)	0.82 (0.06)	0.99 (0.01)	0.99 (0.001)	0.99 (0.01)

TABLE 3 | Statistical results of the analyses of the dwell times and the pupil sizes. Details of the models are described in the Results section.

Effect	df	Dwell times (p)	Pupil size (p)
text category	1	0.967	0.483
item type	1	0.503	0.060^
IP	1	0.015*	<0.001***
IA	2	0.001***	0.200
funniness	1	0.464	0.005**
trial	1	0.988	<0.001***
performance	1	0.202	0.451
Sense of humor	1	0.317	0.101
text category: item type	1	0.957	0.133
text category: IP	1	0.632	<0.001***
item type: IP	1	0.770	<0.001***
text category: IA	2	0.021*	0.003**
item type: IA	2	0.001***	0.834
IP: IA	2	<0.001***	0.519
text category: item type: IP	1	0.766	0.994
text category: item type: IA	2	0.665	0.273
text category: IP: IA	2	0.062^	0.624
item type: IP: IA	2	<0.001***	0.525
text category: item type: IP: IA	2	0.817	0.063^

Text category = Joke/Revision; Item type = Experimental/Control; IA, interest area, Target/Competitor/Distractor; IP, interest period, before/after. Significance values: *** 0.001 ** 0.01 * 0.05 ^ 0.10.

TABLE 4 | Statistical results of the analyses of the effects of funniness on the pupil sizes for the experimental texts. Details of the model are described in the Results section.

Effect	Chisq	Df	p
text category	0.4851	1	0.606
IP	76.0140	1	<0.001***
IA	1.3996	2	0.601
funniness	5.8860	1	0.055^
trial	36.9286	1	<0.001***
performance	0.0000	1	0.814
Sense of humour	4.8333	1	0.043*
text category: IP	0.0616	1	0.850
text category: IA	0.8284	2	0.663
IP: IA	2.7547	2	0.289
text category: funniness	0.0105	1	0.738
IP: funniness	10.9686	1	<0.001***
IA: funniness	3.7296	2	0.167
text category: IP: IA	2.8273	2	0.172
text category: IP: funniness	0.2846	1	0.572
text category: IA: funniness	0.3368	2	0.878
IP: IA: funniness	0.2320	2	0.778
text category: IP: IA: funniness	0.9381	2	0.671

Text category = Joke/Revision; IA, interest area, Target/Competitor/Distractor; IP, interest period, before/after. Significance values: *** 0.001 ** 0.01 * 0.05 ^ 0.10.

factors $[\chi2 \ (1) = 37.77, \ p < 0.0001]$. The model estimates in **Table 2** show that jokes yielded substantially higher funniness ratings than the other three text categories.

3.1.2 Picture Matching Task

In the picture matching task participants picked the correct target picture in 91.4% of the trials. In 7.3% of the trials the competitor picture was chosen, while the distractor was picked in only 2.3% of the trials. The logistic regression model explained 22.4% of the variance with fixed effects, and 61.5% with both random and fixed effects. The statistical analysis of the accuracy rates also showed a significant main effect for text category $[\chi2 \ (1) = 11.85, \ p = 0.0006]$ and item type $[\chi2 \ (1) = 14.86, \ p = 0.0001]$ as well as a significant interaction between the two factors $[\chi2 \ (1) = 8.50, \ p = 0.0035]$. The model estimates in **Table 2** indicate that while revision stories, revision control stories and joke control stories yielded a very high performance (99% estimated accuracy), the joke condition led to a higher error rate with an estimated accuracy of 82%.
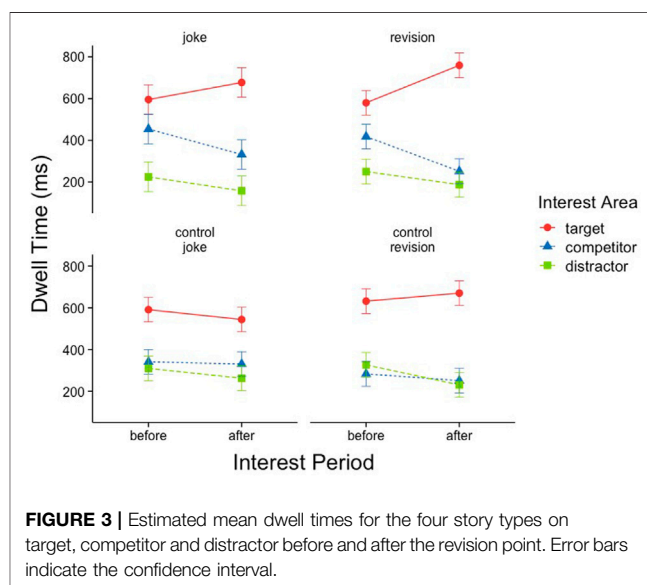
3.2 Eye Movement Data

Results of the mixed-effects regression models for the dwell times and the average pupil size are presented in **Table 3**.
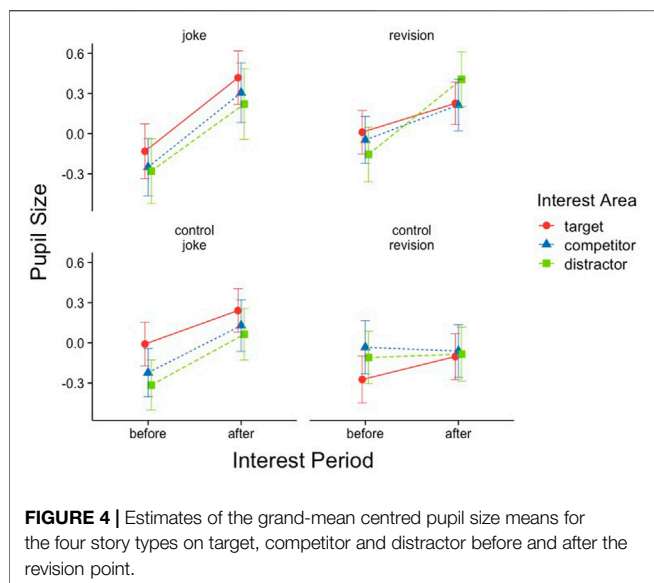

FIGURE 3 | Estimated mean dwell times for the four story types on target, competitor and distractor before and after the revision point. Error bars indicate the confidence interval.

3.2.1 Dwell Times

In the mixed-effects regression model for the analysis of the dwell times, the fixed effects explained 14.5% of the variance in dwell

FIGURE 4 | Estimates of the grand-mean centred pupil size means for the four story types on target, competitor and distractor before and after the revision point.

FIGURE 5 | Estimates of the grand—mean centred pupil size means as a function of interest period and the trial-based individual funniness ratings.

times while the model including random effects explained 24.2%. The Anova analysis (**Table 3**) showed significant main effects of IA and IP, as well as a significant two-way interaction of IA and IP. Inspection of the model estimates displayed in **Figure 3** shows that the target dwell times were higher than those of the distractor and competitor. This effect was more pronounced after the revision point than before. Crucially, the 2-way interaction of IA and item type, as well as a significant three-way interaction of IP, IA and item type indicate that the viewing patterns change mostly for the experimental texts (jokes and revision stories), but less so for the control texts.

Pairwise comparisons confirm this observation: Target dwell times increased for the experimental stories ($z = -4.3$, $p < 0.0001$), but not for the control stories ($z = 0.15$, n. s.); and competitor dwell times decreased for the experimental stories ($z = 4.7$, $p < 0.0001$), but not for the control stories ($z = 0.8$, n. s.). The distractor dwell times decreased for both item types (experimental stories: $z = 2.1$, $p < 0.05$; control stories: $z = 2.5$, $p < 0.05$).

The model also yielded a significant two-way interaction of text category and IA. Across both interest periods and averaged across experimental and control stories, there was a tendency for target pictures to be looked at longer for revision compared to joke stories ($z = -1.9$, $p = 0.06$), while the competitor pictures were viewed longer for the joke stories compared to the revision stories ($z = 2.4$, $p < 0.05$). There were no differences for the distractor pictures ($z = -0.38$, n. s.).

Finally, the results for the dwell times were independent of the trial position, the performance in the picture matching task, and the funniness ratings. None of the respective effects reached significance.

3.2.2 Pupil Size
The fixed effects of the multi-level regression model were able to explain 19.7% of the pupil size variance, fixed and random effects explained 37.9%. The Anova analysis over the model coefficients yielded a marginally significant main effect of IA,
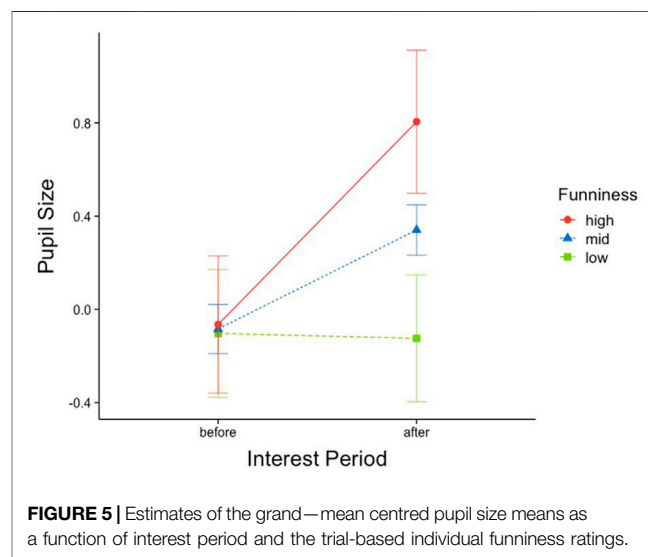
and a two-way interaction of IA with text category. The model estimates in **Figure 4** indicate that for the joke stories (both experimental and control), the pupil sizes were larger on the target pictures than on the competitor ($t = 2.4$, $p = 0.046$) and the distractor ($t = 3.2$, $p = 0.005$), with no difference between competitor and distractor ($t < 1$, n. s.). There was no effect of IA on the pupil sizes for the experimental and control revision stories (t's < 1, n. s.).

Importantly, the pupil sizes changed from before to after the revision point. The significant main effect for IP and an interaction of IP and text category was found. Inspection of the model estimates presented in **Figure 4** shows that experimental jokes and joke control texts elicited a larger increase of the pupil sizes than the revision and revision control stories taken together. Similarly, the main effect of item type and the interaction between item type and IP indicates larger pupil size increases for the experimental items than for the control texts. Despite the fact that both of these effects are driven by the lack of an increase in the revision control stories ($t < 1$, n. s.),—compared to highly significant increases for the other three conditions ($|t|$'s > 4.7, p's < 0.0001)—, the corresponding 3-way interaction between IP, text category and item type was far from significant.

The performance in the picture matching task did not influence pupil size. However, a main effect for trial (serial order) was present as mean pupil sizes decreased significantly over the course of the experiment. Most importantly, the perceived funniness of the trials influenced pupil dilation, as indicated by a significant main effect. In trials with higher funniness ratings, the participant's pupils were larger.

To investigate whether this effect was due to a reaction to the crucial information in the punchline, we conducted a further analysis for the experimental trials only (jokes and revision texts). The experimental texts elicited a wide range of funniness ratings, while over 85% of the control texts were rated very low (1-3). The model corresponds to the previous one but replaces the

categorical variable item type by the funniness ratings. Its fixed effects account for 18.8% of the variance, including random effects yields 36.0%.

The results of this analysis are presented in **Table 4**. The effect of sense of humour indicates that participants who gave higher ratings overall had larger pupil sizes. The main effects of IP, funniness and trial were replicated. Furthermore, the highly significant interaction between funniness and IP confirmed that pupil sizes dilated more as a reaction to crucial information when the story was later rated as funnier. This effect is illustrated in **Figure 5**, where the continuous variable funniness is combined to three categories for sake of readability.

Text category did not have an effect, nor did interact with other variables. Thus, the distinction between experimental jokes and revision stories was not as important as the individually experienced funniness of the story.

4 DISCUSSION

In the present study a visual world paradigm was used to analyse the time course of situation model updating during the comprehension of verbal humour. Jokes were contrasted with non-humorous revision texts that also contained an incongruity, and with straightforward stories without the need for revision. The viewing patterns were comparable for jokes and the structurally identical revision stories.

4.1 Funniness Ratings and Picture Matching Performance

As in the studies of Siebörger (2006) and Ferstl et al. (2017) jokes were rated funnier than revision and control stories, despite the structural similarity of jokes and revision stories. This finding confirms that incongruency might be necessary for making a text funny, but that it is not sufficient. Thus, a differentiation of the cognitive and the affective aspects of humour processing became possible. The differences in funniness between the two text categories are consistent with differences in script opposition (Suls, 1972; Raskin, 1979). For example, in the revision story presented in **Table 1** the two narrative schemes only differ with respect to the agent of the scene (the cook or his wife preparing the meal). In the joke story on the other hand, Mark is introduced as a womanizer, but eventually turns out to be a thief. This more pronounced script opposition elicits an affective reaction.

Picture matching was expected to be a straightforward task. And indeed, it did not pose any problems for the revision stories and the coherent control stories. Surprisingly, however, the accuracy was considerably lower in the joke condition. In previous reading studies with similar materials, the revision stories had been most difficult (e.g., Ferstl et al., 2017). One explanation for this finding is that in some jokes the alternative situations reflect different perspectives on the same state of affairs, i.e., the joke is based on a misunderstanding between two protagonists. Consider the following example:

The mother explains to her son: Today mom and dad are married for 10 years. The son: And for how much longer do you have to stay together?

In this joke, the correct choice in the picture matching task is the meaning implicated by the punchline: the son views his parents' marriage as an unpleasant commitment. However, the funniness of the story depends on the incompatibility of the son's attitude with the attitude of his mother who regards her anniversary as a joyful event. Thus, the initial situation remains active during the processing of the punchline, so that both the competitor and the target are plausible choices in the picture matching task.

To control for the individual differences in picture matching performance incorrect trials were excluded and the participant's picture matching performance for jokes was added as a continuous covariate in all statistical analyses. As this covariate did not yield any significant effects, we are confident that the eye tracking results were independent of task induced strategies.

4.2 Eye Tracking: Viewing Times

The dwell times for the unambiguous coherent control texts showed that, as expected, the differentiation between the target picture and the competitor was completed before the revision point in the punchline. Because only the target picture was relevant for the interpretation of these stories, the competitor picture was viewed as little as the distractor picture. As the punchline did not contain any incongruous information, there were no systematic changes in the fixations patterns at the revision point. In the jokes and revision stories on the other hand, participants fixated both the competitor and the target pictures before the crucial information was presented at the revision point. A rise of target fixations and a switch from the initial to the final meaning of the story occurred within 1,500 ms after the revision point. Using ERPs, Canal et al. (2019) found that the processing sequence consisting of incongruency detection, incongruency resolution and later interpretation was completed within this timeframe.

The time window is also in line with a study by Fiacconi and Owen (2015) who located the moment of insight, i.e., the instant a verbal joke is understood, at around 800 ms after reading the punchline. In their study participants also heard jokes and unfunny but ambiguous control texts with an incongruity arising at the very end of each story. Examples are:

Joke: What did the teddy bear say when he was offered a dessert? No thank you, I'm stuffed!
Control: What was the problem with the other coat? It was very difficult to put it on with the paint roller!

EMG measurements on the musculus zygomaticus showed that first indications of a smile, interpreted an evidence for joke appreciation, were found 800 ms after reading the punchline of jokes. Fiacconi and Owen (2015) also reported a cardiovascular reaction about 5,000 ms after the punchline that correlated with the perceived funniness.

It would be desirable to describe the temporal sequence of processing in a more fine-grained way. However, the exact

timecourse depends on properties of the texts and the experimental procedure. Independent of the funniness of the texts, situation model building and inferencing, as required for the comprehension of texts, depend on the exact wording, the context and the experimental setting.

The fact that target fixations in experimental jokes and revision stories in the present study were already high before the revision point indicates that the correct meaning was anticipated before the occurrence of the incongruity. This anticipation was possibly elicited by the visual material that depicted both alternative interpretations.

The viewing patterns of the two types of experimental texts were indistinguishable. There was no evidence for faster incongruity resolution or longer target dwell times in jokes. Although this null effect cannot be interpreted as providing evidence for similarity—but only as a lack of difference in this particular experiment—it is consistent with the interpretation that the linguistic revision drives the viewing pattern. This result, together with the finding of comparable first pass reading times in the previous study, suggests that the observed differences between these texts in self-paced reading might have reflected meta-cognitive strategies. The affective reaction in jokes presumably provided intrinsic feedback about the correctness of the incongruity resolution, leading to shorter overall reading times for jokes compared to revision stories in the previous reading study (Ferstl et al., 2017).

4.3 Eye Tracking: Pupil Dilation

The analysis of the pupil sizes added important information to the viewing patterns. Higher mean pupil sizes were present on the target picture compared to competitor and distractor pictures for experimental jokes and revision stories. Somewhat surprisingly, a larger pupil size on the target IA was also observed for joke control texts that neither contained an incongruity nor a funny punchline that could account for an increase in pupil size on the target.

The revision control texts, on the other hand, showed pupil size means below the participant's average for all three IA. As all control texts were created in exactly the same way, by changing the context of the experimental text to activate the final situation model, these differences are likely to be due to the visual displays. Although the visual salience was comparable, the content of the target and competitor scenes created for the joke texts reflected the script opposition and were thus more distinct, and possibly more interesting, than those created for the revision texts. It is also possible that the picture material used for the experimental jokes as well as for the matching joke control texts was more funny in general, triggering an affective reaction for the control stories even though the corresponding text did not contain a funny punchline.

The two-way interactions of interest period with text category and item type, respectively, indicated that pupil sizes increased to some extent for all texts, except for the revision control texts. This finding is in line with a sensitivity to revision demands, as well as to affective aspects. And in fact, the main effect of the funniness ratings on the pupil sizes during text comprehension confirms that changes in pupil size were influenced by content that the participants reported to elicit affective reactions (e.g., Partala and Surakka, 2013). The pupils dilated more for funnier stories. An

additional analysis on the experimental trials confirmed that the pupil dilation varied with the funniness of the story, rather than with the predefined text category. A similar result was reported by Mayerhofer and Schacht (2015) who measured pupil diameters during the reading of jokes and control texts. The pupil sizes increased about 800 ms after humourous endings, compared to coherent control texts. Similar to our findings, this effect correlated with the funniness of the jokes, as determined in independent ratings.

The estimated mean pupil sizes decreased over the course of the experiment which could indicate that the cognitive load decreased as participants got used to the task requirements and stories. Moreover, as participants might have anticipated the punchline of jokes more strongly over the course of the experiment based on these learning effects, the strength of the affective reaction might have decreased as well, as the funniness response in jokes is assumed to be a result of the surprising occurrence of the incongruity (Dynel, 2009; Canal et al., 2019).

4.4 The Visual World Paradigm

The present study confirms that the visual world paradigm is a useful tool for studying situation model updating, and, in particular, the processing of humorous language. The fine-grained temporal resolution of the viewing patterns allowed us to describe the time course of comprehension in detail. Auditory presentation is a rather natural presentation modality, and reading strategies did not play a role. Finally, appropriate control conditions eliminated lexical or sentence level effects and deconfounded joke appreciation and cognitive revision. Moreover, the analysis of pupil sizes provided additional information about the affective component of humour processing. The perceived funniness of the story was accompanied by larger pupil dilations, and joke and joke control texts—sharing more interesting picture displays—also elicited larger increases in pupil sizes. Applying this method to other issues in humor comprehension research can aid in further developing theories of humor processing that take into account the interplay between cognitive demands and affective reactions.

4.5 Limitations

Although the visual world paradigm is clearly appropriate for studying verbal humour, the unexpected differences between the two types of control texts indicate that the content of the visual displays influenced processing of the verbal input which was presented later. Further research is needed to understand the interplay between features of the visual scenes with the auditory language input during higher level language comprehension (Hüttig et al., 2011).

A further limitation of the present study is the small number of items used. An increase in experimental power would be desirable. However, the number of eight trials per condition is not unusual for visual world experiments. In addition to the constraints for designing materials for studies on joke comprehension, a visual world study requires visualizable texts for which both competing interpretations can be visualized in comparable pictures.

Nevertheless, due to the repeated-measures design of the study, a substantial number of observations was collected for the analysis

of the behavioural and eye-tracking data. Because of the rather small number of observations on the subject level though, we could not explore the effects of gender on the eye-tracking dynamics of joke comprehension. As effects of gender on joke comprehension have been reported in previous studies though (Ferstl et al., 2017), future studies should also investigate these.

Because of the complex experimental design, the hierarchical structure of the repeated-measures data and the requirements of the visual world paradigm, we implemented a simplified analysis of the temporal sequence of fixations. This choice allowed us to pinpoint the effects of the information presented at the revision point. A more fine-grained analysis of the timecourse (for example using growth curve analysis, Mirman et al., 2008) would enable us to compare the present results to those using different methods, such as ERPs (Canal et al., 2019) or EMG (Fiacconi and Owen, 2015). The present study, however, can be viewed as a first exploratory analysis of the temporal progression of joke comprehension using the visual world paradigm and can be used for generating new hypotheses in this new field of research.

5 CONCLUSION

The present visual world study showed that jokes and non-funny revision texts involve similar cognitive processes. The viewing patterns confirmed that after the occurrence of an incongruity the initial situation model is revised und replaced by the globally correct one. As there was no facilitation effect for jokes, the processing advantages for jokes found in reading studies are likely to be due to meta-cognitive evaluation processes. Importantly, pupil size analyses can shed light on the affective component of the processing of verbal humour. The combination of these information sources from visual world experiments provides a promising tool for studying the interplay between cognitive and affective aspects of humor processing.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in the OSF repository (https://osf.io/qk95j/).

AUTHOR CONTRIBUTIONS

The experiment was conducted by LI in partial fulfilment of the requirements of a M. Sc. degree in Cognitive Science at the Albert-Ludwigs-University in Freiburg. All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

REFERENCES

Adobe Inc. (2019). *Adobe Illustrator*. Retrieved from https://adobe.com/products/illustrator.

Altmann, G. T. M., and Kamide, Y. (2007). The Real-Time Mediation of Visual Attention by Language and World Knowledge: Linking Anticipatory (And Other) Eye Movements to Linguistic Processing. *J. Mem. Lang.* 57, 502–518. doi:10.1016/j.jml.2006.12.004

Altmann, G. T. M. (2011). "The Mediation of Eye Movements by Spoken Language," in *The Oxford Handbook of Eye Movements*. Editors S. P. Liversedge, I. D. Gilchrist, and S. Everling. Oxford University Press, 979–1003.

Attardo, S., and Raskin, V. (1991). Script Theory Revis(it)ed: Joke Similarity and Joke Representation Model. *Humour* 4 (3), 293–347. doi:10.1515/humr.1991.4.3-4.293

Attardo, S. (1997). The Semantic Foundations of Cognitive Theories of Humor. *Humour* 10 (4), 395–420. doi:10.1515/humr.1997.10.4.395

Barr, D. J. (2008). Analyzing 'visual World' Eyetracking Data Using Multilevel Logistic Regression. *J. Mem. Lang.* 59, 457–474. doi:10.1016/j.jml.2007.09.002

Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). lme4: Linear Mixed-Effects Models Using Eigen and S4. R Package Version 1.1-7. Available at: http://CRAN.R-project.org/package=lme4.

Berends, S. M., Brouwer, S. M., and Sprenger, S. A. (2015). "Eye-tracking and the Visual World Paradigm," in *Designing Research on Bilingual Development: Behavioural and Neurolinguistic Experiments*. Editor M. Schmid, et al. (Cham: Springer), 55–80. doi:10.1007/978-3-319-11529-0_5

Bradley, M. M., Codispoti, M., Cuthbert, B. N., and Lang, P. J. (2001). Emotion and Motivation I: Defensive and Appetitive Reactions in Picture Processing. *Emotion* 1, 276–298. doi:10.1037/1528-3542.1.3.276

Bradley, M. M., Miccoli, L., Escrig, M. A., and Lang, P. J. (2008). The Pupil as a Measure of Emotional Arousal and Autonomic Activation. *Psychophysiology* 45, 602–607. doi:10.1111/j.1469-8986.2008.00654.x

Bressler, E. R., and Balshine, S. (2006). The Influence of Humor on Desirability. *Evol. Hum. Behav.* 27 (1), 29–39. doi:10.1016/j.evolhumbehav.2005.06.002

Buchowski, M. S., Majchrzak, K. M., Blomquist, K., Chen, K. Y., Byrne, D. W., and Bachorowski, J. A. (2007). Energy Expenditure of Genuine Laughter. *Int. J. Obes. (Lond)* 31, 131–137. doi:10.1038/sj.ijo.0803353

Canal, P., Bischetti, L., Di Paola, S., Bertini, C., Ricci, I., and Bambini, V. (2019). 'Honey, Shall I Change the Baby? - Well Done, Choose Another One': ERP and Time-Frequency Correlates of Humor Processing. *Brain Cogn.* 132, 41–55. doi:10.1016/j.bandc.2019.02.001

Cooper, R. M. (1974). The Control of Eye Fixation by the Meaning of Spoken Language. *Cogn. Psychol.* 6, 84–107. doi:10.1016/0010-0285(74)90005-x

Coulson, S. (2006). Constructing Meaning. *Metaphor and Symbol* 21 (4), 245–266. doi:10.1207/s15327868ms2104_3

Coulson, S., and Kutas, M. (1998). Frame-shifting and Sentential Integration. *UCSD Cogn. Sci. Tech. Rep.* 98 (3), 1–32.

Coulson, S., and Kutas, M. (2001). Getting it: Human Event-Related Brain Response to Jokes in Good and Poor Comprehenders. *Neurosci. Lett.* 316, 71–74. doi:10.1016/s0304-3940(01)02387-4

Coulson, S., and Lovett, C. (2004). Handedness, Hemispheric Asymmetries, and Joke Comprehension. *Cogn. Brain Res.* 19, 275–288. doi:10.1016/j.cogbrainres.2003.11.015

Coulson, S., Urbach, T., and Kutas, M. (2006). Looking Back: Joke Comprehension and the Space Structuring Model. *Humour* 19 (3), 229–250. doi:10.1515/humor.2006.013

Coulson, S., and Williams, R. F. (2005). Hemispheric Asymmetries and Joke Comprehension. *Neuropsychologia* 43, 128–141. doi:10.1016/j.neuropsychologia.2004.03.015

Dynel, M. (2009). Beyond a Joke: Types of Conversational Humour. *Lang. Linguistics Compass* 3 (9), 1284–1299. doi:10.1111/j.1749-818x.2009.00152.x

Engelhardt, P. E., Ferreira, F., and Patsenko, E. G. (2010). Pupillometry Reveals Processing Load during Spoken Language Comprehension. *Q. J. Exp. Psychol.* 63, 639–645. doi:10.1080/17470210903469864

Experiment Builder [Computer software] (2014). Available at: http://www.sr-research.com.

EyeLink Data Viewer User's Manual (2011). *Document Version 1.11.1*. Mississauga, ON, Canada: SR Research Ltd.

Ferstl, E. C., Israel, L., and Putzar, L. (2017). Humor Facilitates Text Comprehension: Evidence from Eye Movements. *Discourse Process.* 54 (4), 259–284. doi:10.1080/0163853X.2015.1131583

Ferstl, E. C., and Kintsch, W. (1999). "Learning from Text: Structural Knowledge Assessment in the Study of Discourse Comprehension," in *The Construction of Mental Models during reading*. Editors H. Oostendorp, and S. Goldman (Mahwah, NJ: Lawrence Erlbaum), 247–277.

Fiacconi, C. M., and Owen, A. M. (2015). Using Psychophysiological Measures to Examine the Temporal Profile of Verbal Humor Elicitation. *PLoS ONE* 10 (9), e0135902. doi:10.1371/journal.pone.0135902

Fox, J., and Weisberg, S. (2011). *An {R} Companion to Applied Regression*. 2nd ed. Thousand Oaks CA: Sage. Available at: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

Giora, R. (1991). On the Cognitive Aspects of the Joke. *J. Pragmatics* 16, 465–485. doi:10.1016/0378-2166(91)90137-m

Giora, R. (2002). "Optimal Innovation and Pleasure," in *Processing of the April Fools' Day Workshop on Computational Humour. Trento, Itlay*. Editors O. Stock, C. Strapparva, and A. Nijholt (Trento, Italy: ITC-itst), 20 11–28.

Goel, V., and Dolan, R. J. (2001). The Functional Anatomy of Humor: Segregating Cognitive and Affective Components. *Nat. Neurosci.* 4, 237–238. doi:10.1038/85076

Grice, H. (1975). "Logic and Conversation," in *Syntax and Semantics*. Editors P. Cole, and J. Morgan (New York: Academic Press), 3, 45–58.

Hess, E. H., and Polt, J. M. (1960). Pupil Size as Related to Interest Value of Visual Stimuli. *Science* 132, 349–350. doi:10.1126/science.132.3423.349

Huettig, F., and Altmann, G. T. (2005). Word Meaning and the Control of Eye Fixation: Semantic Competitor Effects and the Visual World Paradigm. *Cognition* 96, B23–B32. doi:10.1016/j.cognition.2004.10.003

Huettig, F., Olivers, C. N. L., and Hartsuiker, R. J. (2011). Looking, Language, and Memory: Bridging Research from the Visual World and Visual Search Paradigms. *Acta Psychologica* 137, 138–150. doi:10.1016/j.actpsy.2010.07.013

Hunger, B., Siebörger, F., and Ferstl, E. C. (2009). Schluss mit lustig: Wie Hirnläsionen das Humorverständnis beeinträchtigen [Done with fun: How brain lesions affect humour comprehension]. *Neurolinguistik* 21, 35–59.

Hüttig, F., Rommers, J., and Meyer, A. S. (2011). Using the Visual World Paradigm to Study Language Processing: A Review and Critical Evaluation. *Acta Psychologica* 137, 151–171. doi:10.1016/j.actpsy.2010.11.003

Hyönä, J., Tommola, J., and Alaja, A. (2007). Pupil Dilation as a Measure of Processing Load in Simultaneous Interpretation and Other Language Tasks. *Q. J. Exp. Psychol.* 48 (A), 598–612. doi:10.1080/14640749508401407

Kant, I. (1790/1977). *Kritik der Urteilskraft. Werkausgabe Band X*. Frankfurt am Main: Suhrkamp Verlag.

Kimata, H. (2004). Laughter Counteracts Enhancement of Plasma Neurotrophin Levels and Allergic Skin Wheal Responses by mobile Phone-Mediated Stress. *Behav. Med.* 29, 149–154. doi:10.3200/bmed.29.4.149-154

Lenth, R. (2014). Lsmeans: Least-Squares Means. R Package Version 2.10. Available at: http://CRAN.R-project.org/package=lsmeans.

Martin, A. (2007). *The Psychology of Humour: An Integrative Approach*. Burlington: Elsevier Academic Press.

Mayerhofer, B. (2015). *Universität Göttingen*. Unpublished dissertation. Available at: http://hdl.handle.net/11858/00-1735-0000-0022-5DC9-4.Processing of Garden Path Jokes: Theoretical Concepts and Empirical Correlates

Mayerhofer, B., and Schacht, A. (2015). From Incoherence to Mirth: Neuro-Cognitive Processing of Garden-Path Jokes. *Front. Psychol.* 6, 550. doi:10.3389/fpsyg.2015.00550

Mirman, D., Dixon, J. A., and Magnuson, J. S. (2008). Statistical and Computational Models of the Visual World Paradigm: Growth Curves and Individual Differences. *J. Mem. Lang.* 59 (4), 475–494. doi:10.1016/j.jml.2007.11.006

Mitchell, H. H., Graesser, A. C., and Louwerse, M. M. (2010). The Effect of Context on Humor: A Constraint-Based Model of Comprehending Verbal Jokes. *Discourse Process.* 47, 104–129. doi:10.1080/01638530902959893

Partala, T., and Surakka, V. (2003). Pupil Size Variation as an Indication of Affective Processing. *Int. J. Human-Computer Stud.* 59, 185–198. doi:10.1016/s1071-5819(03)00017-x

Pexman, P. M., Rostad, K. R., McMorris, C. A., Climie, E. A., Stowkowy, J., and Glenwright, M. R. (2011). Processing of Ironic Language in Children with High-Functioning Autism Spectrum Disorder. *J. Autism Dev. Disord.* 41 (8), 1097–1112. doi:10.1007/s10803-010-1131-7

Prehn, K., Kazzer, P., Lischke, A., Herpertz, S. C., Lischke, A., Heinrichs, M., et al. (2013). Effects of Intranasal Oxytocin on Pupil Dilation Indicate Increased Salience of Socioaffective Stimuli. *Psychophysiol* 50, 528–537. doi:10.1111/psyp.12042

Pyykkönen, P., Hyönä, J., and van Gompel, R. P. G. (2010). Activating Gender Stereotypes during Online Spoken Language Processing: Evidence from Visual World Eye-Tracking. *Exp. Psychol.* 57, 126–133.

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Raskin, V. (1979). Semantic Mechanisms of Humor. *Bls* 5, 325–335. doi:10.3765/bls.v5i0.2164

Salverda, A. P., and Tanenhaus, M. K. (2018). "The Visual World Paradigm," in *Research Methods in Psycholinguistics and the Neurobiology of Language: A Practical Guide*. Editors A. M. B. de Groot, and P. Hagoort (Malden, MA: Wiley-Blackwell), 89–110.

Schmidt, S. R. (2002). The Humour Effect: Differential Processing and Privileged Retrieval. *Memory* 10, 127–138. doi:10.1080/09658210143000263

Searle, J. (1979). *Expression and Meaning: Studies in the Theory of Speech Acts*. Cambridge: Cambridge University Press.

Siebörger, F. T. (2006). *Funktionelle Neuroanatomie des Textverstehens. Kohärenzbildung bei Witzen und anderen ungewöhnlichen Texten [*Functional neuroanatomy of text comprehension: Coherence building in jokes and other unusual texts*]*. Dissertation. Universität Leipzig, Leipzig: MPI Series in Human Cognitive and Brain Sciences.

Sirois, S., and Brisson, J. (2014). Pupillometry. *Wires Cogn. Sci.* 5, 679–692. doi:10.1002/wcs.1323

Strick, M., Holland, R. W., van Baaren, R., and van Knippenberg, A. (2010). Humor in the Eye Tracker: Attention Capture and Distraction from Context Cues. *J. Gen. Psychol.* 137, 37–48. doi:10.1080/00221300903293055

Suls, J. M. (1972). "A Two-Stage Model for the Appreciation of Jokes and Cartoons: An Information-Processing Analysis," in *The Psychology of Humour. Theoretical Perspectives and Empirical Issues*. Editors P. Goldstein, and J. McGhee (New York: Academic Press), 81–100. doi:10.1016/b978-0-12-288950-9.50010-9

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., and Sedivy, J. C. (1995). Integration of Visual and Linguistic Information in Spoken Language Comprehension. *Science* 268 (5217), 1632–1634. doi:10.1126/science.7777863

van Dijk, T. A., and Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York, NY: Academic Press.

Volkmann, B., Siebörger, F., and Ferstl, E. C. (2008). *Spass Beiseite? [All Joking Aside?* Hofheim: NAT-Verlag.

Winn, B., Whitaker, D., Elliott, D. B., and Phillips, N. J. (1994). Factors Affecting Light-Adapted Pupil Size in Normal Human Subjects. *Invest. Ophthalmol. Vis. Sci.* 35 (3), 1132–1137.

Wong, A. Y., Moss, J., and Schunn, C. D. (2016). Tracking reading Strategy Utilisation through Pupillometry. *Ajet* 32 (6), 45–57. doi:10.14742/ajet.3096

Zwaan, R. A., Stanfield, R. A., and Yaxley, R. H. (2002). Language Comprehenders Mentally Represent the Shapes of Objects. *Psychol. Sci.* 13 (2), 168–171. doi:10.1111/1467-9280.00430