Title
Pupillary Indices of Cognitive Workload in (In-)Efficient Reference Production

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors
Tourtouri, Elli
Gotzner, Nicole

Publication Date
2026

Pupillary Indices of Cognitive Workload in (In-)Efficient Reference Production

Elli N. Tourtouri (elli.tourtouri@{uni-osnabrueck.de, mpi.nl}[1, 2] & Nicole Gotzner[1]
[1]Institute of Cognitive Science, Osnabrück University, [2]Max Planck Institute for Psycholinguistics, Nijmegen

Abstract

Speakers' use of redundant information in referential communication has been the subject of much debate, as it seemingly defies Grice's Maxim of Quantity (Grice, 1975). While some research argues for an egocentric view on referential redundancy (i.e., that it results from a concern to ease production effort), other work supports an audience-design view (i.e., that speakers strive for communicative efficiency). This research has mainly leveraged listeners' comprehension as an indirect index of production strategy, while it is yet unclear whether audience-design indeed incurs increased cognitive effort for the speaker. In this study, we used pupillometry to directly assess the cognitive load associated with different referential production strategies, finding evidence for modulated effort based on speakers' engagement in reasoning about the efficiency of their utterances.

Keywords: production of referring expressions; redundant adjectives; communicative efficiency; pupillometry

Introduction

When determining what information to encode in a referring expression, speakers do not always act according to Gricean expectations: While they indeed rarely use less information than necessary, they do frequently say more than what is actually needed to specify a target object among a set of potential referents (i.a. Arts et al. 2011a, Davies & Katsos, 2013; Engelhardt et al., 2006; Rubio-Fernández, 2019; Tourtouri et al., 2019). Why speakers may seemingly behave so irrationally is the matter of much debate, especially when their referentially redundant utterances may jeopardize listeners' comprehension.

Two main views have been put forward. According to the **egocentric view** speakers' use of redundancy may be the outcome of a general strategy to ease the effort associated with referential production. In other words, in referential communication speakers may commonly start planning an utterance before they have exhaustively scanned the visual scene for competitor objects, and therefore before they have established which contrasts are relevant and what information is necessary to fully specify the target referent (e.g., Belke & Meyer, 2002). In doing so, speakers rely on describing aspects of the target referent that broadly speaking seem to be relevant, e.g., color in a multi-color scene.

The **audience-design view**, on the other hand, holds that speakers' use of unnecessary information about a target referent may in fact aim at facilitating their addressees' referential processing (i.e., the real-time comprehension and mapping of the referring expression onto the visual scene). As far as we can tell, two accounts fit under this view. According to one, speakers overwhelmingly use **color information redundantly** (Jara-Ettinger & Rubio-Fernández, 2021; Rubio-Fernández, 2016; Gilles, Mollica, & Rubio-Fernández, 2026). This propensity is presumably associated with the visual salience of color as a feature in referential communication tasks (Koolen et al., 2013; but it might also be linked to other factors, e.g., the omnipresence of color modifiers in such tasks since childhood; Ramscar et al. 2010). In this sense, while redundant color information may facilitate the listener's task, it is at the same time easy for the speaker to visually identify and encode it. Based on a stronger version of the audience-design view, speakers may utilize redundant modifiers in contexts where these help exclude a larger number of competitor objects. In doing so, speakers reason about the addressees' visual search for the target referent, and use **redundancy rationally** as a means to guide this search efficiently (Degen et al., 2020; Fukumura, 2018; Tourtouri et al., 2019). It must be noted, however, that the evidence supporting this rational redundancy account also attest to the special status of color.

It is well possible that both the egocentric and audience-design (in either color or rational redundancy version) views hold, and that speakers select production strategies based on the settings or the goal of the interaction (Arts et al., 2011b; Sikos et al. 2021), or based on their own cognitive resources and prior experience (Vogels et al., 2020). However, it is yet unclear whether audience design incurs increased cognitive effort for the speaker, especially when it involves reasoning about how to manage the listener's task efficiently.

In addition, previous research has mainly used an indirect index (if any) of production strategy, leveraging addressees' comprehension: If referential redundancy was shown to impede comprehension, this was taken as indication that the use of redundant information must be the result of an egocentric production strategy—and vice versa, facilitated comprehension was taken as a sign of audience design (cf. Engelhardt et al., 2006; Arts et al., 2011a).

In this study, we used pupillometry to directly assess speakers' cognitive workload during speech planning (see Sauppe, 2017). We aimed to understand whether audience-design (in both versions) requires more cognitive effort as compared to egocentric planning, and whether over and above that there are differences between the two audience-design strategies. In order to elicit utterances corresponding to the different production strategies, we used the materials from a study that reported a tripartite split based on speakers' use of redundant color and pattern adjectives in different visual contexts (Tourtouri et al., 2019, Exp.2). We, therefore, aimed to: (a) identify speaker groups based on their production strategy; and (b) assess whether the observed strategies are associated with distinct cognitive workload profiles during speech planning as indexed by the speakers' pupillary responses.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

Experiment

In a 2x3 experimental design, the factors **Necessary adjective** and **Efficient adjective** were fully crossed. "Necessary adjective" manipulated the target referent's distinguishing property in two levels: color and pattern. "Efficient adjective" was operationalized as the adjective that restricted the set of potential target referents (referential entropy) early on in the expression. That is, in each visual display, either color or pattern was necessary to fully specify the target referent, and (regardless of which adjective was necessary) color or pattern or neither of the two restricted the referential set efficiently. This yielded six experimental conditions: a. Color-Necessary/Color-Efficient, b. Color-Necessary/Pattern-Efficient, c. Color-Necessary/Neither-Efficient, d. Pattern-Necessary/Pattern-Efficient, e. Pattern-Necessary/Color-Efficient, and f. Pattern-Necessary/Neither-Efficient.

We generally expected to replicate the earlier findings (cf. Tourtouri et al., 2019) that speakers' use of redundant adjectives follows distinct production strategies: Some speakers were expected to use both color and pattern adjectives redundantly regardless of how efficient they were in a given context (egocentric redundancy). Another group of speakers was expected to use only color redundantly, regardless of its efficiency (color redundancy). A third group of speakers was expected to use color adjectives redundantly when these adjectives were efficient in the given context (i.e., they helped restrict the set of competitor referents early on; rational redundancy).

Moreover, we hypothesized that, due to the added difficulty of theorizing about another's perspective, audience-design production strategies (color redundancy and rational redundancy) should incur overall higher cognitive workload compared to egocentric production. Specifically, we expected increased effort (mean pupil size changes) for both color redundancy and rational redundancy strategies compared to egocentric redundancy. Because a rational redundancy strategy is linked with reasoning about the utility of a redundant (color) adjective in a given context, we additionally expected increased effort associated with this strategy compared to the color redundancy strategy.

Method

Participants Forty-nine pairs of native speakers of German (mean age=22.2, 76 female) were recruited through the participant database of the Department of Psychology and of the Institute of Cognitive Science at Osnabrück University. Participants had normal or corrected-to-normal vision and no problems with color perception. They were compensated with course credit for their participation.

Materials The stimuli from the production experiment of Tourtouri et al. (2019) were used as materials in this study. The stimuli featured 18 everyday objects in three colors (blue, green, red) and three patterns (checkered, dotted, striped). Experimental items comprised a set of six visual displays, each portraying a different arrangement of the same objects, yielding the six experimental conditions (see Fig.1). Conditions differed in the feature needed to be mentioned for unique identification of the target referent (Necessary adjective; levels: color, pattern), and in which adjective was most efficient for resolving reference, i.e., for reducing the number referential competitors (Efficient adjective; levels: color, pattern, equal). Thus, in experimental items the mention of one adjective was always necessary, and a second adjective would be redundant. In fillers, the same object

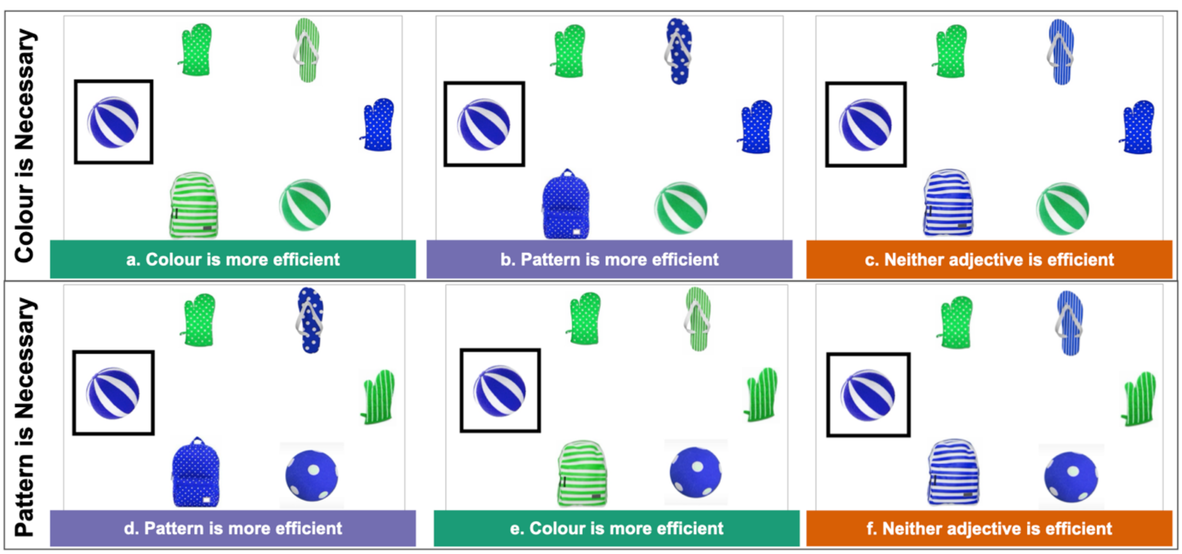

Figure 1: Sample Speaker stimuli for an experimental item. The black frame indicates the target referent. The top panel presents conditions where a color adjective (*blue*) is necessary to specify the target object. The bottom panel presents conditions where a pattern adjective (*striped*) is necessary to specify the target object. In conditions a and e (green labels), color is more efficient than pattern in restricting the referential domain, while in conditions b and d (purple labels) the pattern is more efficient than color. In conditions c and f (orange labels), neither color nor pattern are efficient, as both adjectives exclude the same number of competitor objects.

referents were placed in either six-object or four-object arrangements, and the target referent required either two or no adjectives in order to be fully specified.

Thirty-six experimental items for the main experiment were split across 6 lists using a Latin-Square design, and were intermixed with fillers. Each list was pseudo-randomized such that trials from the same condition were not consecutive, and that experimental trials interleaved with at least one filler. A total of 576 visual displays were used. Half of these were flipped vertically only on the Listener display.

Perceptibility norming Previous research has shown that perceptibility (i.e., how easy it is to characterize an object as carrying a specific property) can influence speakers' choice to include redundant adjectives in their referring expressions (i.a. Koolen et al., 2013; Viethen et al., 2012). In three experiments, Kursat and Degen (2021) found evidence that the asymmetry in the redundant use of color and material adjectives is related to the perceptibility differences between color and material features—i.e., that the visual salience of color may drive the inclusion of redundant color adjectives in referential expressions.

To account for the potential influence of perceptibility in the use of redundant color and pattern modifiers in the present experiment, we conducted a norming study measuring the perceptibility of the six distinguishing attributes used (blue, green, red, checkered, dotted, striped) for each of the object referents. We recruited 180 native speakers of German with normal or corrected-to-normal vision and no problems with color perception on Prolific (https://www.prolific.co; Palan & Schitter, 2018). The data from one participant was not saved due to a technical issue. The remaining 179 participants (average age=33, 50 female) saw a set of 54 visual displays each, portraying four objects of the same type that differed in color and pattern (e.g., four balls, two of which were dotted, one checkered and one striped), and a matching or mismatching description underneath (e.g., "3 striped balls"). The participants' task was to click a "yes"/"no" button on the screen indicating whether the description was correct or incorrect. The datasets from 10 participants who failed to respond correctly to four "catch" trials were not further considered. Responses were analyzed for accuracy and latency in R (R version 4.4.1, R Core Team, 2024).

Overall, results showed that descriptions containing a pattern adjective triggered less accurate ($p<.05$) and slower ($p<.05$) responses compared to descriptions containing a color adjective. Specifically, compared to descriptions with "blue", those with "red" and "green" were found to be equally accurate or fast ($p>.05$), while those with "checkered" and "dotted" were both less accurate and slower. Descriptions with "striped" were also responded to slower than those with "blue" ($p<.05$), but responses did not differ in accuracy ($p>.05$). These results indicate that, while color attributes are indeed perceptually salient, other attributes (such as striped) may be equally easy to perceive. We, therefore, computed a **perceptibility score** for each object-adjective combination (e.g., "dotted ball") based on the mean response accuracy it received in the norming study. This perceptibility score was included as predictor in the statistical models of the main experiment (see below).

Procedure After giving informed consent, participants were assigned to the roles of "Information-Seeker" (Speaker) and "Information-Giver" (Listener), and were seated in adjacent sound-isolating booths, at a 60 cm distance from a 24-inch Dell monitor. Both monitors were connected to a Dell Precision 3581 laptop running Windows 11, placed on the experimenter station outside the booths. Two sets of headphones (AKG K-92) and two microphones (Behringer B5) connected through a sound mixer (Behringer Xenyx 1204USB) established an audio link for participants to communicate. A Tobii Pro Fusion 120 Hz eye-tracker (Tobii AB, Danderyd, Sweden) was attached to the bottom of the Speaker display. Stimuli were presented using the E-Prime 3.0 software (Psychology Software Tools, Inc., Pittsburgh, PA, USA). The Listener had a keyboard placed in front of them, through which they carried out the task and controlled the pace of the experiment. Luminance conditions were held constant across participants.

At the start of each trial, a small cross was presented in the middle of the Speaker's screen, which disappeared as soon as it was fixated. During this time the Listener's screen was blank. A visual display of the object referents was then presented on both participants' screens. In half of the trials, the Listener's display was a mirrored version of that shown to the Speaker. Their joint task was to find out whether a target referent was placed on the same side of both their screens. Specifically, 3 seconds after the presentation of the visual display, the target referent was marked on the Speaker's screen only (see black frame in Fig.1). At the same time, a sound was played over the headphones, indicating to the Listener that the Speaker was made aware of the target object. The Speaker's task was therefore to ask a question about the horizontal position of the target object (e.g., "Is the blue ball on the left?"). It is important to note that Speakers were instructed to produce full grammatical sentences, and avoid using fragments (e.g., "blue ball left"). Listener had to reply by pressing a "yes"/"no" key on the keyboard in front of them (the keys were marked with the corresponding labels). Any time the Listener made a key press, auditory feedback was provided over the headphones: a bell sound for correct responses, and a buzzer sound for incorrect responses. The feedback was also heard by the Speaker and the experimenter.

An experimental session started with participants reading the instructions on their screens. Next, the experimenter explained the instructions orally, making sure that the participants understood their respective tasks and the cooperative nature of the experiment. Before the start of the experiment, a preview of the individual object pictures used in the visual stimuli was shown, and Speakers were asked to name the object types out loud, ensuring name agreement between the participants already on the first trial. Participants then went through a short practice phase, during which the experimenter remained in the Speaker booth and provided feedback as necessary. After a successful 9-point calibration and validation sequence, the experimenter exited the booth, and the experiment started. The experimenter could monitor the interaction and the pace of the experiment over a set of headphones connected to the sound mixer. Speakers' eye-tracking data as well as their overt verbal responses (questions) were recorded.

The experiment consisted of four blocks, in-between which participants could take a break. After the end of the experiment, participants were debriefed. The approximate duration of a full session was 45 minutes.

Data coding Speakers' utterances were transcribed and annotated for specificity of the referring expression by an independent annotator. Minimally-specified expressions were coded as "0", over-specified expressions were coded as "1", and under-specified expressions (cases where the adjective necessary for fully specifying the target referent was not encoded) were coded as "-1". Trials containing self-repairs (1.5%), syntactic revisions (1.7%), or under-specified referring expressions (2.7%) were excluded from analyses. In order to make sure that the use of a redundant adjective on a given trial was not due to priming from an immediately preceding filler requiring two adjectives, we also excluded from analyses trials in which the same number of adjectives were produced in the exact same order as in the preceding trial (11.8%).[1]

Pupil data processing Pupil data was processed using the PupillometryR package (version 0.0.5; Forbes, 2020). Pupil size data from both eyes was obtained and regressed against each other for smoothing (Sirois & Brisson, 2014). The mean pupil size was calculated, and passed through an 11-sample Hanning filter. Linear interpolation was performed to account for missing data (blinks), and the data was downsampled to 50 Hz to reduce residual autocorrelation (van Rij et al., 2019). The mean pupil size per participant and trial within the first 500 ms after the target object was revealed served as baseline, and was subtracted from all timepoints (Mathôt et al., 2018).

Analysis We performed two types of analysis: (a) generalized linear mixed (GLM) models for the rate of Speakers' use of redundant adjectives, and (b) Growth Curve Analysis (GCA; Mirman, 2014) for the time course of the pupil response per Speaker group after picture onset. For redundancy rates, we ran separate GLM models per Speaker group, including as fixed effects the factors "Necessary adjective" (treatment-coded, reference level: "Pattern") and "Efficient adjective" (treatment-coded, reference level: "Color"), and crossed random slopes and intercepts for participants and items. When the maximal model did not converge, the random effect structure was simplified as needed (Barr et al., 2013). For GCA, mixed-effects linear regression models included as fixed factors orthogonal polynomials up to the third order and Speaker group (Helmert-coded) to test the contribution of each time term, speaker group and their interaction on the time course of the pupil response. The random effects structure included random intercepts for participants and random slopes and intercepts for items. The individual contribution of each fixed effect and interaction on model fit was evaluated based on Log-likelihood model comparisons, and their results are reported below (see Table 2). Perceptibility was centered (z-transformed) and included as a control predictor in all models. Perceptibility results are reported when model fit was improved. All analyses were carried out in R (version 4.4.1, R Core Team, 2024; lme4 package, Bates et al., 2015).

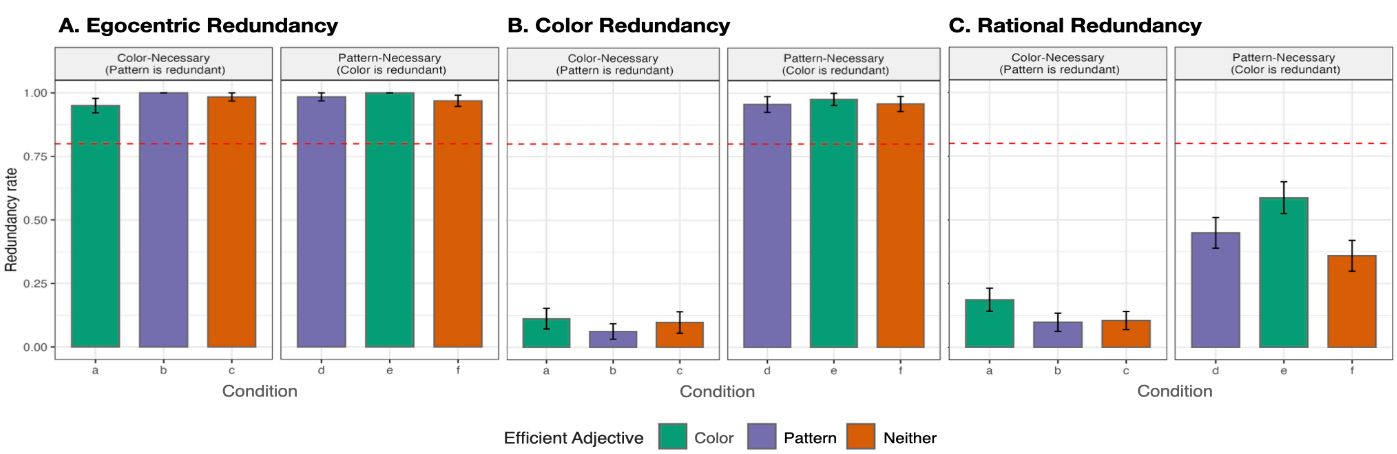

Figure 2: Redundancy rates per group. In all panels, the plot on the right presents redundancy rates in Pattern-Necessary conditions (where color is redundant), and the plot on the left presents redundancy rates in Color-Necessary conditions (where pattern is redundant). Adjective efficiency is color-coded: Green bars demonstrate conditions where color is efficient (a and e), purple bars conditions where pattern is efficient (b and d), and orange bars conditions where neither adjective is efficient (c and f). The red line represents the 80% redundancy rate threshold. Error bars represent *SE*.

[1] Note that this step was taken after identifying participants who produced a redundant adjective in more than 80% of the trials (Egocentric redundancy group; see Results).

Table 1: Production of redundant adjectives for speakers in the Colour redundancy and Rational redundancy groups

	Color redundancy (N=12)				**Rational Redundancy** (N=15)			
	β	*SE*	*z*	*p*	β	*SE*	*z*	*p*
Intercept	4.28	1.16	3.67	<0.001 ***	1.65	5.55	0.29	0.77
Necessary adjective								
Color	-6.98	1.48	-4.70	<0.001 ***	-2.19	0.45	-4.90	<0.001 ***
Efficient adjective								
Pattern	-0.64	1.28	-0.50	0.62	-0.73	0.39	-1.89	0.058 .
Neither	-0.69	1.28	-0.54	0.58	-1.12	0.40	-2.77	<0.01 **
Necessary:Efficient								
Color:Pattern	0.09	1.46	0.06	0.95	-0.07	0.64	-0.11	0.91
Color:Neither	0.87	1.48	0.58	0.55	0.35	0.64	0.55	0.58
Perceptibility	-	-	-	-	-1.14	5.72	-0.20	0.84

Results

Redundancy rate

Participants whose referring expressions were predominantly minimally-specified (redundancy rate <10%; N=9) were excluded from further analyses. The remaining participants were categorized into three groups based on their redundancy rate (Fig. 2). Participants who produced a redundant adjective in more than 80% of all trials regardless of the condition were categorized in a first group (N=11). In a second group (N=12), we categorized participants who produced a redundant color adjective in more than 80% of all Pattern-Necessary trials regardless of its efficiency. In a third group (N=15), we categorized participants who produced a *redundant color* adjective more frequently when it was more efficient in reducing the number of referential competitors compared to the *necessary pattern* adjective (see Fig.1e).[2]

The model could not estimate condition effects in the first group, as redundancy rates were nearly at ceiling (>95%) for all conditions (see Fig. 2). These high rates in all conditions indicate that the use of redundant adjectives by these speakers was unrelated to contextual factors, but potentially motivated by a concern to ease production processes (**egocentric redundancy**). As shown in Table 1, participants in the second group produced redundant adjectives more frequently overall in the Pattern-Necessary than in the Colour-Necessary conditions. In other words, for these participants the perceptual salience of color alone seemed to affect the use of redundant adjectives (**color redundancy**)—regardless of the distribution of other features in the visual scene. Speakers in the third group seemed to consider the distributional properties of the visual scene in order to encode information that would benefit the listeners' visual search of the target referent (**rational redundancy**; cf. Table 1: decreased use of redundant color adjectives when neither adjective is efficient compared to when color is efficient, and a trend in the same direction for the Pattern-Efficient). In other words, speakers in this group utilized redundancy strategically when the redundant adjective could help restrict the referential space early in the utterance—but only when this information was also perceptually salient (cf. Table 1: decreased use of redundant pattern adjectives when color was necessary).

Pupil response

Figure 3 illustrates the time course of the pupil response per Speaker group. Results from model comparisons are reported in Table 2. All polynomials (linear: β=56.9, *SE*=4.3; quadratic: β=-13.5, *SE*=2.2; cubic: β=-12.17, *SE*=1.9), and Perceptibility (β=.005, *SE*<.001) improved model fit. The contrast between the Egocentric redundancy group and the Color and Rational redundancy groups yielded marginally significant effects on the linear (β=-18.2, *SE*=9.4) and cubic terms (β=8.1, *SE*=4.3), indicating a steeper rise and fall of the pupil responses for the latter two groups. No effect was observed on the quadratic term (β=3.8, *SE*=4.9) for this contrast. The contrast between the Color redundancy and Rational redundancy groups showed no differences on the linear (β=-4.7, *SE*=10.1) and cubic (β=5.4, *SE*=4.7) terms, but

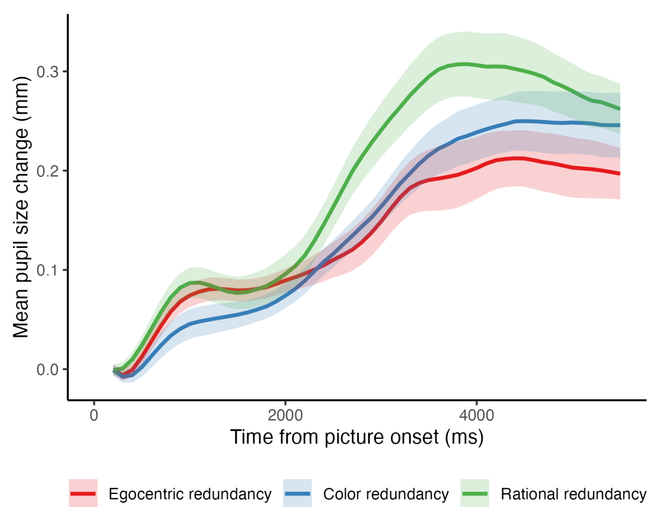

Figure 3. Time course of Speakers' baseline-corrected mean pupil size from the start of the picture onset. The target object was revealed at 3000 ms. The red line corresponds to the Egocentric redundancy group, the blue line to the Color redundancy group, and the green line to the Rational redundancy group. Ribbons around the three lines represent *SE*.

[2] The data from two participants who produced redundant adjectives in a few of the Pattern/Neither-Efficient trials were not included in any group.

Table 2: Log-likelihood model comparisons for GCA

	$\chi 2$	*df*	*p*
Group	4.9	2	.08.
$Time^1$	64.1	1	<.001 ***
$Time^2$	24.7	1	<.001 ***
$Time^3$	25.8	1	<.001 ***
Perceptibility	349.1	1	<.001 ***
Group:$Time^1$	3.9	2	.14
Group:$Time^2$	6.8	2	.03 *
Group:$Time^3$	4.8	2	.09.

Note. $Time^1$ = linear polynomial; $Time^2$ = quadratic polynomial; $Time^3$ = cubic polynomial

a significant difference on the quadratic term (β=-4.1, SE=1.5), indicating that the curvature peak was higher for the Rational redundancy compared to Color Redundancy group.

General Discussion

In this study, we examined whether different production strategies are associated with distinct cognitive workload profiles, depending on the extent to which speakers reason about the addressee's comprehension. Specifically, we aimed to assess whether, compared to egocentric speech planning, tailoring utterances to the addressees' needs (audience design) is related to increased cognitive effort for the speaker.

To elicit egocentric and addressee-oriented utterances, we exploited the referential communication task, where participants have to describe a target object within a set of referents to a task-partner. A recurring finding in research using this task is that speakers frequently encode redundant information about the target object. While the rate at which speakers produce redundant information may vary, this typically consists of redundant color adjectives. Earlier work has argued both that this prolixity is the result of egocentric production processes (utterance planning begins based on simple heuristics and before speakers have identified the distinguishing property of the target referent), and that it is in fact a demonstration of audience-design (redundant adjectives—especially color—may facilitate the addressees' visual search of the target referent). Two flavors of the view that redundancy serves audience-design purposes can be identified: i.e., that redundant color adjectives facilitate referential processing in general, or when they are used rationally so as to efficiently restrict the referential set size.

While their pupil responses were recorded, we presented participants with displays of objects differing in color and pattern. Manipulating both which property was necessary and which property was more efficient in managing visual search for the target object, we elicited responses that allowed us to categorize participants into three groups: (a) an Egocentric redundancy group, where speakers produced both color and pattern adjectives redundantly regardless of which one was necessary and which one was more efficient; (b) a Color redundancy group, where speakers produced redundant color adjectives when pattern was necessary, and regardless of how efficient color was in reducing the amount of competitors; (c) a Rational redundancy group, in which speakers used color adjectives redundantly, but predominantly when they were more efficient than the necessary pattern adjective in restricting the referential set size.

Two comparisons were of interest regarding how the speakers' cognitive workload (as measured by the pupil response) evolved over time: (a) between the Egocentric redundancy group and both the Color and Rational redundancy (audience-design) groups, indicating whether there is an overall difference in the cognitive effort expended by egocentric vs. addressee-oriented speakers during referential language planning; (b) within the two audience-design groups, indicating whether the cognitive effort associated with audience design in referential language production is modulated by the extent to which speakers reason about how to efficiently manage the addressees' visual search for the target referent.

The findings indicated that speakers' planning-related cognitive effort was modulated based on how they utilized redundant adjectives in their referring expressions. The two audience-design strategies incurred distinct levels of workload (as indexed by a difference in pupil response on the quadratic term): Workload was higher for speakers who used redundant color adjectives rationally (when they were more efficient than the necessary pattern adjectives) compared to speakers who generally favored the use of color adjectives (regardless of their efficiency). This result supports the view that reasoning about the efficiency of redundant information for the addressees' task (and, thereby, the joint goal) is more costly than a blanket use of redundant color information. It should, however, be noted that due to the visual salience of color, the use of redundant color information does not impede addressees' referential processing (cf. Rubio-Fernández, 2021). In this respect, color redundancy is both beneficial for the listener and not too costly for the speaker.

We found only tentative evidence that speakers who encoded both adjectives redundantly exhibited reduced cognitive effort compared to both audience-design groups. It is possible, however, that the analytical approach, in combination with the low number of observations in the Egocentric redundancy group compared to the two audience-design groups, may have obscured potential differences between the production strategies (especially between egocentric and rational redundancy). If supported by further evidence, this finding would suggest that the production strategy of the Egocentric redundancy group was motivated by a concern to facilitate production-internal processes rather than to assist the addressee, as indeed the unnecessary use of information other than color has been found to hinder listeners' processing (e.g., Tourtouri et al., 2021). Analyses incorporating eye-tracking data could further enrich our understanding of the different production strategies.

In all, in this study we confirmed earlier findings of distinct production strategies in referential communication, and provided first-time evidence of the cognitive workload associated with each, showing that reasoning about utterance efficiency may result in increased processing effort for the speakers.

Acknowledgements

We thank Benjamin Peters and Hardika Potey for assistance with the perceptibility norming, and Mitra Gholami for assistance with data collection and annotation. A preregistration for this study was presented at the XPrag.it 2024 conference in Venice.

References

Arts, A., Maes, A., Noordman L., & Jansen C. (2011a). Overspecification facilitates object identification. *Journal of Pragmatics*, 43(1), 361–374.

Arts, A., Maes, A., Noordman, L., & Jansen, C. (2011b). Overspecification in Written Instruction. *Linguistics*, 49(3), 555–74.

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, 255–278.

Bates, D., Maechler, M., Bolker, B. & Walker S. (2015). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-35.3.

Belke, E., & Meyer, A. S. (2002). Tracking the Time Course of Multidimensional Stimulus Discrimination: Analyses of Viewing Patterns and Processing Times during "Same"-"Different" Decisions. *European Journal of Cognitive Psychology*, *14*(2), 237–266.

Davies, C., & Katsos, N. (2013). Are speakers and listeners 'only moderately Gricean'? An empirical response to Engelhardt et al. (2006). *Journal of Pragmatics*, 49(1): 78106.

Degen, J., Hawkins, R. D., Graf, C., Kreiss, E., Goodman, N. D. (2020). When Redundancy Is Useful: A Bayesian Approach to 'Overinformative' Referring Expressions. *Psychological Review*, 4, 591–621.

Engelhardt, P., Bailey, K. & Ferreira, F. (2006). Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, 54(4), 554–573.

Forbes, S. (2020). PupillometryR: An R package for preparing and analysing pupillometry data. *Journal of Open Source Software*, *5*(50): 2285.

Fukumura, K. (2018). Ordering Adjectives in Referential Communication. *Journal of Memory and Language*, *101,* 37-50.

Giles, M., Rubio-Fernández, P. & Mollica, F. (2026) Search Efficiency Drives Reference Production across Modalities, but Colour Is Special. *Open Mind, 10,* 236–260.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. III, pp. 41–58). New York: Academic.

Jara-Ettinger, J., & Rubio-Fernandez, P. The Social Basis of Referential Communication: Speakers Construct Physical Reference Based on Listeners' Expected Visual Search. *Psychological Review,* 129.

Koolen, R., Goudbeek, M., & Krahmer, E. (2013). The effect of scene variation on the redundant use of color in definite reference. *Cognitive Science*, 37(2), 395–411.

Kursat, L. & Degen, J. (2021). Perceptual Difficulty Differences Predict Asymmetry in Redundant Modification with Color and Material Adjectives. *Proceedings of the Linguistic Society of America*, *6*(1), 676–688.

Mathôt, S., Fabius, J., Van Heusden, E., Van der Stigchel, S. (2018). Safe and sensible preprocessing and baseline correction of pupil-size data. *Behavior Research Methods. 50*(1)*,* 94-106

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. Chapman and Hall/CRC.

Palan, S., & Schitter, C. (2018). Prolific.ac – A subject pool for online experiments. *Journal of Behavioral and Experimental Finance*, 17, 22–27.

Ramscar, M., Yarlett, D., Dye, M., Denny, K. & Thorpe, K. (2010). The Effects of Feature-Label-Order and Their Implications for Symbolic Learning. *Cognitive Science*, 34, 909-957.

Rubio-Fernández P. (2016). How Redundant Are Redundant Color Adjectives? An Efficiency-Based Analysis of Color Overspecification. *Frontiers in Psychology, 7*, 153

Rubio-Fernandez, P. (2019) Overinformative Speakers Are Cooperative: Revisiting the Gricean Maxim of Quantity. *Cognitive Science, 43*(11): e12797.

Rubio-Fernandez, P. (2021). Color Discriminability Makes Over-Specification Efficient: Theoretical Analysis and Empirical Evidence." *Humanities and Social Sciences Communications*, *8*:147.

Sauppe, S. (2017) "Symmetrical and Asymmetrical Voice Systems and Processing Load: Pupillometric Evidence from Sentence Production in Tagalog and German." Language, 93(2), 288–313.

Sikos, L., Venhuizen, N. J., Drenhaus, H., & Crocker, M. W. (2021). Reevaluating Pragmatic Reasoning in Language Games. *PLoS ONE* 16(3): e0248388.

Sirois, S. & Brisson, J. (2014). Pupillometry. *WIREs Cognitive Science, 5*(6), 679–692.

Tourtouri, E. N., Delogu, F. Sikos, L. & Crocker, M. W. (2019). Rational over-specification in visually-situated comprehension and production. *Journal of Cultural Cognitive Science*, 3(2), 175–202.

Tourtouri, E.N., Delogu, F., & Crocker, M.W. (2021). Rational Redundancy in Referring Expressions: Evidence from Event-related Potentials. *Cognitive Science,* 45:e13071

Van Rij, J., Hendriks, P., Van Rijn, H. Baayen, R. H. & Wood S. N. (2019) Analyzing the Time Course of Pupillometric Data. *Trends in Hearing, 23*, 1-22.

Viethen, J., Goudbeek, M., & Krahmer, E. (2012). The Impact of Colour Difference and Colour Codability on Reference Production. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, 1084-1089.

Vogels, J., Howcroft, D.M., Tourtouri, E., & Demberg, V. (2020). How Speakers Adapt Object Descriptions to Listeners under Load. *Language, Cognition and Neuroscience*, 35(1), 78–92.