The Quarterly Journal of Experimental Psychology

ISSN: 1747-0218 (Print) 1747-0226 (Online) Journal homepage: http://www.tandfonline.com/loi/pqje20

Pupillometry reveals increased pupil size during indirect request comprehension

Johanne Tromp, Peter Hagoort & Antje S. Meyer

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2016
Vol. 69, No. 6, 1093–1108, http://dx.doi.org/10.1080/17470218.2015.1065282

Pupillometry reveals increased pupil size during indirect request comprehension

Johanne Tromp[1,2], Peter Hagoort[1,3], and Antje S. Meyer[1,3]

[1]Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands
[3]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

(Received 17 December 2014; accepted 15 June 2015; first published online 19 August 2015)

Fluctuations in pupil size have been shown to reflect variations in processing demands during lexical and syntactic processing in language comprehension. An issue that has not received attention is whether pupil size also varies due to pragmatic manipulations. In two pupillometry experiments, we investigated whether pupil diameter was sensitive to increased processing demands as a result of comprehending an indirect request versus a direct statement. Adult participants were presented with 120 picture–sentence combinations that could be interpreted either as an indirect request (a picture of a window with the sentence "it's very hot here") or as a statement (a picture of a window with the sentence "it's very nice here"). Based on the hypothesis that understanding indirect utterances requires additional inferences to be made on the part of the listener, we predicted a larger pupil diameter for indirect requests than statements. The results of both experiments are consistent with this expectation. We suggest that the increase in pupil size reflects additional processing demands for the comprehension of indirect requests as compared to statements. This research demonstrates the usefulness of pupillometry as a tool for experimental research in pragmatics.

Keywords: Pupillometry; Experimental pragmatics; Indirect requests; Speech act; Language comprehension.

Fluctuations in pupil size have been shown to reflect variations in processing demands during a number of cognitive tasks (Kahneman, 1973; for an overview see Beatty, 1982; Sirois & Brisson, 2014). For example, Hess and Polt (1964) recorded the pupil size of participants solving maths problems that varied in difficulty. Their findings indicated that solving more difficult problems was accompanied by larger pupil diameters. Pupil size has also been shown to vary with processing effort during visual search and counting tasks (Porter, Troscianko, & Gilchrist, 2007), digit list recall (e.g., Piquado, Isaacowitz, & Wingfield, 2010), and with working memory load (Attar, Schneps, & Pomplun, 2013).

More recently, pupillometry has been used to study language processing, and pupillary responses have been taken as an index of increases in processing demands during sentence comprehension

Correspondence should be addressed to Johanne Tromp, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH, Nijmegen, The Netherlands. E-mail: Johanne.Tromp@mpi.nl

We thank Markus van Ackeren for making the stimulus materials available, Svetlana Gerakaki for help with analysing the pupillometry data and Ronald Fisher for technical support.

(e.g., Just, Carpenter, & Miyake, 2003). These demands can arise in a number of ways. For instance, processing negative sentences as compared to affirmative sentences resulted in an increase in pupil size (Beatty, 1982). Also, Schluroff (1982) observed a larger pupil diameter during the processing of grammatically complex sentences as compared to their simpler counterparts. He suggested that pupil size may be of considerable use as an online monitor of cognitive load imposed by grammatical complexity. Similarly, Just and Carpenter (1993) observed a larger change in pupil diameter for object relative sentences (e.g., *The senator that the reporter attacked admitted the error*) than for subject relative sentences (e.g., *The senator that attacked the reporter admitted the error*), within 1.2 s after the critical verb.

In addition to grammatical complexity, pupil diameter is sensitive to lexical and syntactical ambiguity (Ben-Nun, 1986; Schluroff et al., 1986) and to prosody manipulations (e.g., Zellin, Pannekamp, Toepel, & Van der Meer, 2011). Engelhardt, Ferreira, and Patsenko (2010) investigated the processing of syntactically ambiguous sentences in relation to prosody and visual context. In their first experiment, participants were presented with garden-path sentences (e.g., "While the woman cleaned the dog that was big and brown stood in the yard") accompanied by correct prosody (i.e., a prosodic break between "cleaned" and "the dog") or conflicting prosody (i.e., no prosodic break). They predicted that, if prosody influences online processing of the sentence, an increase in pupil diameter should be observed for the incongruent sentence-prosody condition, indicating more cognitive effort. This prediction was confirmed, as pupil size reliably increased when prosodic structure was inconsistent with syntactic structure. In their second experiment, a visual context was added to the prosody manipulation. The visual context could be either congruent or incongruent with the correct interpretation of the sentence. The results indicated an interaction between prosody and visual context. When visual context was consistent with the correct interpretation of the sentence, prosody had little effect on processing effort. In contrast, when the visual context was inconsistent with the correct interpretation, prosody had an effect on processing effort. This suggests that, in addition to prosody, visual context affected online processing load as measured by pupil diameter change. In sum, there is strong evidence that pupil diameter during sentence comprehension is sensitive to differences in cognitive load resulting from increases in sentence complexity or ambiguity. An issue that has not received any attention is whether pupil size is also sensitive to pragmatic manipulations. In the current study, we investigated whether pupil diameter was sensitive to increased processing demands for nonconventional indirect requests compared to direct statements. During natural conversation, communication is often indirect. We might hint at what we want rather than expressing it directly. For example, in an appropriate context "It's cold in here" may be a request to shut the window, rather than a statement about the room temperature (Holtgraves, 1994). The way in which we comprehend the intended meaning of indirect speech acts, such as the indirect request in the example above, has been a topic of much debate (Holtgraves, 2002). However, most theories agree that understanding this type of request requires some form of intention recognition (Austin, 1962; Holtgraves, 1994; Levinson, 2000; Searle, 1975, 1979; Sperber & Wilson, 1986, 1995). The listener has to infer that the speaker intents to request something in order to interpret the utterance correctly. Indirect requests vary in their conventionality (e.g., Gibbs, 1986; Holtgraves, 1994, 2002). For example, "Can you pass the salt?" is an indirect request, but it is conventional. It has a literal meaning ("I ask if you are able to pass the salt") and an indirect meaning ("I request you to pass the salt"). Usually this type of request can be performed by asserting or questioning the felicity conditions that underlie requests (e.g., question the hearer's ability). Also, it contains the request-based propositional content (e.g., "pass the salt") and it allows the preverbal insertion of "please" (e.g., "Can you please pass the salt?"; Holtgraves, 2002). Research suggested that these conventional indirect requests are recognized fast (Gibbs, 1981, 1983) and immediately (Clark, 1979). Gibbs (1983) proposed that people do not always have to retrieve the literal meaning of conventional indirect requests first, but rather they can compute the indirect meaning automatically (see also Holtgraves, 1994). Thus, even though these types of sentences have two meanings, retrieving the indirect one does not seem to require additional processing effort on the part of the listener. Indirect requests that cannot be characterized by the features mentioned above are categorized as nonconventional (Holtgraves, 2002), and they are the focus of the current research. One common form of this type is a *negative state remark*, where the speaker asserts or questions a negative state (e.g., "It's cold in here"), which can be eliminated or lessened by the hearer (e.g., by closing the window; Holtgraves, 1994). Research showed that participants took longer to comprehend this type of request than the conventional type (e.g., Gibbs, 1981; Holtgraves, 1994), and it has been suggested that comprehending these requests involves an inference process (Holtgraves, 1994, 2002). During this process, listeners take into account information from other sources than the linguistic code, for example contextual factors and prior knowledge of the conversational partner's intentions (Holtgraves, 1994). That an additional inference process is necessary during comprehension of certain types of indirect speech acts is also supported by two recent neuroimaging studies. Bašnáková, Weber, Petersson, Van Berkum, and Hagoort (2014) conducted a functional magnetic resonance imaging (fMRI) study to investigate the neural underpinnings of inferring speaker meaning (i.e., the message of the speaker). Participants listened to sentences (e.g., "It's hard to throw a good party") that had different meanings depending on the dialogue and final question that preceded it. For example, the sentence mentioned above is a direct reply to the question "How hard is it to throw a party?", but it can also be an indirect reply to the question "Will you throw a party for your graduation?". Furthermore, it can be an indirect reply to the question "Did you enjoy yourself at my party?". In the latter case, the motivation of the speaker for using an indirect reply is "face saving", or to mutually protect another's public self (e.g., Brown & Levinson, 1987). For indirect replies as compared to direct replies, increased activation was found in areas relevant for discourse-level processing (bilateral prefrontal cortex and right temporal regions) and areas involved in mentalizing and empathy (medial frontal cortex, MFC; right temporoparietal junction, TPJ; and anterior insula). For face-saving replies, there was additional activation in regions involved in affective and social cognitive processing, such as insula and anterior cingulate cortex (ACC). Bašnáková and colleagues concluded, based on the activation pattern for the indirect replies, that when inferring speaker meaning, listeners take the speaker's perspective on both cognitive (theory of mind) and affective levels. Thus, comprehending indirect replies seems to rely on inferences made by the listener.

Most relevant for the present purposes is a study by Van Ackeren, Casasanto, Bekkering, Hagoort, and Rueschemeyer (2012). They investigated the neural correlates of indirect request (IR) comprehension. Participants were presented with picture–sentence combinations as shown in Figure 1. In each item set, the two sentences were combined with each of the two pictures, such that in one combination the utterance could be interpreted as an indirect request, whereas in the remaining combinations it was a statement. For example, "It's very hot here" in combination with a picture of a window may be interpreted as an indirect request to open the window, while the sentence "It's very nice here" with the same picture would most likely be interpreted as a mere statement. First, Van Ackeren et al. (2012) observed increased activation in cortical motor areas for indirect requests as compared to statements. In addition, they expected increased activation in theory of mind (ToM) areas, such as medial prefrontal cortex (mPFC) and TPJ (Gallagher & Frith, 2003), for indirect requests as compared to statements, since

Picture	Utterance	Condition
	"It is very hot here" *Action Utterance (AU)*	Indirect request (AP/AU) *Action Picture / Action Utterance*
Action Picture (AP)	"It is very nice here" *No-action Utterance (NU)*	Picture control (AP/NU) *Action Picture / No-action Utterance*
	"It is very hot here" *Action Utterance (AU)*	Utterance control (NP/AU) *No-action Picture / Action Utterance*
No-action Picture (NP)	"It is very nice here" *No-action Utterance (NU)*	Picture-utterance control (NP/NU) *No-action Picture / No-action Utterance*

Figure 1. *Illustration of the design with a single item set.*

making inferences about mental states of others has often been associated with having ToM. This prediction was confirmed: Both mPFC and left TPJ were sensitive to indirect requests versus statements. The authors concluded that, quite probably, these regions were crucial for making inferences about the communicative intent of speaker during IR comprehension.

In the present study, we used a subset of the stimuli created by Van Ackeren et al. (2012) and recorded the participants' pupil size while they listened to the sentences and viewed the pictures. One quarter of the scene–sentence combinations could be interpreted as indirect requests [e.g., a picture of a window (scene) and the sentence "It is very hot here"]. The other combinations served as controls for the indirect requests and could only be interpreted as statements [e.g., a picture of a window (scene) and the sentence "It is very nice here"]. In Experiment 1, following each picture–sentence pair, the participant had to indicate whether or not the utterance was an indirect request. In Experiment 2, a control experiment, participants were asked to make an affirmative response when they heard a direct statement.

Our first aim was to test whether pupil size would be sensitive to this difference in the implied meaning of the utterances. Since pupillometry is a relatively cheap noninvasive tool, it would be useful to demonstrate its applicability for studies of pragmatics. Our second aim was to investigate the cognitive effort involved in IR comprehension. Although the abovementioned fMRI studies are informative regarding the neural infrastructure supporting the comprehension of indirect utterances, they provide little information about the processing costs involved in understanding them. As noted above, there is strong evidence that pupil size is a good indicator of mental effort (e.g., Beatty, 1982; Engelhardt et al., 2010; Piquado et al., 2010). Thus, if deriving the meaning of indirect requests involves cognitive effort beyond the effort entailed in understanding mere statements, we should see this reflected in the participants' pupil size, which should be larger for the indirect requests than for the control combinations.

EXPERIMENT 1

Method

Participants

Forty-nine native speakers of Dutch participated in the study (nine men, mean age = 20.8 years, range = 18–26 years). All participants had normal hearing, normal or corrected-to-normal vision, and no history of language disorders. All but one were right-handed. Informed consent was obtained from all participants. They were paid for taking part in the experiment. Ethical approval for the study was granted by the ethics board of the Social Sciences Faculty of Radboud University.

Materials and design

Materials consisted of a subset[1] of the materials used by Van Ackeren et al. (2012), namely 120 images of visual scenes and 120 spoken sentences. The visual scenes were collected from publically available online search engines (e.g., flickr.com), and the sentences were recordings of a native speaker of Dutch. The stimuli were divided into 60 item sets. Each set (see Figure 1) consisted of two pictures and two sentences. Pictures were labelled "action picture" (AP) when they could appear in the IR (action) condition or "no-action picture" (NP) when they could only be in the statement conditions. The same was done for the two utterance types (AU = "action utterance," NU = "no-action utterance"). The pictures and sentences could be combined in four different ways, which resulted in four experimental conditions: indirect request (AP/AU), picture control (AP/NU), utterance control (NP/AU), and picture–utterance control (NP/NU). In a pretest, Van Ackeren et al. (2012) confirmed that the IR (AP/AU) sentence–scene combinations were interpreted as indirect requests more often than items in the other conditions. The control conditions were included to control for the unique effects of picture and utterance separately.

Since pupil size is sensitive to luminosity, luminosity values of the pictures were adjusted so that all pictures had values between 140 and 160. Luminosity was measured using the luminosity tool in Adobe Photoshop, Version 11.0.2. Picture size was kept relatively small (250 × 250 pixels), so that the larger part of the computer screen was white.

Each participant saw two scene–sentence combinations from each item set. For example, from the set in Figure 1, Participant 1 would see the AP/AU (indirect request) scene–sentence combination and the NA/NU (picture–utterance control) combination. Participant 2 would see the remaining combinations. Thus, two lists were created, and items were never repeated within participants. Each participant viewed 30 combinations per condition, resulting in a total of 120 trials. The sentence–scene combinations were distributed over four blocks (30 items per block), and they were pseudorandomized so that combinations from the same condition were never presented more than twice in a row. Between blocks, the participant was encouraged to take a short break. Before the experimental blocks, participants completed 12 practice trials.

Apparatus and procedure

Participants were seated in a medium-lit soundproof booth. The lighting was kept constant for all participants. Stimuli were presented using Experiment Builder Version 1.10.1025 (SR Research Ltd., Mississauga, Canada). The sentences were presented through Sennheiser HD201 lightweight over-ear binaural headphones. The pupil diameter of each participant's right eye was measured with an Eyelink 1000 Tower Mount eye-tracker (SR Research Ltd., Mississauga, Canada). In Eyelink 1000, pupil size is measured in arbitrary units that have a linear relation to the recorded pupil diameter (see Eyelink user manual; Einhäuser, Stout, Koch, & Carter, 2008). Before the start of the experiment, randomized target order calibration and validation routines were performed using EyeLink 1000 software (SR Research Ltd., Mississauga, Canada). Button presses were recorded by means of a button box. Each trial started with a fixation cross that was presented for 1000 ms, after which the visual scene appeared on the screen. Two-hundred ms later, the sentence was presented through the headphones. Then, the fixation cross appeared again for 2500 ms followed by the statement: "The person made a request." Participants then indicated whether or not they thought the statement was true (right button press) or false (left button press). After participants made their choice, the fixation cross appeared again for 2500

[1]Four item sets from the original study by Van Ackeren et al. (2012) were removed because in these sets the critical word was repeated, which could influence pupil dilation (e.g., Otero, Weekes, & Hutton, 2011). We randomly selected three of these four item sets to use in the practice block.

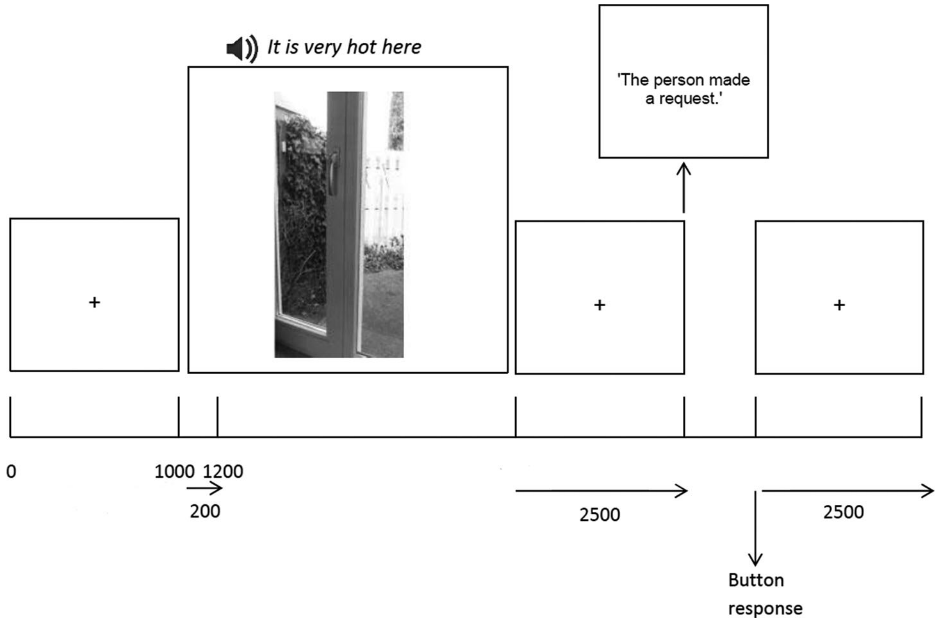

Figure 2. *Example of the time-course of a single trial (time in ms).*

ms to give the pupil enough time to return to baseline before the next trial (see Figure 2).

Behavioural data analysis

A one-sample *t* test (test value = 0.5) was conducted on the correct responses to the AP/AU combinations to assess whether participants performed above chance in identifying the indirect requests. Further analysis of comprehension accuracy was conducted with logit mixed models in R (Jaeger, 2008). Predictors were mean-centred. The model included the fixed effects picture (action, no-action), and utterance (action, no-action) and the interaction. Also it contained random intercepts and slopes for picture and utterance by participant and a by-item (picture) random intercept for the effect of utterance. This was the maximal random structure justified by the data leading to convergence (Barr, Levy, Scheepers, & Tily, 2013).

Reaction times were analysed with linear mixed effects models in R (Version 3.0.3; The R foundation for statistical computing; lme4 package, Bates, Maechler, Bolker, & Walker, 2014). The model included the fixed effects picture (action, no-action), and utterance (action, no-action) and the interaction. The random structure of the model was the same as that for the comprehension accuracy analysis described above.

Pupillometry data analysis

Pupillometry data were pre-processed and analyzed in R (version 3.0.3). The R-scripts for the signal pre-processing procedure were developed by Gerakaki, Sjerps, and Meyer (in press). Pupil dilation was originally measured with a sampling rate of 500 Hz, for the analysis the signal was down-sampled to 50 Hz. To detect and remove outliers, the change in pupil diameter was assessed from sample to sample. Based on Piquado et al. (2010), all data points with a ratio that differed more than one standard deviation from the mean pupil change of the trial, were categorized as outliers. Outliers were treated as missing values and linear interpolation was used to replace them. Trials were completely removed from the analysis if more than 25% of values were missing (3.4% of the data).

On a trial by trial basis, absolute pupil diameter was transformed to relative pupil diameter by means of baseline-correction and normalization. This was done to correct for tonic changes in pupil dilation and to allow for a comparison between participants (e.g., Van Rijn, Dalenberg, Borst, & Sprenger, 2012). First, the baseline pupil size of a given trial was subtracted from each sample in the trial. These values were then divided by the baseline to calculate the pupil size change. The baseline was defined as the average pupil size during the first 1000 ms of a trial. In this time window, a fixation cross was presented on the screen. To plot the task-evoked pupillary responses (TEPRs), which represent the percentage of pupil diameter change (PDC) over time, the value of pupil diameter change was multiplied by 100. For the statistical analyses, each trial was partitioned into four parts; baseline (0 to 1000 ms), audio (1000 ms to critical word onset (M = 2337)), critical (1500 ms window from critical word onset), end of trial (from end of critical window to trial offset). The choice for a critical window of 1500 ms from word onset was based on a study by Just and Carpenter (1993), which found the largest peak in pupil size ~1.2 seconds after the critical word offset in an ambiguous sentence. We took a time window of 1.5 seconds starting from critical word onset, adopting the 1.2 seconds window of Just and Carpenter (1993) plus 300 ms for word recognition (see Engelhardt et al., 2010 for a similar approach).

Mean pupil size was analysed using linear mixed effects models in R (Version 3.0.3; The R foundation for statistical computing; lme4 package, Bates et al., 2014), which allow for simultaneous inclusion of items and participants as random variables (Baayen, Davidson, & Bates, 2008). Statistical analyses were performed only for the critical time window. The predictors (picture, utterance) and the random structure were the same as those in the logit mixed model for the comprehension accuracy data. To assess the effects of utterance, picture, and the interaction, a backwards elimination procedure was used in which models were compared using a likelihood ratio test. The same procedure was followed for the peak pupil size data. Peak pupil size was included as an additional dependent variable in this study, since this measure has been shown to be less dependent on the number of observations in the critical time window than mean pupil size (Beatty & Lucero-Wagoner, 2000). This is important, since in our study indirect requests occur less frequently than statements (ratio: 1:3). Although recent papers (e.g., Wierda, van Rijn, Taatgen, & Martens, 2012, for an overview see Sirois & Brisson, 2014) proposed more advanced analyses of the time-course of pupillary responses, the analyses reported here are sufficient for the purposes of the current research.

Results

Behavioural results

Utterances were categorized as requests more often in the IR condition than in the control conditions (see Table 1). In line with the results reported by Van Ackeren et al. (2012), a one-sampled t test (test value = 0.5) confirmed that participants were able to correctly identify the AP/AU combinations as indirect requests, $t(48) = 14.64$, $p < .001$ ($M =$ 76.97%, $SE = 1.12\%$).

The logit mixed-effect model for comprehension accuracy ($n = 49$) indicated a significant effect of picture ($\beta = -1.072$, $SE = 0.123$, $z = -8.750$, $p < .001$). Accuracy was lower for the action pictures ($M = 74.89\%$, $SE = 0.82\%$) than for the no-action pictures ($M = 84.42\%$, $SE = 0.68\%$). There was no effect of utterance ($\beta = 0.017$, $SE = 0.225$, $z = 0.076$, $p = .94$), nor an interaction between picture and utterance ($\beta = 0.411$, $SE = 0.278$, $z = 1.475$, $p = .14$).

There was substantial variation in accuracy rates across participants. For the first analysis of the pupillometry data reported below, we selected only participants with accuracy rates of 70% or higher for each condition. This criterion was used to make sure that we captured IR comprehension in the AP/AU condition and to ensure that after removal of incorrect trials, participants still contributed similar numbers of data points to the analysis. For this group of 22 participants, the pattern of comprehension accuracy across conditions was

Table 1. *Experiment 1: Percentage of IR responses ("yes" to the statement "The person made a request") per condition for the entire sample (top row, n = 49) and for the subset (bottom row, n = 22)*

		Indirect request (AP/AU)		*Picture control (AP/NU)*		*Utterance control (NP/AU)*		*Picture–utterance control (NP/NU)*	
Condition	*n*	*%*	*SE*	*%*	*SE*	*%*	*SE*	*%*	*SE*
IR responses	49	76.97	1.12	27.22	1.19	16.56	0.99	14.62	0.93
IR responses	22	80.97	1.55	16.93	1.50	8.78	1.12	9.06	1.14

Note: AP = action picture; NP = no-action picture; AU = action utterance; NU = no-action utterance.

similar to the pattern for the entire sample (see bottom row of Table 1). In the statistical analysis of the comprehension accuracy data, we again only found an effect of picture ($\beta = -0.789$, $SE = 0.198$, $z = -3.981$, $p < .001$). Accuracy was lower for action pictures ($M = 82.00\%$, $SE = 1.08\%$) than for no-action pictures ($M = 91.08\%$, $SE = 0.80\%$). There was no effect of utterance ($\beta = -0.069$, $SE = 0.245$, $z = -0.281$, $p = .78$), nor an interaction between picture and utterance ($\beta = -0.257$, $SE = 0.330$, $z = -0.778$, $p = .44$). For the reaction times, the best fitting model included the interaction between picture and utterance ($\beta = -2.134$, $SE = 36.918$, $t = -2.134$). There was no evidence for a main effect of picture ($t < 1.4$), nor of utterance ($t < 1$). Closer inspection of the interaction revealed a trend ($p = .07$) for shorter reaction times for the AP/AU ($M = 524$, $SE = 16$) than for the AP/NU ($M = 587$, $SE = 18$) combinations. There was no difference for the no-action pictures—that is, between NP/AU ($M = 599$, $SE = 18$) and NP/NU ($M = 568$, $SE = 13$).

Pupillometry results

The results reported here are based on the trials with correct responses. However, the same pattern is present when all trials are included in the analysis. Visual inspection of the TEPRs, which show the percentage of pupil diameter change (PDC) per condition (see Figure 3), suggested a larger pupil diameter for the AP/AU combinations (indirect requests) than for all all other conditions in the critical time-window (1.5 s after critical word onset).

No difference between conditions was observed in the preceding time window (audio). In the last window (end of trial), a larger mean pupil was still observed for the AP/AU combinations than for all other conditions. Also, the AP/NU condition showed a slightly higher mean than the other control conditions.

The statistical analyses revealed that, for the critical time-window, the optimal model for the mean pupil size contained the interaction between picture and utterance ($\beta = 0.012$, $SE = 0.005$, $t = 2.051$). Including the interaction significantly improved model fit, $\chi^2(1) = 4.182$, $p < .05$. There was no evidence for the fixed effects of picture and/or utterance (all $t < 1$). Further inspection of the interaction, based on planned comparisons, revealed an effect of utterance type (action versus no-action) for the action pictures, $t(21) = 2.78$, $p < .02$, but not for the no-action pictures, $t(21) = -1.234$, $p = .23$. For the action pictures, mean pupil size was larger for action utterance—that is, indirect requests ($M = 0.051$, $SE = .003$)—than for no-action utterances ($M = 0.041$, $SE = .004$).

In the model for the peak pupil size there was no evidence for a main effect of picture ($\beta = 0.004$, $SE = 0.004$, $t = 1.197$), or utterance ($\beta = 0.002$, $SE = 0.005$, $t = .043$). However, as for the mean pupil diameter, there was evidence for the interaction between picture and utterance ($\beta = 0.014$, $SE = 0.006$, $t = 2.222$). Including the interaction improved the fit of the model, $\chi^2(1) = 4.960$, $p < .03$. Further examination of the interaction revealed a difference between utterance types (action versus no-action) for the action pictures, $t(21) = 3.434$, $p < .005$, but not for the no-action pictures, $t(21) = -1.131$,

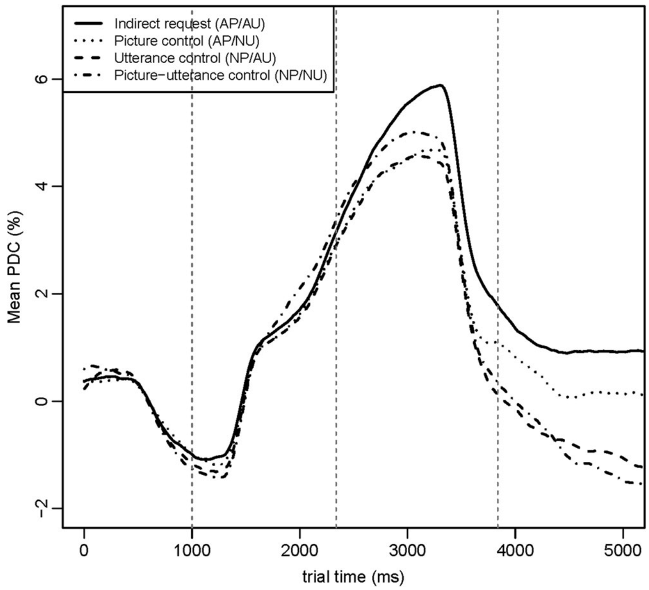

Figure 3. *Average percentage of pupil diameter change (PDC) as a function of condition (indirect request, IR; picture control, PC; utterance control, UC; picture–utterance control, PUC). Trial time (ms) is represented on the* x*-axis and pupil diameter change (%) on the* y*-axis. The vertical lines represent the different time windows: baseline (0 to 1000 ms), audio (1000 ms to critical word onset, M = 2337), critical (critical word onset + 1500-ms window after critical word onset), and end of trial (from end of the critical time window to end of trial).*

$p = .27$. In line with the results of the analysis of the mean pupil diameter, the peak for the AP/AU combinations (i.e., the indirect requests) was larger ($M = 0.080$, $SE = 0.003$) than the peak for the AP/NU combinations ($M = 0.067$, $SE = 0.003$). In sum, the results indicated that there was an effect of utterance type for the action pictures, but not for the no-action pictures, for both the mean and peak pupil diameter. Namely, indirect requests, or the unique combination of action utterances with action pictures, resulted in a larger mean and peak pupil diameter than control combinations.

In the above analyses, we compared the pupil sizes for correct responses over the four predefined experimental conditions, or, picture–sentence combinations. In the following analyses we compared the pupil sizes for the two response types (indirect request vs. statement) regardless of the stimulus condition. In other words, we compared trials where participants did versus did not indicate that they had heard an indirect request. All trials ("yes" response = 1923, "no" response = 3757) from all 49 participants were included.

The preprocessing procedure was the same as that for the first analysis, except that the data were split according to response rather than experimental condition (see Figure 4). The mixed-effects model thus contained only one predictor: response (indirect request versus statement). The random structure of the model was the same as that in all other analyses, and models were again compared using a likelihood ratio test.

Pupillometry results: Response-split

Visual inspection of the PDC (see Figure 4) suggested a higher mean and peak pupil size for indirect requests than for statements. This was confirmed by the analysis. The optimal model for the mean pupil size during the critical time-window included the fixed effect of response (indirect request versus statement), $\beta = 0.006$, $SE = 0.001$, $t = 3.263$. Indirect requests resulted in a larger mean pupil diameter than statements (indirect request: $M = 0.042$, $SE = 0.001$; statement: $M = 0.039$, $SE = 0.001$). Including the effect improved the fit of the model, $\chi^2(1) = 10.492$, $p < .01$.

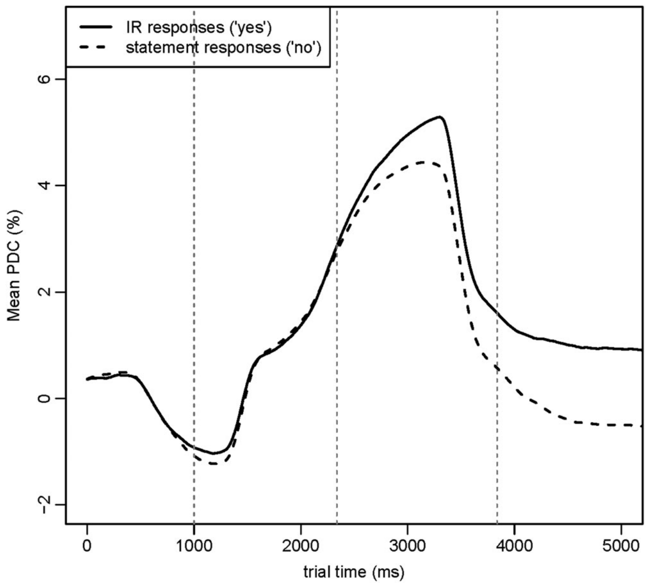

Figure 4. *Experiment 1: Average percentage of pupil diameter change (PDC) as a function of response (indirect request, IR, versus statement), n = 49. Trial time (ms) is represented on the* x*-axis and pupil diameter change (%) on the* y*-axis. The vertical lines represent the different time windows: baseline (0 to 1000 ms), audio (1000 ms to critical word onset, M = 2337), critical (critical word onset + 1500-ms window after critical word onset), and end of trial (from end of the critical time window to end of trial).*

The model for the peak pupil size was also optimal with the inclusion of response ($\beta = 0.008$, $SE = 0.002$, $t = 3.917$). Including the effect improved model fit, $\chi^2(1) = 13.953$, $p < .001$. Peak pupil size was larger for indirect requests ($M = 0.067$, $SE = 0.002$) than for statements ($M = 0.065$, $SE = 0.001$). In sum, these results indicated that response type, reflecting whether participants thought they heard an indirect request or a statement, predicted mean and peak pupil size, regardless of the stimulus conditions. Mean and peak pupil size were larger for indirect request than for statement responses.

Discussion

Based on the hypothesis that understanding indirect utterances requires additional inferences by the listener, we predicted a larger mean pupil size for indirect requests than for control items (Holtgraves, 1994, 2008; Searle, 1975, 1979; Van Ackeren et al., 2012). In the analyses of trials with correct responses, this prediction was confirmed by an interaction between utterance and picture for the mean and peak pupil size in the 1.5-s time window following the critical word. In other words, the unique combination of action pictures with action utterances in the IR condition resulted in a larger mean and peak pupil size. The hypothesis that this increase in pupil size was related to additional inferences leading to the decision that a request was intended is supported by the second set of analyses, where the data set was split depending on the response type rather than the condition: Larger mean and peak pupil sizes were observed for utterances classified as indirect requests than for statements. Thus, pupil size appears to capture the effort leading to the decision that a request was made.

An interesting finding of the current experiment is the discrepancy between the pupil data and the accuracy scores. For the accuracy scores, we only observed an effect of picture (action versus no-action), but no interaction with utterance. Participants were less accurate on trials containing an action picture, regardless of the type of utterance that these pictures were combined with. This comprehension accuracy pattern is plausible since the action-pictures were designed to be more ambiguous. The trials on which these pictures appeared should allow for an IR interpretation or a statement interpretation. In contrast, the no-action pictures did not have to contain this ambiguity as they were designed to only accompany a statement. Consequently, participants were less accurate for the action pictures than for the no-action pictures. However, this difference in response accuracy due to picture type was not reflected in the pupil dilation. Rather, larger mean and peak pupil size was observed for the action pictures in combination with an action utterance (AP/AU), not for the action picture combined with a no-action utterance (AP/NU).

Most probably, the accuracy scores reflected uncertainty in response selection, whereas pupil size captured the effect of processing an indirect request online. Comprehending and classifying a scene–sentence combination as an indirect request required online inferences that entailed additional processing effort, reflected in the increase in pupil diameter.

This hypothesis—that the increase in pupil size reflects processing effort related to interpreting an utterance as an indirect request—allows for a prediction for the pupillometry data of the entire group of participants, including those who did not pass the threshold of 70% correct responses for each condition. The prediction is that regardless of being correct or not, all IR responses should be associated with an increase in pupil size. The results of the response-split analyses support this hypothesis. Regardless of whether or not the response was correct, sentences classified as indirect requests were associated with a larger mean and peak pupil diameter than were statements, suggesting that only when participants made an inference did pupil diameter increase.

A possible confound for this interpretation is that in Experiment 1, participants always pressed the right button, labelled "yes", to indicate that they heard an indirect request and the left button, labelled "no", to indicate that they heard a statement. It is conceivable that the observed differences in pupil size for the response-split analysis were not related to differences in the processing of the utterances, but rather to making positive or negative decisions or to choice of a response hand. This mapping of IR responses to the dominant hand can also explain the shorter reaction times for indirect requests observed in the behavioural data (see Van Ackeren et al., 2012, for a similar explanation). To rule out the possibility that affirmative responses with the right hand could explain the differences in pupil diameter observed in the response-split analysis, a control experiment was conducted where responses were reversed such that a right-hand "yes" response indicated a statement and a left-hand "no" response an indirect request.

EXPERIMENT 2

Method

Participants

Twelve native speakers of Dutch participated in the study (1 male, mean age = 21.4 years, range = 18–25 years). All participants were right-handed, had normal hearing, normal or corrected-to-normal vision, and no history of language disorders. Informed consent was obtained from all participants. Ethical approval was granted by the ethics board of the Social Sciences Faculty of Radboud University. Participants were paid for taking part in the experiment.

Materials and design

Materials and design were the same as those for Experiment 1.

Apparatus and procedure

The same equipment and procedure were used as those in Experiment 1 except that participants now saw the sentence "The person made a statement" following each item. They were asked to push the right button (dominant hand) if they thought this was true ("yes" responses) and the left button if they thought this was not true ("no" responses). The experimenter indicated during

Table 2. *Experiment 2: Percentage of IR responses ("no" to the statement "The person made a statement") per condition*

	Indirect request (AP/AU)		*Picture control (AP/NU)*		*Utterance control (NP/AU)*		*Picture-utterance control (NP/NU)*	
Condition	*%*	*SE*	*%*	*SE*	*%*	*SE*	*%*	*SE*
IR responses	76.11	5.41	29.17	5.51	3.00	3.01	11.94	2.86

the instruction that sometimes the speaker "means something more" with his statement—for example, he might be asking the listener, in an indirect way, to perform an action. Thus, participants were asked to respond "no" if they thought this was the case.

Analyses

Comprehension accuracy scores were analysed using a one-sampled t test to confirm that participants were able to discriminate between statements and indirect requests. The analysis of the reaction times was the same as that in Experiment 1.

Mean pupil size and peak pupil size were analysed using linear mixed-effects models in R (Baayen et al., 2008). All models contained the same random structure as that in Experiment 1. For this control experiment, we only examined whether or not pupil size was related to the participants' response type ["yes" (statement) versus "no" (indirect request)]. To assess the effect of response (statement versus indirect request), models were compared via a backwards elimination, using a likelihood ratio test. The predictor was mean-centred.

Results and discussion

Behavioural results

Requests were identified more often in the IR condition than in the control conditions (see Table 2). A one-sampled t test (test value = 0.5) confirmed that participants were able to identify the indirect requests correctly in the AP/AU condition, $t(11) = 4.83$, $p < .01$, ($M = 76.11\%$, $SE = 5.41\%$).

Comprehension accuracy scores were slightly higher than those in Experiment 1, especially for the control conditions. This might be due to the fact that in the current experiment participants were not explicitly asked to identify indirect requests, although they knew that indirect requests were present in the experiment. Possibly, participants tried less hard to identify sentences as indirect requests, resulting in lower false-alarm rates. The model for the reaction times revealed no reliable effects and/or interactions (all $t < 1.4$).

Pupillometry results: Response-split

Visual inspection of Figure 5 shows that mean pupil diameter was larger for "no" responses (indirect requests) than "yes" responses (statements). This was confirmed by the analysis. The optimal model for the mean pupil size ($n = 12$) included the fixed effect of Response (statement versus indirect request), $\beta = -0.010$, $SE = 0.005$, $t = -2.097$. "No" responses (indirect requests) resulted in a larger mean pupil diameter than "yes" responses (indirect request: $M = 0.049$, $SE = 0.004$; statement: $M = 0.044$, $SE = 0.002$). Including the effect improved the fit of the model, $\chi^2(1) = 4.252$, $p < .05$.

The model for the peak pupil size was also optimal with the inclusion of response ($\beta = -0.012$, $SE = 0.005$, $t = -2.308$). Peak pupil size was larger for indirect requests ($M = 0.079$, $SE = 0.004$) than for statements ($M = 0.072$, $SE = 0.003$). The predictor improved model fit, $\chi^2(1) = 5.012$, $p < .05$.

In sum, as in Experiment 1, pupil size was larger when participants categorized the utterances as indirect requests than when they categorized them as statements. This was true even though IR responses were now left-hand "no" responses. Together, the two response-split analyses indicated that the differences in pupil size between the IR

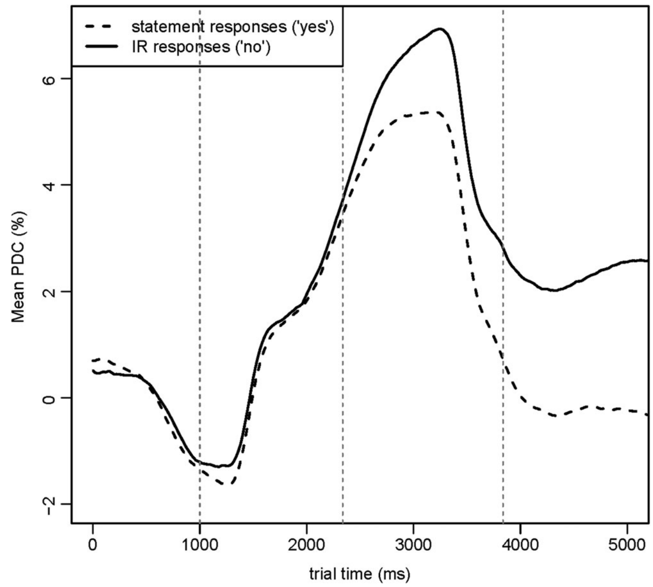

Figure 5. *Experiment 2: Average percentage of pupil diameter change (PDC) as a function of response (statement versus indirect request, IR), n = 12. Trial time (ms) is represented on the* x*-axis and pupil diameter change (%) on the* y*-axis. The vertical lines represent the different time windows: baseline (0 to 1000 ms), audio (1000 ms to critical word onset, M = 2337), critical (critical word onset + 1500-ms window after critical word onset), and end of trial (from end of the critical time window to end of trial).*

condition and the control conditions can be related to the processing of the picture–utterance combinations rather than the choice of a response hand or the selection of an affirmative or negative response.

GENERAL DISCUSSION

Previous research has demonstrated that pupil diameter is sensitive to increases in processing demands as a consequence of, for example, higher memory load (e.g., Piquado et al., 2010), syntactic anomalies (Schluroff, 1982), sentence complexity (Just & Carpenter, 1993), syntactic ambiguity (e.g., Engelhardt et al., 2010), and lexical ambiguity (Ben-Nun, 1986). To our knowledge, the present study is the first to test whether pupil diameter is also sensitive to pragmatic factors, specifically the processing of indirect requests versus direct statements. We presented participants with combinations of sentences and pictures, chosen such that in one out of four combinations the sentence could be interpreted as an indirect request, whereas in the remaining combinations the sentences were mere statements. In Experiment 1, participants were asked to make a yes-response with their right hand when they thought they heard an indirect request and a no-response with their left hand for statements. In Experiment 2 the response choices were reversed, such that right-hand affirmative responses were to be given to statements.

In a time window of 1.5 s following the critical word ("hot" or "nice" in "It is hot/nice here"), we observed a larger mean pupil diameter and a larger peak pupil size for indirect requests than for statements. As explained above, this was true when we compared the pupil size in the experimental conditions of Experiment 1 (indirect request versus control) and also, in both experiments, when we compared the pupil size for indirect request and statement responses regardless of the experimental condition. Thus, we observed an increase in pupil size whenever participants inferred that the utterance was an indirect request, which demonstrates the sensitivity of pupil size to our pragmatic manipulation. Measuring pupil diameter allowed us to observe a unique pattern for indirect requests compared to control statements. From a methodological point of view, this is an encouraging result, as it demonstrates that pupillometry can be used to study the processing of the pragmatic implications of utterances.

Evidently, pupillometry, like any other technique, has certain limitations. The most obvious ones are that participants have to wear eye-tracking equipment and that the visual environment must be carefully controlled in order to minimize changes in pupil sizes due to variations in luminosity (e.g., Janisse, 1977). In addition, the temporal resolution of pupil size measures may be seen as relatively poor, in comparison to, for instance, electroencephalography (EEG). Pupillary responses are relatively slow. In the present study the peak pupil size was reached on average 1000 ms after the onset of the critical word, and it took more than 3000 ms for the pupil diameter to return to baseline. However, although the pupil reaction was quite slow, the temporal resolution of the measure was sufficient for the purposes of the current study. Analyses of pupil diameter in the predefined critical window of 1.5 s after critical word onset allowed us to observe reliable differences as a result of our manipulation.

Our second aim was to investigate the cognitive effort involved in IR comprehension. In the current experimental setting, we observed an increase in mean and peak pupil diameter, reflecting an increase in processing demands for indirect requests as compared to control statements. This supports the view that identifying (and presumably understanding) nonconventional indirect requests is not an automatic process but requires processing effort beyond that needed to process mere statements. This conclusion is in line with behavioural studies on nonconventional indirect requests (e.g., Gibbs, 1981; Holtgraves, 1994) and with an event-related potential (ERP) study on this type of indirect request. In this study, Coulson and Lovett (2010) observed transient processing costs for indirect requests in the form of a larger late positivity component (LPC) for indirect requests than for literal statements. Based on our own findings and the findings by Van Ackeren et al. (2012), which were obtained with a similar stimulus set to that used in the current experiment, we propose that the differences in pupil diameter observed for indirect requests compared to statements in our study reflect the cognitive effort involved in inference processes. Van Ackeren et al. (2012) observed increased activation in ToM areas (mPFC and left TPJ) for indirect requests compared to statements and proposed that these regions were important for making inferences about the communicative intents of the speaker. We suggest that the increases in pupil diameter in our data reflect this inference process as well. This interpretation is consistent with theories of pragmatic processing that postulate that drawing pragmatic inferences requires time and effort (e.g., Sperber & Wilson, 1995). However, it should be noted that in the paradigm used here, the requests were nonconventional (Holtgraves, 2002), and they were difficult to identify, as evidenced by the relatively low comprehension accuracy scores. Also, in natural conversations, interlocutors experience a shared context and have access to each other's speech, gesture, and facial expressions, which has been shown to influence language comprehension (e.g., Özyürek, Willems, Kita, & Hagoort, 2007; Van Berkum, Van den Brink, Tesink, Kos, & Hagoort, 2008). Moreover, Holtgraves (1994) demonstrated that information about the status of the speaker influenced IR comprehension. Thus, in everyday contexts, identifying indirect requests may be easier or harder than in our laboratory setting, depending on the availability of more or less (nonlinguistic) information. Finally, in the present study, participants were asked to provide explicit judgements concerning the implied meaning of the utterances. This is not the case in most everyday contexts and may have increased the processing load. Consequently, on the basis of the present data we cannot make claims about the processing costs incurred during IR processing in other contexts. However, we can say that in some situations identifying and understanding indirect requests entails processing effort beyond the effort needed to understand mere statements, and that this additional effort is reflected in the listeners' pupil size. An important direction for future research is to study the determinants of processing costs for different types of indirect utterances in different contexts. The current study has demonstrated that pupillometry is a useful tool in this endeavour.

REFERENCES

Attar, N., Schneps, M., & Pomplun, M. (2013). Pupil size as a measure of working memory load during a complex visual search task. *Journal of Vision*, *13*(9), 160–160.

Austin, J. L. (1962). *How to do things with words.* Oxford: Oxford University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

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.

Bašnáková, J., Weber, K., Petersson, K. M., Van Berkum, J., & Hagoort, P. (2014). Beyond the language given: The neural correlates of inferring speaker meaning. *Cerebral Cortex*, *24*(10), 2572–2578.

Bates, D. Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1–7.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*, 276–292.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. *Handbook of Psychophysiology*, *2*, 142–162.

Ben-Nun, Y. (1986). The use of pupillometry in the study of on-line verbal processing: Evidence for depths of processing. *Brain and Language*, *28*, 1–11.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge: Cambridge University Press.

Clark, H. H. (1979). Responding to indirect speech acts. *Cognitive psychology*, *11*(4), 430–477.

Coulson, S., & Lovett, C. (2010). Comprehension of non-conventional indirect requests: An event related brain potential study. *Psychology*, *11*, 430–477.

Einhäuser, W., Stout, J., Koch, C., & Carter, O. (2008). Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(5), 1704–9.

Engelhardt, P. E., Ferreira, F., & Patsenko, E. G. (2010). Pupillometry reveals processing load during spoken language comprehension. *Quarterly Journal of Experimental Psychology*, *63* (4), 639–645.

Gallagher, H. L., & Frith, C. D. (2003). Functional imaging of 'theory of mind'. *Trends in Cognitive Sciences*, *7*(2), 77–83.

Gerakaki, S., Sjerps, M. J., & Meyer, A. S. (in press). *Planning speech while listening to speech affects memory of heard words*.

Gibbs, R. W. (1981). Your wish is my command: Convention and context in interpreting indirect requests. *Journal of Verbal Learning and Verbal Behavior*, *20*(4), 431–444.

Gibbs, R. W. (1983). Do people always process the literal meanings of indirect requests? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*(3), 524–533.

Gibbs Jr, R. W. (1986). What makes some indirect speech acts conventional? *Journal of Memory and Language*, *25*(2), 181–196.

Hess, E. H., & Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science*, *143*(3611), 1190–1192.

Holtgraves, T. (1994). Communication in context: Effects of speaker status on the comprehension of indirect requests. *Journal of Experimental Psychology: Learning Memory and Cognition*, *20*(5), 1215–1218.

Holtgraves, T. (2002). *Language as social action.* New Jersey: Lawrence Erlbaum Associates.

Holtgraves, T. (2008). Automatic intention recognition in conversation processing. *Journal of Memory and Language*, *58*(3), 627–645.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446.

Janisse, M. P. (1977). *Pupillometry: The psychology of the pupillary response*. Washington, DC: Hemisphere Publishing Corporation.

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology-revue Canadienne De Psychologie Experimentale*, *47*, 310–339.

Just, M. A., Carpenter, P. A., & Miyake, A. (2003). Neuroindices of cognitive workload: Neuroimaging, pupillometric and event-related potential studies of brain work. *Theoretical Issues in Ergonomics Science*, *4* (1), 56–88.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, MA: MIT Press.

Otero, S. C., Weekes, B. S., & Hutton, S. B. (2011). Pupil size changes during recognition memory. *Psychophysiology*, *48*(10), 1346–53.

Özyürek, A., Willems, R., Kita, S., & Hagoort, P. (2007). On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience*, *19*(4), 605–616.

Piquado, T., Isaacowitz, D., & Wingfield, A. (2010). Pupillometry as a measure of cognitive effort in younger and older adults. *Psychophysiology*, *47*(3), 560–569.

Porter, G., Troscianko, T., & Gilchrist, I. D. (2007). Effort during visual search and counting: Insights from pupillometry. *The Quarterly Journal of Experimental Psychology*, *60*(2), 211–229.

Schluroff, M. (1982). Pupil responses to grammatical complexity of sentences. *Brain and Language*, *17*(1), 133–145.

Schluroff, M., Zimmermann, T. E., Freeman, R. B., Hofmeister, K., Lorscheid, T., & Weber, A. (1986). Pupillary responses to syntactic ambiguity of sentences. *Brain and Language*, *27*(2), 322–344.

Searle, J. R. (1975). Indirect speech acts. *Syntax and Semantics*, *3*, 59–82.

Searle, J. R. (1979). *Expression and meaning: Studies in the theory of speech acts*. Cambridge, UK: Cambridge University Press.

Sirois, S., & Brisson, J. (2014). Pupillometry. *Wiley Interdisciplinary Reviews: Cognitive Science*, *5*(6), 679–692.

Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Cambridge, MA: Harvard University Press.

Sperber, D., & Wilson, D. (1995). *Relevance: Communication and cognition* (2nd ed.). Oxford: UK Blackwell Publishers.

Van Ackeren, M. J., Casasanto, D., Bekkering, H., Hagoort, P., & Rueschemeyer, S. A. (2012). Pragmatics in action: Indirect requests engage theory of mind areas and the cortical motor network. *Journal of Cognitive Neuroscience*, *24*(11), 2237–2247.

Van Berkum, J. J., Van den Brink, D., Tesink, C. M., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591.

Van Rijn, H., Dalenberg, J. R., Borst, J. P., & Sprenger, S. A. (2012). Pupil dilation co-varies with memory strength of individual traces in a delayed response paired-associate task. *PloS One*, *7*(12), 1–8.

Wierda, S. M., van Rijn, H., Taatgen, N. A., & Martens, S. (2012). Pupil dilation deconvolution reveals the dynamics of attention at high temporal resolution. *Proceedings of the National Academy of Sciences*, *109* (22), 8456–8460.

Zellin, M., Pannekamp, A., Toepel, U., & van der Meer, E. (2011). In the eye of the listener: Pupil dilation elucidates discourse processing. *International Journal of Psychophysiology*, *81*(3), 133–141.