Replicating five pupillometry studies of Eckhard Hess

de Winter, J. C.F.; Petermeijer, S. M.; Kooijman, L.; Dodou, D.

DOI
10.1016/j.ijpsycho.2021.03.003

Publication date
2021

Document Version
Final published version

Published in
International Journal of Psychophysiology

International Journal of Psychophysiology 165 (2021) 145–205

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Replicating five pupillometry studies of Eckhard Hess

J.C.F. de Winter *, S.M. Petermeijer, L. Kooijman, D. Dodou

Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, the Netherlands

ARTICLE INFO

Keywords:
Replication study
Pupil dilation
Interest
Arousal
Gender differences
Mental demands

ABSTRACT

Several papers by Eckhard Hess from the 1960s and 1970s report that the pupils dilate or constrict according to the interest value, arousing content, or mental demands of visual stimuli. However, Hess mostly used small sample sizes and undocumented luminance control. In a first experiment ($N = 182$) and a second preregistered experiment ($N = 147$), we replicated five studies of Hess using modern equipment. Our experiments (1) did not support the hypothesis of gender differences in pupil diameter change with respect to baseline (PC) when viewing stimuli of different interest value, (2) showed that solving more difficult multiplications yields a larger PC in the seconds before providing an answer and a larger maximum PC, but a smaller PC at a fixed time after the onset of the multiplication, (3) did not support the hypothesis that participants' PC mimics the pupil diameter in a pair of schematic eyes but not in single-eyed or three-eyed stimuli, (4) did not support the hypothesis of gender differences in PC when watching a video of a male trying to escape a mob, and (5) supported the hypothesis that arousing words yield a higher PC than non-arousing words. Although we did not observe consistent gender differences in PC, additional analyses showed gender differences in eye movements towards erogenous zones. Furthermore, PC strongly correlated with the luminance of the locations where participants looked. Overall, our replications confirm Hess's findings that pupils dilate in response to mental demands and stimuli of an arousing nature. Hess's hypotheses regarding pupil mimicry and gender differences in pupil dilation did not replicate.

1. Introduction

In the 1960s and 1970s, psychologist and ethologist Eckhard Hess published a number of papers in which he advanced the theory that the pupils dilate or constrict in response to visual stimuli of different interest value, arousing content, mental demands, or taste (e.g., Hess, 1965, 1968, 1972, 1973a, 1975b; Hess and Goodwin, 1974; Hess and Polt, 1960, 1964, 1966; Hess et al., 1965; Polt and Hess, 1968). The first study by Hess on pupil response was published in *Science* in 1960 (Hess and Polt, 1960). The results of that paper showed that the pupils of the female participants dilated when viewing an image of a mother and a baby, a baby, or a partially naked male, whereas the male participants exhibited pupil dilation when viewing a partially naked female. Hess and Polt concluded that "*there is a clear sexual dichotomy in regard to the interest value of the pictures, with no overlap between sexes*" (p. 350).

The works of Hess appear to have a considerable influence on what researchers have come to believe about pupil response. Janisse (1977) pointed out: "*Psychology's debt to Hess lies in his discovery and popularization of applications for pupillometry to current research issues*" (p. 19). As of today, Google Scholar lists more than 900 citations to Hess and Polt (1960), with 65% in the last ten years. Hess's work on pupillometry is often cited in psychology and psychophysiology handbooks (e.g., Andreassi, 1980; Stern et al., 2001). In a well-cited review, Laeng et al. (2012) commented: "*The measurement of pupil diameter in psychology* (*in short,* "*pupillometry*") *has just celebrated 50 years. The method established itself after the appearance of three seminal studies (Hess and Polt, 1960, 1964; Kahneman and Beatty, 1966)*" (p. 18). Similarly, in a more recent review, Mathôt (2018) stated: "*Since the seminal studies by Hess and Polt (1960, 1964, Hess et al., 1965) and Kahneman and Beatty (1966), whose conclusions by and large still hold, there has been little theoretical development in this area*". Hess himself kept newspaper items about his work: in the Drs. Nicholas and Dorothy Cummings Center for the History of Psychology, at the University of Akron, Ohio, where Hess's work is archived (Appendix A), we retrieved more than 100 newspaper items about his findings. In recent times, the topic of pupillometry still draws the regular attention of science journalists and popular press worldwide (e.g., Dovey, 2014; Lewis, 2016; Martinez, 2015).

The pupillometry research of Hess is not without criticism. One recurring point of critique concerns possible differences in luminance between visual stimuli (Goldwater, 1972; Loewenfeld and Lowenstein,

* Corresponding author.
E-mail address: j.c.f.dewinter@tudelft.nl (J.C.F. de Winter).

https://doi.org/10.1016/j.ijpsycho.2021.03.003
Received 17 February 2020; Received in revised form 1 March 2021; Accepted 9 March 2021
Available online 22 March 2021
0167-8760/

1993) and between different locations within the same image, for example, when shifting gaze from a darker to a lighter area of an image (Janisse, 1977). Janisse (1977): "*If a study used a picture of a white male, wearing only dark trousers, the pupil would be larger if the subject looked at the trousers than if he looked at the face. If two subjects, one male and one female each preferred to look at a different part of the picture, they would have different pupillary responses* (*one dilation and one constriction*)" (p. 6). Another criticism concerns the plausibility of the bidirectionality of pupil response. Loewenfeld (1966) argued that there is no physiological evidence that any stimulus other than light can cause the pupil to constrict. Similarly, Nunnally et al. (1967), Peavler and McLaughlin (1967), Janisse (1973), and Garrett et al. (1989) argued that the pupil responds by dilating to pleasant as well as aversive stimuli. Hess has also been criticized for using small sample sizes (Skinner, 1980; Woodmansee, 1966; Zuckerman, 1971), and for the fact that he did not report statistical analyses but just based his conclusions on the observed mean pupil dilation (Janisse, 1977). An overview of prior criticisms of Hess's work is provided in Appendix B.

Given the impact of Hess's work, it seems worthwhile to examine whether the findings of Hess replicate. We selected five studies for replication: three highly cited and two lesser-known ones. The highly cited ones ('Images of five themes', 'Multiplications', and 'Schematic eyes') were included because they are among the most seminal and influential works of Eckhard Hess. The other two studies ('Western' and 'Visually presented words') are less influential but also relate to Hess's hypothesis about the association between visual interest and pupil dilation. The Western study is methodologically interesting, as the stimulus is a movie, which poses specific challenges for pupillometry research. The visually presented words are also interesting because these stimuli are offered in text-only form and likely free from visual confounders, such as differences in luminance between the stimuli.

1.1. Study 1. Images of five themes

In the aforementioned *Science* paper by Hess and Polt (1960; 930 citations in Google Scholar as of April 12, 2021), four males and two females looked at five images. The authors reported that the area of the pupils of the males increased by 18% when viewing an image of a partially nude female, whereas females exhibited only 5% pupil dilation. Females, on the other hand, showed a mean pupil dilation of 20% when viewing an image of a partially nude male, compared to a 7% dilation of male participants. Moreover, females exhibited a mean pupil dilation of 25% for an image portraying a mother with a baby and 17% for an image of a baby, a response not observed in the males, who exhibited only 5% and 0% dilation, respectively. No substantial difference in pupil response between male and female participants was found for an image of a landscape.

1.2. Study 2. Multiplications

In a second *Science* publication, Hess and Polt (1964; 1080 citations) reported that pupil size relates to mental effort. Five participants (four males, one female) were asked to solve four multiplications that were presented orally. The authors reported a mean increase in pupil diameter of 10.8% for the easiest multiplication (7×8) up to 21.6% for the most difficult one (16×23). Many other studies have established the phenomenon of pupil dilation during cognitively demanding tasks (e.g., Boersma et al., 1970; Bradshaw, 1968; Payne et al., 1968; Schaefer et al., 1968; see Van der Wel and Van Steenbergen, 2018 for a review). Ahern and Beatty (1979) showed pupil dilation during multiplication tasks, and Klingner et al. (2008) and Marquart and De Winter (2015) successfully replicated this finding. Herein, we aimed to replicate whether the difficulty level of the multiplication is associated with the degree of pupil dilation, as reported by Hess and Polt (1964). A limitation of previous research on this topic (Ahern and Beatty, 1979; Klingner et al., 2008; Marquart and De Winter, 2015) is that participants were given a fixed time to solve the multiplication. It can be expected that participants solve easier multiplications more quickly, resulting in earlier constriction back to baseline levels while awaiting the next multiplication, thus yielding a relatively low average dilation over the whole calculation period. In the present replication, we aimed to correct for the confounding of pupil dilation and task completion time by asking participants to press the spacebar and give their answer as soon as they had solved the problem.

1.3. Study 3. Schematic eyes

Hess (1975a; 268 citations) investigated whether images of schematic eyes evoke a pupil response. This study was first mentioned in a brief conference summary (Hess, 1969), after which it was presented in Hess (1973c) and Hess and Goodwin (1974) and summarized in Hess (1975b) and Hess and Petrovich (1987). Hess showed participants (ten males, ten females) slides with one, two, or three horizontally aligned schematic eyes with three sizes of the inner circle, representing the pupil. Hess reported that participants' pupil response did not vary systematically as a function of the pupil size of the single and triple schematic eyes but did dilate more for larger pupils when the eyes were presented as a pair, that is, for the representation that mostly resembled eyes of a human. Hess (1975a) argued that his findings had an evolutionary basis, a "*behavior that is innate or perhaps learned very early in life*" (p. 112).

1.4. Study 4. Western

Hess (1975b, pp. 193–197; the book in which this study appears is cited 291 times) presented findings from 100 participants (50 males, 50 females; sample size reported in Hess and Goodwin, 1974) who watched a 30-min episode of a TV series. Based on the audio recording of a conference talk (Hess, 1973b) and a description of the episode in Hess (1975b), we deduced that the episode is called "*Survival*" from the TV series "*A man called Shenandoah*", a Western aired between 1965 and 1966 (Sagal, 1965). Hess (1975b) highlighted a specific scene (between 880 s and 930 s) from that 30-min episode, where the hero of the series is harassed by a crowd, tries to escape, but is eventually caught. Hess (1975b) reported that "*during this time the men's pupils get bigger and the women's pupils decrease in diameter. When he is actually caught the mens pupils constrict sharply, while there is a brief period of dilation for the women subjects*" (p. 197) (Fig. 1). Hess (1975b) suggested that these findings point to a fundamental difference between men and women: "*The men like to see the man get away; the women like to see the man caught*" (p. 196).

1.5. Study 5. Visually presented words

Polt and Hess (1968; 18 citations) investigated the effect of (1) the size of visually presented words and (2) the emotional content of these words on pupil response of male versus female participants. The sample consisted of nine males and six females. Four words (i.e., 'hostile', 'squirm', 'flay', and 'nude') were presented two times each, once with large and once with small font. Polt and Hess presented no hypotheses. The participants' pupils slightly constricted (mean = −0.4%) and slightly dilated (mean = 0.1%) when viewing the large versus small font, respectively; the effect of font size was not statistically significant. The results section reported that there were no significant differences between men and women: "*While there are distinct sex differences in responses, none of these proved to be significant at the .05 level*" (p. 389). However, the authors hinted that the observed dilation for the words 'flay' and 'nude' was because these words are "*both rich and individualistic in imagery related arousal*" (p. 390) and that threatening words cause pupil constriction.

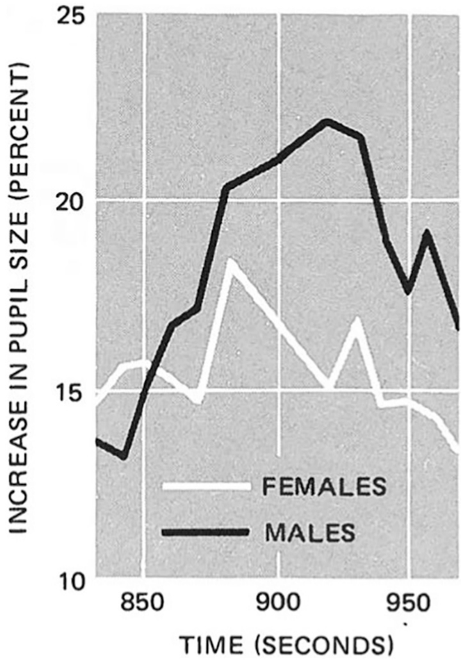


Fig. 1. Mean pupil diameter change of 50 male and 50 female participants for a scene from an episode of a Western TV series (graph taken from Hess, 1975b). At 830 s, the pupil diameter increase with respect to baseline is 13.7% for males and 14.6% for females. At 920 s, this value has become larger for males (22.1%) than for females (15.2%). At 950 s, the gender difference has diminished again to 17.6% for males and 14.8% for females.

1.6. Aim and approach of this study

This study aimed to replicate the above five studies using modern equipment. Hess used a Bell and Howell Slider Master slide projector (Hall, 1959) for presenting the visual stimuli (Appendix C). Using the same projector and a replica of the presentation equipment, we found that slide changes yield a 1-s period of darkness (see Appendix D) and a corresponding increase in pupil diameter, see Fig. 2 and Appendix E. Fig. 2 further illustrates that when a new slide is presented (at around 0 s, 10 s, and 20 s), the pupils constrict rapidly (within a second), followed by slight re-dilation. We decided to prevent these luminance effects by using a computer monitor instead of a slide projector. Furthermore, Hess used a camera that recorded at a frequency of 0.5 Hz (except for the Western, where one measurement was taken every 10 s), which might not be sufficient for capturing rapid changes in pupil diameter. We used an eye tracker that recorded the pupil diameter at 2000 Hz. Finally, Hess reported pupil size only. We recorded eye movements to examine whether gender differences in pupil diameter can be explained by gender differences in the extent to which participants focused on darker or brighter parts of the stimulus.

The replications of the above five studies were performed by means of two experiments. In Experiment 1, Studies 1 and 2 were replicated, and in Experiment 2, Studies 3, 4, and 5 were replicated, and Study 1 was performed again with modifications after applying lessons learned from Experiment 1. More specifically, because Experiment 1 showed that luminance had strong stimulus-specific effects, we decided to use line drawings instead of images, to ensure that luminance was constant regardless of visual stimulus. Experiment 1 was not preregistered, as we were still unsure about confounders such as luminance and eye movements. In Experiment 2, we preregistered our hypotheses, stimuli, experimental protocol, data processing, and statistical analyses in the Open Science Framework (OSF) repository (Kooijman et al., 2018). Preregistration is a recommended solution for preventing problems related to biases in human reasoning such as hindsight bias (Nosek et al.,

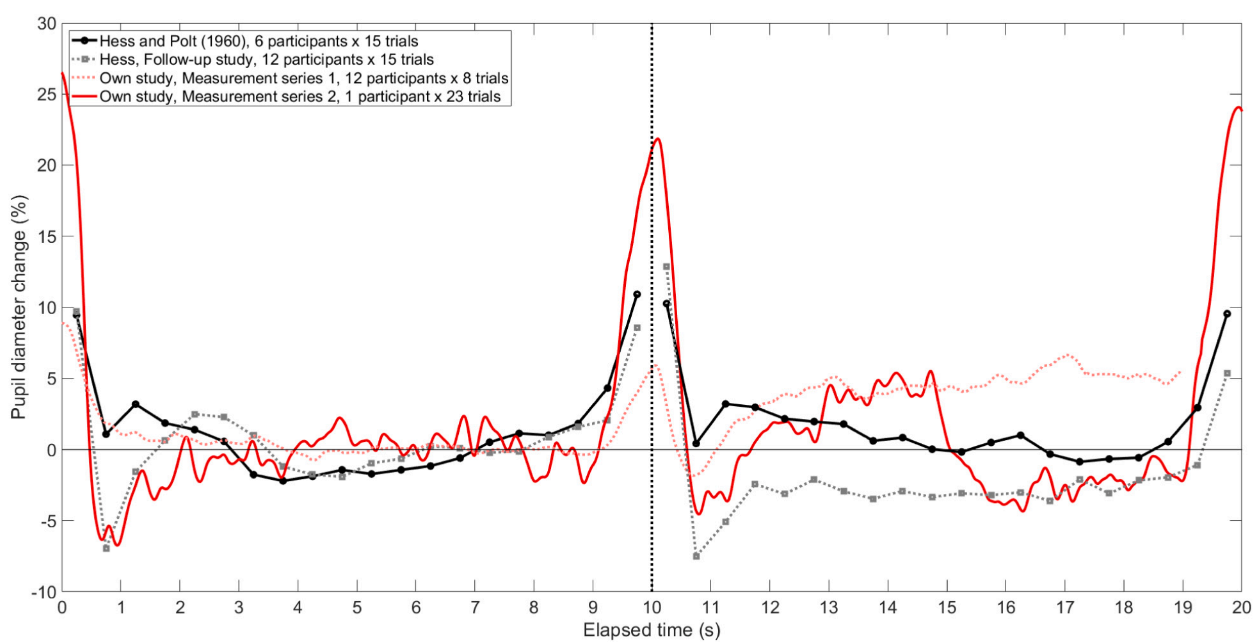


Fig. 2. Mean pupil diameter change (%) of participants as a function of time, calculated using raw data from Hess and Polt (1960; Study 1), raw data from a follow-up study by Hess, and two measurement series conducted with our replica of Hess's pupil apparatus. A positive value indicates pupil dilation; a negative value indicates pupil constriction. The dotted vertical line indicates the moment of transition (defined in Appendix E as the end of the 'full darkness' period) from a control slide to a stimulus slide for the Hess and Polt (1960) data and Measurement series 1, and from a control slide to another control slide for Measurement series 2. The differences in peak pupil diameter change around 10 s between the experiments are likely related to differences in the luminance of the slides (the slides in Measurement series 1 were darker than the slides in Measurement series 2). The increase in pupil diameter change after 10 s for Measurement series 1 is due to mental effort while solving multiplication problems (see Appendix E). Note that Hess and Polt (1960) discarded the first and last second of data for the control and stimulus slides. Further information is provided in Appendices D and E.

2018). Through preregistration, a clear distinction is established between hypothesis generation based on existing observations (i.e., prediction) and hypothesis testing using new observations. According to Brandt et al. (2014), any convincing replication should include a priori registration of the materials and methods.

In summary, this work aimed to replicate five works of Hess, using the same stimuli and measures as the original studies. Because we used modern equipment for stimulus presentation and measurement, and extra stimuli such as line drawings, the current replications do not qualify as direct replications. However, apart from necessary methodological modifications, our replications are as direct as possible. We refrained from performing a conceptual replication of Hess's theories, because, according to Zwaan et al. (2018), "*It is always possible to attribute a failed conceptual replication to the changes in procedures that were made. … Direct replications do not have this interpretational ambiguity*" (p. 8, see also Simons, 2014).

2. Methods

2.1. Overview of studies

Two experiments were performed. Experiment 1 consisted of two studies in the following order: Images of five themes (Replication of Study 1) and Multiplications (Replication of Study 2). An additional study aimed to determine the pupillary response due to screen luminance, and is described in Appendix J. Experiment 2 consisted of four studies in the following order: Schematic eyes (Replication of Study 3), Western (Replication of Study 4), Visually presented words (Replication of Study 5), and Line drawings of five themes (Replication of Study 1). A subsequent study about a visual inspection time task is described elsewhere (Eisma and De Winter, 2020).

2.2. Participants

Table 1 provides an overview of the characteristics of the participants. They were all engineering students at the Delft University of Technology, and mostly male (70%). In comparison, Hess used 67%, 75%, 50%, 50%, and 60% males in Studies 1, 2, 3, 4, and 5, respectively. In Hess and Polt (1960; Study 1), participants' mean age was approximately 24 years (see also Appendix F), and in Polt and Hess (1968; Study 5), participants' ages ranged between 24 and 45 years. The mean ages for Hess's Studies 2–4 are unavailable. The experiments were approved by the Human Research Ethics Committee of the TU Delft. All participants provided written informed consent. None of the participants in Experiment 2 had taken part in Experiment 1.

Table 1
Participant characteristics in Experiments 1 and 2.

	Experiment 1	Experiment 2
No of participants	182; 129 (71%) males, 53 (29%) females	147; 102 males (69%), 45 (31%) females
Mean age (*SD*)	23.2 years (1.81)	23.3 years (2.13)
No of additional participants excluded due to data logging errors	3	1
Seeing aids	None: 125 (69%) Glasses: 17 (9%) Contact lenses: 39 (21%)	None: 112 (76%) Glasses: 13 (9%) Contact lenses: 22 (15%)
Caffeine in the past two hours	No: 125 (69%), yes: 56 (31%)	No: 87 (59%), yes: 60 (41%)
Smoked in the past two hours	No: 172 (95%), yes: 9 (5%)	No: 142 (97%), yes: 5 (3%)

Note. For Experiment 1, information about seeing aids, caffeine use, and smoking is unavailable for one participant. For Experiment 2, the number of participants wearing glasses during the experiment was smaller than 13, as some of them were asked to remove their glasses to enhance eye-tracking quality.

2.3. Power analysis (Experiments 1 and 2)

In a previous power-analysis (two-tailed, alpha = 0.05), we calculated that, for 160 participants, the achieved power for detecting a 3% difference in pupil diameter is between 86% (a worst-case scenario in a between-subjects design) and 100% (a best-case scenario in a within-subject design) (Kooijman et al., 2018). To place the 3% difference in pupil diameter in perspective: for Study 1, the gender differences in pupil diameter change reported by Hess and Polt (1960) (excluding the 'control' image of the landscape) ranged between 6% and 11%; for Study 2, the pupil diameter change between the easiest and most difficult multiplication differed with 10.8%; for Study 3, Hess (1975a) reported a pupil diameter change difference of 3.8% between the smallest and the largest pairs of schematic pupils; for Study 4, Hess (1975b) reported gender differences in pupil diameter change of 2–7%; and for Study 5 (Polt and Hess, 1968), pupil diameter change with respect to baseline measurements for the words 'flay' and 'nude' was about 2.5% (Kooijman et al., 2018).

2.4. Apparatus (Experiments 1 and 2)

We used an EyeLink 1000 Plus desktop eye tracker (SR Research Ltd., version II CL v5.08; Fig. J1) for acquiring data at 2000 Hz of the right eye, except for one participant in Experiment 2 for whom the left eye was recorded instead. Tracking mode was set to 'pupil-CR' and pupil tracking to 'Centroid'. The EyeLink records the pupil diameter in arbitrary units. The pupil diameter in millimeters was obtained through a multiplication factor based on a calibration with printed circles of known diameter.

The visual stimuli were presented using a computer running 'SR Research Experiment Builder' (version 1.10.1386), using a 64-bit Windows 7 Professional operating system and Intel Core i7-4790K CPU @ 4.00 GHz, NVIDIA GeForce GTX 970 graphics card, and ASUS Xonar DS Audio Device. Experiment 1 used a 24-inch monitor (Model: BenQ XL2420Z) with a resolution of 1920 × 1080 pixels (display area 531 × 298 mm), whereas Experiment 2 used a 25-inch monitor (Model: BenQ XL2540-B) with the same resolution (display area 544 × 303 mm). The screen refresh rate was set to 60 Hz and 144 Hz for Experiments 1 and 2, respectively. The distance between the monitor and the table edge was approximately 950 mm. The distance between the camera and the head support was approximately 540 mm. For a distance of 910 mm between the monitor and the eyes, the display subtended an approximately 33° horizontal and 19° vertical viewing angle.

2.5. Control slide (Experiments 1 and 2)

Each stimulus was preceded by a control slide, which was used to obtain a baseline pupil diameter. Hess used control slides containing five numbers, likely in portrait format (Appendix G). Our control slides were similar to Hess in terms of the layout of the numbers. Because our monitor had a wider aspect ratio than Hess's slides, we used nine instead of five numbers. Our control slide consisted of the numbers 1 to 9, presented in a black outline of 2-pixel thickness, in Mangal font with a height of 44 pixels (0.8°) and a width between 20 pixels (0.4°) and 30 pixels (0.5°) (see Fig. G3).

All stimuli and control slides were presented on a gray background, with a grayscale value of 50%, or 127 on an 8-bit scale from 0 (black) to 255 (white). The order of the stimuli within each study was random and different for each participant.

2.6. Stimuli (Experiments 1 and 2)

2.6.1. Replication of Study 1 (Hess and Polt, 1960): Images of five themes (Experiments 1 and 2)

Ten images were used. Five images were selected from a presentation by Hess in 1962, where each image was accompanied by a bar plot with the same results as in Hess and Polt (1960; Appendix H). The other five images were modern equivalents retrieved from the internet, cropped and mirrored to resemble the original images. Modern images were added because the original images may not evoke arousal due to cultural change (Greenfield, 2017). The images were adjusted to all have a mean grayscale level of 50%, and a similar standard deviation of the grayscale level between the original and modern version of each image. Note that after the completion of Experiment 1, we discovered that Hess and Polt (1960) used different images of the same themes in their experiment (Appendix I).

The Replication of Study 1 of Experiment 1 made use of stimuli that were also used by Hess and Polt (1960). A limitation of this approach is that it does not prevent the pupillary light reflex. Although the mean grayscale levels of the control slide and the stimulus slides were all close to 50%, there were strong variations in grayscale levels between different parts of the stimulus slides. In Experiment 2, we used line drawings instead of images to prevent pupillary light reflexes. This decision is consistent with a recommendation on pupillometry by Janisse (1974): "*If visual stimuli are used, they should be of minimal contrast and be line drawings, words, numbers or other symbols*" (p. 3). We used ten line drawings, two for each of the five themes (see Appendix H). The drawings were obtained from stock photo databases and adjusted to give a uniform appeal in drawing style. The '*Image Trace*' tool (option: '*Line Art*') in Adobe Illustrator was used to equalize the thickness of all lines to 2 pixels.

2.6.2. Replication of Study 2 (Hess and Polt, 1964): Multiplications (Experiment 1)

Participants were presented with twelve multiplication problems: 7 × 8, 9 × 8, 6 × 7, 8 × 13, 7 × 14, 6 × 16, 13 × 14, 12 × 14, 9 × 17, 16 × 23, 15 × 17, and 16 × 18. Four of these multiplications (7 × 8, 8 × 13, 13 × 14, 16 × 23) were used by Hess and Polt (1964). The eight additional multiplications had similar difficulty levels as the multiplications in Hess and Polt (based on a classification method by Marquart and De Winter, 2015). The multiplications were presented in a black outline of 2-pixel thickness, in Mangal font with a height of 204 pixels (3.6°) and a width between 547 pixels (9.5°) and 842 pixels (14.6°), see Fig. J2 for an example. We used an outline to minimize the effect of luminance on pupillary response.

2.6.3. Replication of Study 3 (Hess, 1975a): Schematic eyes (Experiment 2)

First, a drawing of a happy face and a drawing of an angry face were shown to introduce the participants to the topic of schematic eyes. These faces were also presented in the same works by Hess where the schematic eyes study was reported (Hess, 1973c, 1975a, 1975b; Hess and Goodwin, 1974; Hess and Petrovich, 1987), see Appendix J for details.

Next, nine stimuli containing schematic representations of eyes were presented (see Fig. J3). The stimuli contained a single eye, two eyes, or three eyes, with three levels of pupil size (i.e., small, medium, and large). The schematic eyes were redrawn from Hess and Goodwin (1974). The diameter of the outer circle was 66 pixels (1.2°), and the diameter of the inner circle was 27, 37, and 45 pixels (0.5, 0.7, and 0.8°) for small, medium, and large pupils, respectively. The center-to-center distance for the two- and three-eyed stimuli was 228 pixels (4.1°), and the line thickness was proportional to the original drawings.

2.6.4. Replication of Study 4 (Hess, 1975b): Western (Experiment 2)

A 75-s video clip of the episode "*Survival*" from the Western TV series "*A man called Shenandoah*" was shown, corresponding to the scene highlighted by Hess (1975b) (Fig. 1; see Fig. J4 for a video frame). The clip was 1348 pixels wide and 1080 pixels high (original size: 720 × 480 pixels). The frame rate was 25 fps.

2.6.5. Replication of Study 5 (Polt and Hess, 1968): Visually presented words (Experiment 2)

Participants were presented with twelve words. Four words (i.e., 'hostile', 'squirm', 'flay', and 'nude') were used by Polt and Hess (1968). The other eight words ('flirt', 'party', 'sadist', 'demon', 'aroma', 'harmonica', 'fragment', and 'standby') were selected from Mohammad (2018), who, using crowdsourcing, rated 20,007 English words on valence, arousal, and dominance. Three of the four words from Polt and Hess were available in Mohammad's list, and all three were characterized by high arousal and low-to-medium valence and dominance. For the four combinations of low and high valence and arousal, we selected two words from Mohammad's list that scored medium in dominance, appeared in the online Dutch dictionary Van Dale (2019), and had the same meaning in English and Dutch. The words were presented in a black outline of 2-pixel thickness, in Mangal font with a height of 253 pixels (4.5°) from the top of the ascenders to the bottom of the descenders (151 pixels or 2.7° when excluding ascenders and descenders) and a width between 387 pixels (6.9°) and 1288 pixels (22.7°) (see Fig. J5 for an example). Table 2 shows the twelve words together with their ratings of valence, arousal, and dominance.

2.7. Light and sound conditions (Experiment 1 and 2)

The windows next to the eye tracker were blinded. Luminescent tube lights mounted to the ceiling lit up the room. In Experiment 2, the participants wore closed-back headphones (Beyerdynamic DT-770 Pro 32 Ohm) to limit the effect of sounds from the environment and to present the sound of the video clip. In Experiment 2, the illuminance in the room at the location where the participant's eyes would be positioned was around 400 lx (as measured with a Konica Minolta T-10MA illuminance meter), and the sound level of the computer was set to 80%.

The lighting conditions in Experiments 1 and 2 were such that the pupil diameter was at a nominal level of about 4 mm. More specifically, in Experiments 1 and 2, the mean of participants' mean pupil diameter while viewing the ten control slides before the images of five themes was 3.96 mm ($SD = 0.48$ mm; $N = 182$) and 3.98 mm ($SD = 0.56$ mm; $N = 147$), respectively. Participants' caffeine consumption and smoking in the two hours prior to the experiment showed no significant point-biserial correlations with pupil diameter in Experiment 1 ($r = 0.12$, $p = 0.100$; $r = 0.01$, $p = 0.864$, respectively) nor in Experiment 2 ($r = 0.04$, $p = 0.653$; $r = 0.08$, $p = 0.307$, respectively).

Table 2
Word stimuli in Experiment 2. Ratings of valence, arousal, and dominance were taken from Mohammad (2018).

	Valence	Arousal	Dominance
Polt and Hess (1968)			
Flay	N/A	N/A	N/A
Hostile	0.188	0.877	0.474
Nude	0.490	0.915	0.200
Squirm	0.235	0.824	0.373
High valence & high arousal			
Flirt	0.792	0.790	0.538
Party	0.948	0.840	0.547
Low valence & high arousal			
Sadist	0.042	0.918	0.500
Demon	0.037	0.908	0.509
High valence & low arousal			
Aroma	0.823	0.235	0.442
Harmonica	0.847	0.235	0.510
Low valence & low arousal			
Fragment	0.211	0.316	0.429
Standby	0.260	0.224	0.386

Note. Ratings range from 0 (*low*) to 1 (*high*). N/A = not available.

2.8. Procedures and instructions (Experiments 1 and 2)

Upon arrival, participants were informed about the aim of the experiment via a consent and procedures form. The form was also available on a student portal for a course taught by the principal investigator.

Participants faced the monitor and adjusted the seat height so that they could comfortably position their head in the support. The eye tracker was then calibrated. Each experimental study was preceded by a slide introducing the upcoming study. The participants were informed that they were not required to do anything but looking at the screen. They were also asked to focus on the nine numbers on the control slide in ascending order. These instructions are in line with instructions by Hess we retrieved from the archive (Appendix K). Before the multiplication study, participants were given the following instructions: "*Each problem will be shown for 30 seconds. When you solve the problem, hit the space bar as fast as possible and call out the answer. Please keep either your left or right hand on the keyboard during the entire block*". Dutch-speaking participants were allowed to give their answers to the multiplications in Dutch instead of English.

In Experiment 1, participants performed a 'drift correction' between each control slide and subsequent thematic image. In Experiment 2, a drift correction was performed before the first control slide of each study. During the drift correction, participants focused on a black circle in the middle of a gray background (grayscale value of 50%) and pressed the spacebar to continue. Note that the drift correction does not affect the calibration; it can only be used to perform a retrospective check of the calibration error (SR Research, 2009).

In Experiment 1, after the participants had completed the studies, they completed a questionnaire about their age and gender, whether they wore seeing aids, and whether they had consumed caffeinated drinks or smoked in the past two hours. In Experiment 2, a similar questionnaire was completed before the calibration.

The stimulus and control slides were shown for 10 s each. Exceptions were the multiplications in Experiment 1 (Replication of Study 2) and the 75-s Western in Experiment 2 (Replication of Study 4). The multiplications were shown until the participant pressed the spacebar or 30 s if the participant did not press the spacebar. In the latter case, an answering time of 30 s was imputed. If participants had the spacebar pressed at the onset of the presentation of the multiplication (which happened in 11 out of 2184 trials), then that trial was omitted from the analysis. Participants spent in the eye tracker approximately 12.5 min in Experiment 1 and 16.5 min in Experiment 2 (excluding the visual inspection time task).

In Experiment 1, all participants were tested by the same male experimenter. In Experiment 2, one female and three male experimenters tested 97, 30, 18, and 2 participants, respectively. The experimenter's role was to summarize the aim of the experiment, provide the participant with the informed consent form, calibrate the eye tracker, and answer questions raised by the participant. During the experiment, the experimenter sat behind a laptop at a separate table, without a direct view of where the participant was looking during the experiment.

2.9. Data processing (Experiments 1 and 2)

First, raw data of pupil diameter and horizontal and vertical gaze coordinates in pixels were filtered using a median filter with 100 ms interval. Pupil diameter data and eye movement data during blinks were linearly interpolated. MATLAB scripts are available in the Supplementary Material.

Polt and Hess (1968) mentioned that "*all scores reflect the per cent difference in mean pupil size during the 20 frames the eye looked at a stimulus (10 sec) with the mean of the pupil size during the previous 10 sec control period*" (p. 389). A protocol retrieved from the archive provides an additional detail, namely that "*the first and last two frames of each sequence were disregarded, to compensate for any variability in light at the time of slide change*" (Box *M4138*, folder *EARLY Pupil Research*). Woodmansee Jr. (1965) confirmed that Hess removed the first and last two frames: "*To reduce the contaminating overlap of data for adjacent stimulus periods, Hess disregards the first two and the last two of the 20 frames of film assigned to a given stimulus-presentation period*" (p. 53). We used the same data analysis approach as Hess. More specifically, for each stimulus slide, the percentage change $PC_{[1,9]}$ between the mean pupil diameter for the stimulus slide $\overline{p}_{s[1,9]}$ and the mean pupil diameter for the preceding control slide $\overline{p}_{c[1,9]}$ was calculated (Eq. 1). In other words, we used the 1–9 s interval instead of the entire 0–10 s interval, as we excluded the first and last 1 s, corresponding to the first two and the last two frames excluded by Hess.

$$PC_{[1,9]} = 100\% \frac{\overline{p}_{S[1,9]} - \overline{p}_{C[1,9]}}{\overline{p}_{C[1,9]}} \tag{1}$$

For the Western (Replication of Study 4; Hess, 1975b), which involved a video instead of static stimulus, the percentage change PC_t was calculated between the pupil diameter at each sampling instant $p_{s,t}$ (2000 Hz) and the mean pupil diameter during the preceding control slide $\overline{p}_{c[1,9]}$ (Eq. (2)).

$$PC_t = 100\% \frac{p_{s,t} - \overline{p}_{c[1,9]}}{\overline{p}_{c[1,9]}} \tag{2}$$

Graphs of PC_t as a function of the elapsed time were created for all five replication studies (preregistered for Experiment 2).

For the multiplications (Replication of Study 2; Hess and Polt, 1964), four alternative metrics were computed. More specifically, (1) the percentage change $PC_{[ans\text{-}2.5,ans]}$ was computed between the mean pupil diameter for the 2.5-s period before an answer was given $\overline{p}_{s[ans\text{-}2.5,ans]}$ (i.e., the 2.5-s period before the spacebar was pressed) and the mean pupil diameter for the 2.5-s period before presenting the multiplication $\overline{p}_{c[7.5,10]}$ (Eq. (3)). This is also the measure used by Hess and Polt (1964): "*the mean size of the pupil of one subject, recorded on five frames immediately before a question is asked, is compared with the mean size of the pupil at the period of maximum dimension, recorded on five frames immediately before the answer is given*" (p. 1191).

$$PC_{[ans-2.5,ans]} = 100\% \frac{\overline{p}_{s[ans-2.5,ans]} - \overline{p}_{c[7.5,10]}}{\overline{p}_{c[7.5,10]}} \tag{3}$$

If the participant answered within 2.5 s, then $PC_{[ans\text{-}2.5,ans]}$ was defined using the entire calculation interval (Eq. (4)).

$$PC_{[ans-2.5,ans]} = 100\% \frac{\overline{p}_{s[0,ans]} - \overline{p}_{c[7.5,10]}}{\overline{p}_{c[7.5,10]}} \tag{4}$$

Hess and Polt (1964) argued that they used the above-mentioned measure because the pupil diameter "*reached a maximum dimension immediately before an answer was given, and then reverted to the previous control size*" (p. 1191). To capture the rationale of Hess and Polt, we therefore additionally calculated (2) the percentage change PC_{max} between the maximum pupil diameter $p_{s,max}$ during the calculation interval and the mean pupil diameter for the 2.5-s period before presenting the multiplication $\overline{p}_{c[7.5,10]}$ (Eq. (5) and (3)) the percentage change PC_{ans} between the pupil diameter when providing the answer $p_{s,ans}$ (i.e., at the moment of pressing the spacebar) and the mean pupil diameter for the 2.5-s period before presenting the multiplication (Eq. (6)). Finally, we computed (4) the pupil diameter change (PC_3) between the pupil diameter 3 s after the presentation of the multiplication problem $p_{s,3}$ and the mean pupil diameter for the 2.5-s period before presenting the multiplication (Eq. (7)), as an indication of pupil dilation at a fixed moment in time.

$$PC_{\max} = 100\% \frac{p_{s,max} - \overline{p}_{c[7.5,10]}}{\overline{p}_{c[7.5,10]}} \tag{5}$$

$$PC_{ans} = 100\% \frac{p_{s,ans} - \overline{p}_{c[7.5,10]}}{\overline{p}_{c[7.5,10]}} \quad (6)$$

$$PC_3 = 100\% \frac{p_{s,3} - \overline{p}_{c[7.5,10]}}{\overline{p}_{c[7.5,10]}} \quad (7)$$

2.10. Statistical tests to examine whether Hess's effects replicate (Experiments 1 and 2)

The following statistical tests were performed at the level of participants. The analyses of Experiment 1 were not preregistered, whereas the analyses for Experiment 2 were (Kooijman et al., 2018). We used an alpha value of 0.05 and two-tailed tests. We opted for simple statistical tests because we were interested in replicating the specific effects of Hess as described in the Introduction. If expected effects were not in full agreement but in partial agreement with Hess, this was interpreted as a partial confirmation of Hess's findings.

2.10.1. Replication of Study 1 (Hess and Polt, 1960): Images of five themes (Experiments 1, not preregistered; Experiment 2, preregistered)

Independent-samples *t*-tests were performed between the $PC_{[1,9]}$ values of male and female participants. For Experiment 1, the *t*-tests were performed for each of the ten images. For Experiment 2, which involved two comparable line drawings per theme, the $PC_{[1,9]}$ value was first averaged between the two drawings per theme. Thus, for Experiment 2, five independent-samples *t*-tests were performed. The findings of Hess and Polt (1960) were confirmed if male participants had a statistically significantly higher $PC_{[1,9]}$ than female participants for the images/drawings of a nude female, and if female participants had a statistically significantly higher $PC_{[1,9]}$ than male participants for the images/drawings of the baby, mother and baby, and nude male.

2.10.2. Replication of Study 2 (Hess and Polt, 1964): Multiplications (Experiment 1, not preregistered)

To investigate the hypothesis of whether the difficulty of the multiplication relates to the degree of pupil dilation, tests of within-subject linear contrasts were performed for $PC_{[ans\text{-}2.5,ans]}$, with the 12 multiplications introduced in the following order: 9 × 8, 6 × 7, 7 × 8, 6 × 16, 8 × 13, 7 × 14, 9 × 17, 12 × 14, 13 × 14, 15 × 17, 16 × 18, 16 × 23. This order was based on the observed average time it took participants to solve the multiplications. A test of within-subject linear contrasts was also performed for $PC_{[ans\text{-}2.5,ans]}$, for the four multiplications used by Hess and Polt (1964), in the following order: 7 × 8, 8 × 13, 13 × 14, and 16 × 23. This order corresponds to the average time it took participants to solve the multiplications and was identical to the difficulty order assumed by Hess and Polt. Support for Hess and Polt's hypothesis that "*there is a complete correlation between difficulty and the mean response of the five subjects*" (p. 1191) was obtained if the contrast analysis for the four multiplications produced a statistically significant result, with more difficult multiplications yielding a higher $PC_{[ans\text{-}2.5,ans]}$.

2.10.3. Replication of Study 3 (Hess, 1975a): Schematic eyes (Experiment 2, preregistered)

One-way repeated-measures ANOVAs of $PC_{[1,9]}$ were performed, with the size of the schematic eyes as a within-subject variable (small, medium, large). The repeated-measures ANOVA was performed separately for one-, two-, and three-eyed stimuli. Hess's hypothesis was confirmed if a statistically significant increase in $PC_{[1,9]}$ as a function of the presented pupil diameter was observed for the two-eyed stimuli but not for the one- and three-eyed stimuli.

2.10.4. Replication of Study 4 (Hess, 1975b): Western (Experiment 2, preregistered)

The difference (d) between the PC_t of male and female participants was computed per sampling instant of the video (2000 Hz). If $d > 0$, then males have higher PC_t than females; if $d < 0$, females have a higher PC_t than males. Support for Hess's hypothesis was obtained if d increased between 16.5 s and 57.0 s (i.e., while the man tries to escape), and if d decreased between 57.0 s and 73.4 s (i.e., the man is caught and subdued until the moment when the scene starts fading).

2.10.5. Replication of Study 5 (Polt and Hess, 1968): Visually presented words (Experiment 2, preregistered)

A two-way repeated-measures ANOVA of $PC_{[1,9]}$ was performed with valence and arousal levels as within-subject variables. The pupil diameter was first averaged between the two words per category. The four words used by Polt and Hess (1968) were analyzed using a one-way repeated-measures ANOVA of $PC_{[1,9]}$. Polt and Hess's implicit hypothesis that arousing words evoke pupil dilation was confirmed if the words with high arousal ratings yielded a statistically significantly higher dilation than words with low ratings of arousal.

2.10.6. Additional non-preregistered analyses (Experiments 1 and 2)

The above-mentioned statistical tests were used to examine whether Hess's effects replicate. We performed several follow-up analyses to gain a more in-depth understanding of the participants' pupil dilation. More specifically, omnibus tests and pairwise comparisons were conducted to examine pupil dilation differences between (categories of) stimuli. Furthermore, as mentioned in the Introduction, viewing behavior is a possible confounder of (gender differences in) pupil dilation. Therefore, additional analyses were conducted to examine whether the different stimuli cause different degrees of pupil dilation and whether these differences in pupil dilation are explained by eye movements and the corresponding local darkness of the stimuli. The local darkness (LD_t) was computed for stimuli with variable luminance, namely the five themes in Experiment 1 (Replication of Study 1), the schematic eyes (Replication of Study 3), and the Western (Replication of Study 4). LD_t was defined based on where participants looked at a particular moment (Bradley et al., 2017). More precisely, LD_t was defined for each time sample as the mean grayscale value on a scale from 0% (white pixels only) to 100% (black pixels only) of a 21 × 21-pixel area around the gaze sample per participant. We use a darkness scale instead of a scale from black to white, because darkness is more intuitively interpretable when presented in graphs together with pupil diameter, as a high level of darkness is expected to yield pupil dilation due to the light reflex. We opted for a narrow region of 21 × 21 pixels (about 0.4° horizontal and vertical) to obtain an indication of foveal stimulation only.

For the images of five themes in Experiment 1 (Replication of Study 1) and the schematic eyes in Experiment 2 (Replication of Study 3), the global darkness (i.e., the mean darkness across the entire image) was constant and close to 50% for the entire 10 s of stimulus presentation. For the Western (Replication of Study 4), however, the global darkness differed per video frame. Therefore, for the Western, we also calculated the global darkness GD_t (i.e., the mean darkness of the entire video frame) at each sampling instant (at 2000 Hz).

The following non-preregistered analyses were conducted:

- *Replication of Study 1* (Hess and Polt, 1960)*: Images of five themes* (*Experiment 1*). As mentioned above, next to the five images retrieved from a presentation by Hess in 1962, we included five modern images. To investigate whether participants responded differently to the old versus the modern images, a two-way repeated-measures ANOVA of $PC_{[1,9]}$ was performed, with image age (original vs. modern) and image theme as within-subject factors. Pairs of stimuli were statistically compared using paired *t*-tests with Bonferroni correction (correction factor = 45). Additionally, we conducted a two-way repeated-measures ANOVA of the local darkness at the onset of the stimulus $LD_0 = ld_{s,0}$, again with image age and image theme as within-subject factors. Significant differences between pairs of stimuli were assessed using paired *t*-tests with Bonferroni correction (correction factor = 45). Pearson's correlation between

LD_0 averaged across participants and the corresponding pupil diameter change 1 s later (PC_1) averaged across participants was computed to examine whether local darkness is predictive of pupil diameter change ($n = 10$ images). Also, heatmaps of the eye-gaze coordinates were created to examine gender differences in viewing behavior, and the duration for which males versus females looked at specific 150 × 150-pixel (2.6 × 2.6°) areas of interest were compared using independent-samples *t*-tests. For the heatmaps, the horizontal and vertical gaze sample coordinates were used, not fixation coordinates. Finally, independent-samples *t*-tests were performed for the mean local darkness (%) between 11 s and 19 s ($LD_{[1,9]} = \overline{ld}_{s[1,9]}$) of male versus female participants to investigate whether there were gender differences in local darkness.

- *Replication of Study 1* (Hess and Polt, 1960)*: Images of five themes* (*Experiment 2*). A repeated-measures ANOVA of $PC_{[1,9]}$ was performed with the five image themes as a within-subject factor. Pairs of stimuli were statistically compared using paired *t*-tests with Bonferroni correction (correction factor = 10). Again, heatmaps of the eye-gaze coordinates were created, and the duration for which males versus females looked at specific 350 × 150-pixel (6.2 × 2.7°) and 150 × 150-pixel (2.7 × 2.7°) areas of interest for Female 1 and Female 2, respectively, were compared through independent-samples *t*-tests.
- *Replication of Study 2* (Hess and Polt, 1964)*: Multiplications* (*Experiment 1*). The same tests of within-subject linear contrasts as in the analysis of $PC_{[ans\text{-}2.5,ans]}$ were performed for PC_{max}, PC_{ans}, and PC_3. The four pupil change measures ($PC_{[ans\text{-}2.5,ans]}$, PC_{max}, PC_{ans}, and PC_3) were plotted against the average time that took to solve each multiplication in Experiment 1, to inspect trends between pupil diameter change and answering time visually.
- *Replication of Study 3* (Hess, 1975a)*: Schematic eyes (Experiment 2)*. A two-way repeated-measures ANOVA of $PC_{[1,9]}$ was conducted with the number of schematic eyes and the depicted pupil sizes as within-subject factors in order to investigate whether the number and size of schematic pupils interact, in line with Hess's hypothesis that humans respond with pupil dilation to two-eyed stimuli only. Furthermore, because the accuracy of pupil diameter measurements may depend on eye movements, and because the schematic eyes were very different from each other (i.e., 1, 2, or 3 salient features present), we performed an analysis of eye movements. The number of saccades was used as a global index of visual scanning and eye movement activity. More specifically, the number of saccades since the start of the stimulus slide was calculated using a velocity threshold of 2000 pixels/s or 35°/s (see Eisma et al., 2018). A two-way repeated-measures ANOVA of the number of saccades was conducted with the number of schematic eyes and the depicted pupil sizes as within-subject factors. Finally, the correlation between $LD_{[1,9]}$ averaged across participants and $PC_{[1,9]}$ averaged across participants was computed to examine whether local darkness is correlated with pupil diameter change ($n = 9$ stimuli).
- *Replication of Study 4* (Hess, 1975b)*: Western* (*Experiment 2*). A repeated-measures ANOVA of pupil diameter change was conducted, with time (pupil diameter at 16.5, 57.0, and 73.4 s) as within-subject factor and gender as a between-subjects factor. Also, the correlation between PC_t, on the one hand, and global darkness GD_t and local darkness LD_t, on the other, was computed at the level of video frames ($n = 1877$).
- *Replication of Study 5* (Polt and Hess, 1968)*: Visually presented words* (*Experiment 2*). As mentioned above, for the two-way repeated-measures ANOVA of $PC_{[1,9]}$, the pupil diameter was averaged between two words per category. This averaging might have masked word-specific effects such as those reported by Polt and Hess (1968) for the words 'flay' and 'nude'. Accordingly, a one-way repeated-measures ANOVA of $PC_{[1,9]}$ with the 12 words as a factor was conducted, and pairs of stimuli were statistically compared using the paired *t*-tests with Bonferroni correction (correction factor = 66).

3. Results

3.1. Replication of Study 1 (Hess and Polt, 1960): Images of five themes (Experiment 1)

3.1.1. Analyses examining whether Hess's results replicate

Fig. 3 shows the PC_t of the participants as a function of viewing time during the control slide and subsequent stimulus slide for the ten images. The pupil constricted from 0.5 s to 1 s after the stimulus onset for each of the ten images. This constriction was image-specific, ranging between about 10% for the 'Male' images and 5% for the 'Mother and baby' images.

Table 3 shows the means and standard deviations of $PC_{[1,9]}$ for

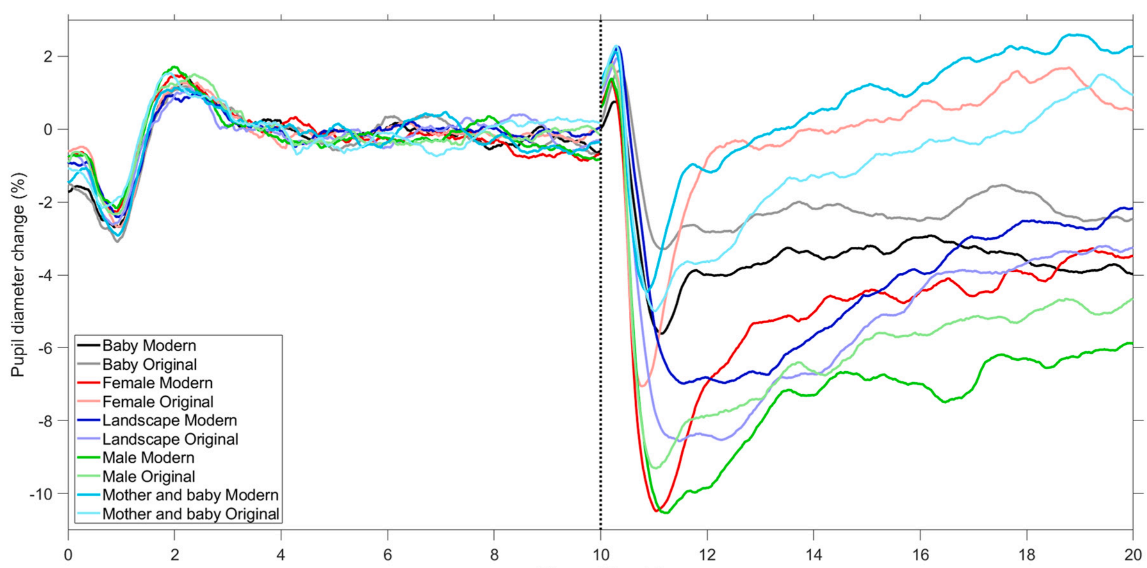


Fig. 3. Mean pupil diameter change (PC_t) with respect to the preceding control slide, for the images of five themes in Experiment 1. The dotted vertical line indicates the moment of transition from the control slide to the stimulus slide. A positive value indicates pupil dilation; a negative value indicates pupil constriction. Note that the small jump in pupil diameter at 10 s is because participants performed a 'drift correction' between the control slide and the stimulus slide.

Table 3
Means (standard deviations in parentheses) of pupil diameter change ($PC_{[1,9]}$, %) for female and male participants, and results of independent-samples *t*-tests, for the images of five themes in Experiment 1.

Stimulus	Females	Males	*t*(180)	Cohen's *d*	*p*
Baby – Modern	−4.33 (5.78)	−3.31 (4.45)	−1.28	−0.21	.202
Baby – Original	−3.32 (4.68)	−1.88 (5.07)	−1.78	−0.29	.077
Female – Modern	−7.05 (5.83)	−4.48 (5.96)	−2.65	−0.43	**.009**
Female – Original	−0.81 (5.15)	0.31 (5.49)	−1.27	−0.21	.205
Landscape – Modern	−5.11 (5.28)	−4.59 (6.34)	−0.52	−0.09	.601
Landscape – Original	−6.26 (5.81)	−5.54 (5.61)	−0.78	−0.13	.438
Male – Modern	−7.17 (4.99)	−7.69 (5.87)	0.57	0.09	.569
Male – Original	−6.06 (5.60)	−6.38 (5.54)	0.36	0.06	.723
Mother and baby – Modern	−0.83 (5.26)	1.29 (5.98)	−2.25	−0.37	**.026**
Mother and baby – Original	−1.96 (6.54)	−1.02 (5.45)	−0.99	−0.16	.321

Note. A positive value indicates pupil dilation; a negative value indicates pupil constriction. Statistically significant *p*-values are indicated in boldface.

female and male participants, together with the results of independent-samples *t*-tests per image. For the 'Female – Modern' image, the results were in agreement with Hess and Polt (1960), with males having a significantly higher $PC_{[1,9]}$ (a less negative value, indicating a smaller constriction) than females. Note that participants on average exhibited pupil constriction, as indicated by the negative $PC_{[1,9]}$ values. In other words, for the 'Female – Modern' image, females had a larger constriction of pupil diameter from the control image to the stimulus image than males. Table 3 also shows a significant difference between male and female participants for the 'Mother and baby – Modern' image, but the direction of this effect was opposite to Hess and Polt. No significant differences between males and females were observed for the other eight images.

3.1.2. Additional analyses

A two-way repeated-measures ANOVA of $PC_{[1,9]}$ with image age (original vs. modern) and image theme as within-subject factors showed a significant difference between original and modern images, $F(1,181) = 19.8$, $p < 0.001$, $\eta_p^2 = 0.10$, and between image themes, $F(4,724) = 96.6$, $p < 0.001$, $\eta_p^2 = 0.35$, as well as a significant 'image age' × 'image theme' interaction, $F(4,724) = 36.7$, $p < 0.001$, $\eta_p^2 = 0.17$. Pairwise comparisons showed that the $PC_{[1,9]}$ of 33 of the 45 pairs of images differed significantly from each other.

To understand these image-specific effects in pupil dilation, we computed local darkness LD_t at each sampling instant (Fig. 4). There were substantial differences in local darkness between images, even though all images had the same global darkness of 50% (see Appendix H). A two-way repeated-measures ANOVA of LD_0 with image age (original vs. modern) and image theme as within-subject factors showed a significant difference between original and modern images, $F(1,181) = 11.0$, $p = 0.001$, $\eta_p^2 = 0.06$, and between image themes, $F(4,724) = 1292$, $p < 0.001$, $\eta_p^2 = 0.88$, as well as a significant 'image age' × 'image theme' interaction, $F(4,724) = 198.6$, $p < 0.001$, $\eta_p^2 = 0.52$. Pairwise comparisons showed that the LD_0 of 41 of the 45 pairs of images differed significantly from each other. The strong effect size for image theme ($\eta_p^2 = 0.88$) indicates that local darkness is theme-specific. For example, the two 'Male' images yielded low LD_0 because participants initially looked at the male's body, which was bright, and not at the dark background. Fig. 5 shows a scatter plot of LD_0 averaged across participants versus PC_1 averaged across participants. The strong correlation (Pearson's $r = 0.89$, $p < 0.001$, $n = 10$ images) suggests that the initial pupil constriction was due to the luminance of the location where people looked when the side appeared.

Additionally, we inspected the heatmaps of the eye-gaze coordinates (see Appendix L). A result that stood out was that males were more likely than females to look at the breast of the female: For the 'Female – Modern' image, females looked on average 0.69 s ($SD = 0.60$ s) at the breast, whereas males looked at that area for 1.05 s ($SD = 0.83$ s). Similarly, for the 'Female – Original' image, females and males looked at the breast for 0.79 s ($SD = 0.75$ s) and 1.23 s ($SD = 0.90$ s), respectively.

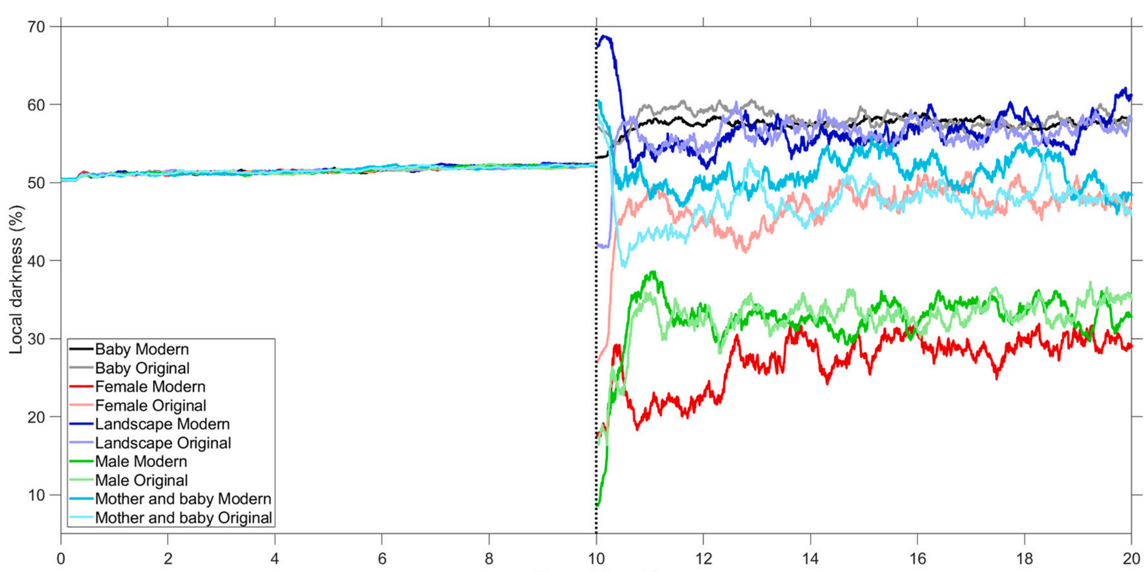


Fig. 4. Mean local darkness (LD_t) for the images of five themes in Experiment 1 and the preceding control slide. The dotted vertical line indicates the moment of transition from the control slide to the stimulus slide. The jump in local darkness occurring at 10 s is due to the appearance of the image, which resulted in a change of local darkness.

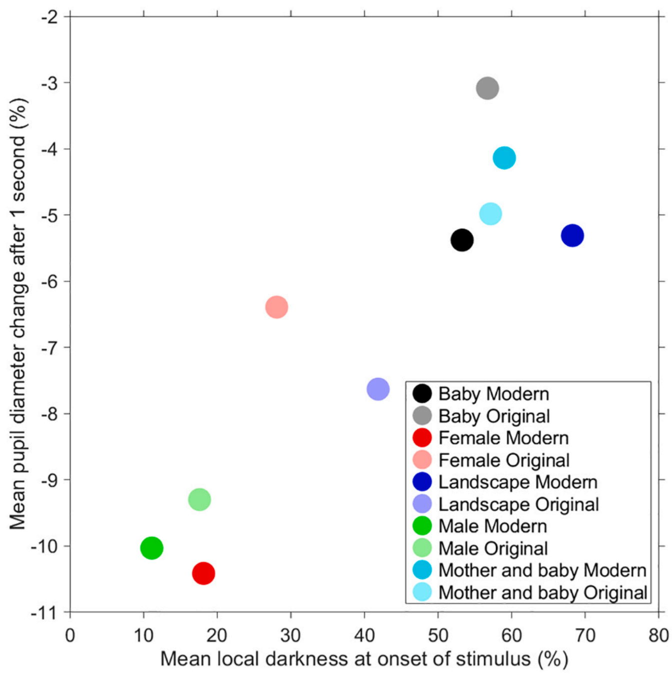


Fig. 5. Local darkness (LD_0) averaged across participants versus pupil diameter change PC_1 averaged across participants, for the images of five themes in Experiment 1.

Cohen's *d* effect sizes between females and males were −0.46 and −0.52 for the 'Female – Modern' and 'Female – Original' images. The differences between males and females were significant, 'Female – Modern': $t(180) = -2.83$, $p = 0.005$, 'Female – Original': $t(180) = -3.17$, $p = 0.002$.

Finally, we compared whether $LD_{[1,9]}$ for the ten images was significantly different between male and female participants (Appendix L). Two statistically significant differences were found, for 'Landscape – Original' and 'Male – Original', with females looking on average at, respectively, lighter and darker areas than males. The same images were not associated, however, with statistically significant gender differences in $PC_{[1,9]}$ (Table 3). Moreover, the two images for which statistically significant gender differences in $PC_{[1,9]}$ were found did not yield significant differences in $LD_{[1,9]}$. In other words, the gender differences in $PC_{[1,9]}$ could not be explained by gender differences in $LD_{[1,9]}$.

3.2. Replication of Study 1 (Hess and Polt, 1960): Images of five themes (Experiment 2)

3.2.1. Analyses examining whether Hess's results replicate

Fig. 6 shows the mean pupil diameter change (PC_t) of participants as a function of elapsed time for the ten line drawings of Experiment 2. Similar to Experiment 1, the pupillary responses showed congruence of the two stimuli of the same theme. Drawings of nude males and females yielded the largest pupil dilation (Table 4). Independent-samples *t*-tests showed no statistically significant gender differences in $PC_{[1,9]}$ ($p > 0.05$ for each of the five tests; Table 4).

3.2.2. Additional analyses

A repeated-measures ANOVA with image theme as a within-subject factor showed a significant difference between the $PC_{[1,9]}$ of the five image themes, $F(4,584) = 70.4$, $p < 0.001$, $\eta_p^2 = 0.33$. Pairwise comparisons showed that the 'Female' and 'Male' line drawings did not significantly differ from each other but yielded significantly larger $PC_{[1,9]}$ than the 'Baby', 'Landscape', and 'Mother and baby' line drawings, which in turn did not significantly differ from each other.

Similar to Experiment 1, the heatmaps of the eye-gaze coordinates showed that males were more likely than females to look at the breast of the nude female (Appendix L). On average, females and males looked at the breast in the 'Female 1' drawing for 1.51 s ($SD = 0.91$ s) and 2.09 s ($SD = 1.14$ s), respectively (Cohen's *d* between females and males =

Table 4
Means (standard deviations in parentheses) of pupil diameter change ($PC_{[1,9]}$, %) for female and male participants, and results of independent-samples *t*-tests, for the line drawings of five themes in Experiment 2.

Stimulus	Females	Males	t(145)	Cohen's d	p
Baby	−0.80 (3.97)	0.58 (4.86)	−1.67	−0.30	.098
Female	4.73 (4.64)	6.20 (5.35)	−1.60	−0.29	.112
Landscape	0.89 (4.94)	−0.76 (4.90)	1.87	0.34	.063
Male	4.95 (4.62)	4.77 (6.36)	0.17	0.03	.866
Mother and baby	0.62 (4.84)	0.77 (4.49)	−0.18	−0.03	.857

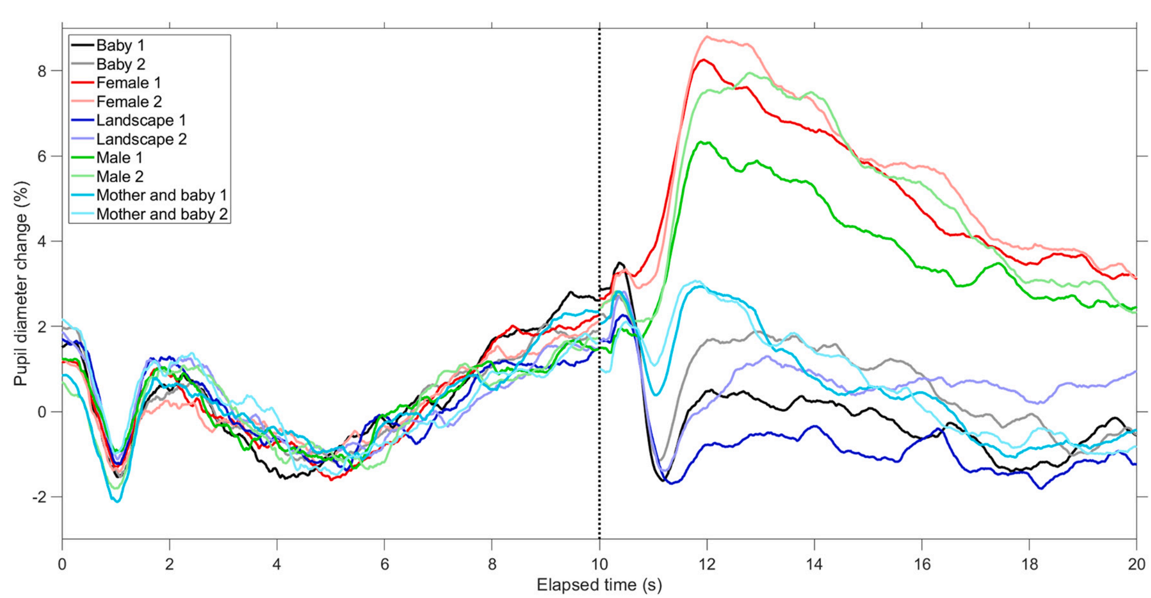


Fig. 6. Mean pupil diameter change (PC_t) for the line drawings of the five themes in Experiment 2 with respect to the preceding control slide. The dotted vertical line indicates the moment of transition from the control slide to the stimulus slide.

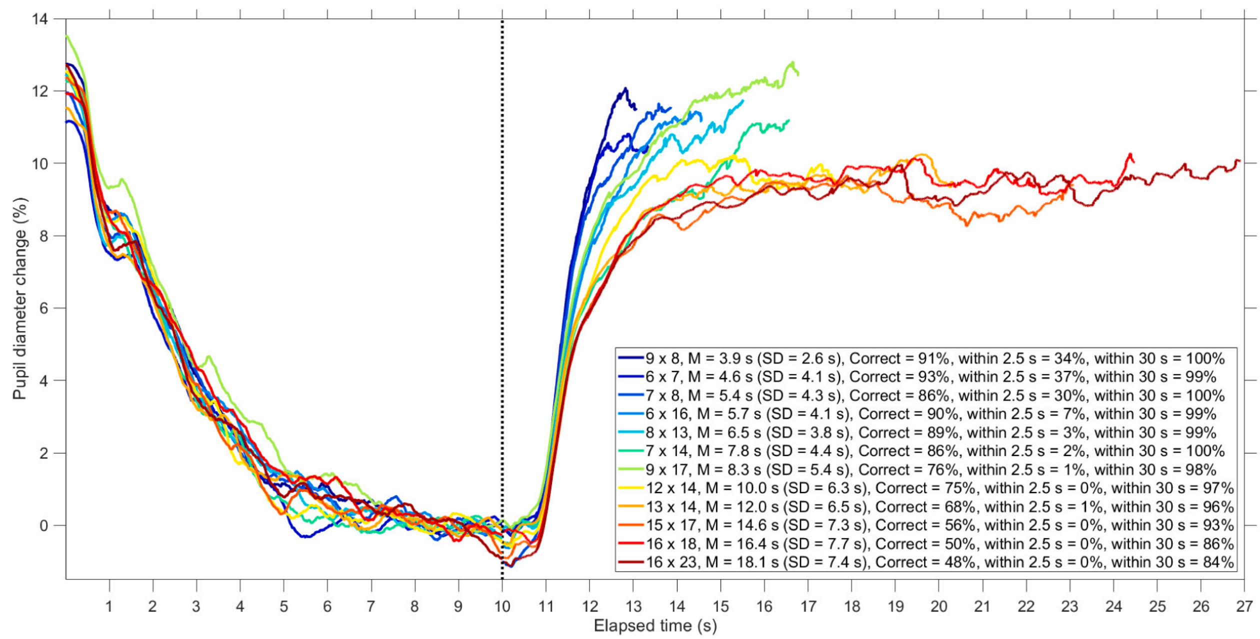


Fig. 7. Mean pupil diameter change (PC_t) for the multiplications in Experiment 1 with respect to the preceding control slide. A positive value indicates pupil dilation; a negative value indicates pupil constriction. Because the trial ended once the participant pressed the spacebar, the sample size decreases with elapsed time. Means are shown up to the point where data for at least 91 of the 182 participants were available. The legend shows the means and standard deviations of the answering times (i.e., elapsed time of pressing the spacebar since the onset of the multiplication), the percentage of participants who provided the correct answer, the percentage of participants who answered within 2.5 s, and the percentage of participants who answered within the time limit of 30 s. The dotted vertical line indicates the moment of transition from the control slide to the stimulus slide.

−0.54, $t(145) = 3.01$, $p = 0.003$). For the 'Female 2' drawing, female and male participants looked at the breast 1.89 s ($SD = 0.94$ s) and 2.53 s ($SD = 1.28$ s), respectively (Cohen's d between females and males = −0.54, $t(145) = 3.04$, $p = 0.003$).

3.3. Replication of Study 2 (Hess and Polt, 1964): Multiplications (Experiment 1)

3.3.1. Analyses examining whether Hess's results replicate

Fig. 7 shows the mean PC_t as a function of the elapsed time for the 12 multiplications. During the control slide, the pupil diameter gradually recovered from the previous multiplication. Strong dilations of about 10% occurred while participants were performing the multiplications. It is worth noting that the mean PC_t rose to higher values for the easier multiplications.

Table 5 provides the results for the twelve multiplications. $PC_{[ans\text{-}2.5,ans]}$, the measure used by Hess and Polt (1964), was lower for easier calculations, consistent with Hess and Polt.

3.3.2. Additional analyses

Next to $PC_{[ans\text{-}2.5,ans]}$, Table 5 shows the results for the twelve multiplications for PC_{max}, PC_{ans}, and PC_3, and Fig. 8 shows the trends that the four pupil change measures follow as a function of the average time it took to solve each multiplication. It can be seen that the direction of the effect between difficulty and pupil diameter change depends on the measure. For easy calculations, the 2.5-s period often included the period before dilation (i.e., 10–11 s in Fig. 7), leading to an (artificially) low $PC_{[ans\text{-}2.5,ans]}$ value. PC_{max} was also lower for easier calculations.

Table 5
Means (standard deviations and sample sizes in parentheses) for four measures of pupil diameter change (%), and results of tests of within-subject linear contrasts, for the multiplications in Experiment 1.

Multiplication	Hess and Polt (1964)	Replication study			
	$PC_{[ans\text{-}2.5,ans]}$	$PC_{[ans\text{-}2.5,ans]}$	PC_{max}	PC_{ans}	PC_3
9 × 8		8.00 (7.76, 180)	14.42 (7.96, 180)	11.61 (8.20, 180)	11.63 (7.84, 97)
6 × 7		7.95 (7.49, 182)	14.07 (7.95, 182)	10.91 (8.24, 182)	10.37 (6.71, 101)
7 × 8	10.8	9.14 (7.52, 179)	15.44 (7.57, 179)	12.50 (7.35, 179)	10.67 (6.85, 115)
6 × 16		8.94 (6.98, 180)	15.15 (7.61, 180)	11.82 (7.93, 180)	10.13 (7.13, 150)
8 × 13	11.3	10.49 (7.62, 182)	16.02 (8.20, 182)	12.17 (7.94, 182)	9.34 (7.42, 168)
7 × 14		10.30 (6.51, 182)	16.02 (7.18, 182)	11.78 (7.89, 182)	8.11 (6.28, 174)
9 × 17		12.88 (7.72, 181)	18.47 (8.27, 181)	14.15 (8.00, 181)	9.71 (7.25, 179)
12 × 14		11.55 (7.83, 181)	16.99 (8.09, 181)	12.86 (8.47, 181)	8.93 (6.51, 179)
13 × 14	18.3	11.09 (8.05, 182)	17.34 (8.25, 182)	11.97 (8.66, 182)	8.18 (6.95, 180)
15 × 17		11.64 (8.16, 181)	18.10 (8.10, 181)	12.11 (8.66, 181)	7.69 (6.31, 181)
16 × 18		12.80 (7.83, 181)	19.21 (7.87, 181)	13.31 (8.55, 181)	8.14 (6.72, 180)
16 × 23	21.6	12.52 (8.83, 182)	19.19 (9.01, 182)	13.18 (9.97, 182)	7.84 (6.84, 182)
Tests of within-subject contrasts (7 × 8, 8 × 13, 13 × 14, 16 × 23)		$F(1,178) = 15.5$, $p < 0.001$, $\eta_p^2 = 0.08$	$F(1,178) = 24.0$, $p < 0.001$, $\eta_p^2 = 0.12$	$F(1,178) = 0.29$, $p = 0.588$, $\eta_p^2 = 0.00$	$F(1,109) = 19.3$, $p < 0.001$, $\eta_p^2 = 0.15$
Tests of within-subject contrasts (all 12 multiplications)		$F(1,176) = 68.7$, $p < 0.001$, $\eta_p^2 = 0.28$	$F(1,176) = 83.2$, $p < 0.001$, $\eta_p^2 = 0.32$	$F(1,176) = 6.77$, $p = 0.010$, $\eta_p^2 = 0.04$	$F(1,64) = 20.2$, $p < 0.001$, $\eta_p^2 = 0.24$

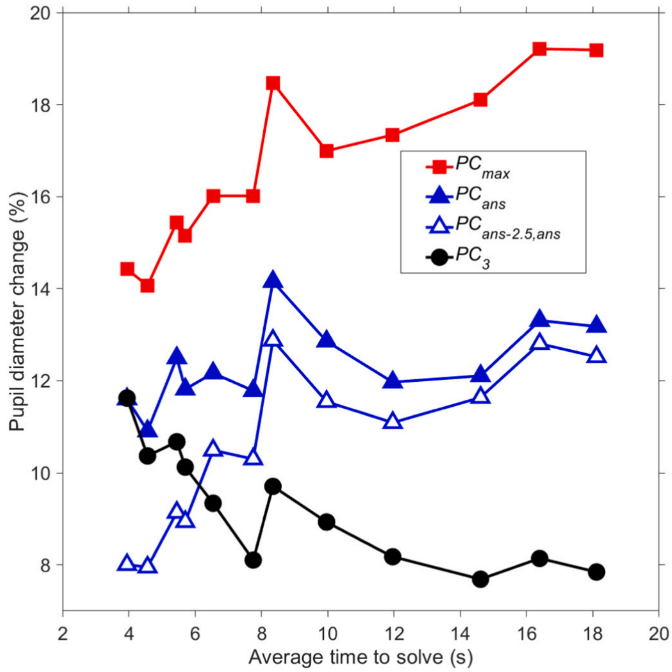


Fig. 8. Pupil diameter change for four measures as a function of the average time to solve the multiplication in Experiment 1.

However, the pupil diameter change at a fixed moment of 3 s after the presentation of the multiplication problem (PC_3) was *larger* for the easier multiplications. Appendix M provides corroborating results for the 65 participants with complete data at 3 s.

3.4. Replication of Study 3 (Hess, 1975a): Schematic eyes (Experiment 2)

3.4.1. Analyses examining whether Hess's results replicate

Fig. 9 shows the mean pupil diameter change (PC_t) of participants as a function of elapsed time, and Table 6 shows the mean and *SD* of $PC_{[1,9]}$ for the nine schematic eyes. It can be seen that the larger the depicted pupil, the larger the participants' $PC_{[1,9]}$. One-way repeated-measures ANOVAs showed that the effect was significant only for the one-eyed stimuli, with one-eyed stimuli: $F(2,292) = 7.87$, $p < 0.001$, $\eta_p^2 = 0.05$; two-eyed stimuli: $F(2,292) = 0.81$, $p = 0.446$, $\eta_p^2 = 0.01$; and three-eyed stimuli: $F(2,292) = 2.15$, $p = 0.118$, $\eta_p^2 = 0.01$. These findings are not consistent with Hess (1975a), who reported that dilations occurred for the two-eyed stimuli only.

3.4.2. Additional analyses

We performed a two-way repeated-measures ANOVA of $PC_{[1,9]}$ with the number of schematic eyes and the depicted pupil sizes as within-subject factors. Results showed a significant effect of the number of schematic eyes, $F(2,292) = 11.5$, $p < 0.001$, $\eta_p^2 = 0.07$ and of depicted pupil size, $F(2, 292) = 8.83$, $p < 0.001$, $\eta_p^2 = 0.06$. There was no significant 'number of eyes' × 'depicted pupil size' interaction, $F(4, 584) = 1.00$, $p = 0.408$, $\eta_p^2 = 0.01$.

We calculated the number of saccades while participants were viewing the schematic eyes. The mean (*SD*) number of saccades was 2.30 (2.54) for one-eyed stimuli, 11.50 (4.81) for two-eyed stimuli, and 11.64 (5.46) for three-eyed stimuli. These results are explained by the fact that when the slide depicted two or three eyes, participants glanced back and forth between those eyes; when the slide depicted one eye, participants showed little eye movement (see the Supplementary Material for a video showing the eye movements). A two-way repeated-measures ANOVA of the number of saccades showed a significant effect of the number of schematic eyes, $F(2,292) = 406.3$, $p < 0.001$, $\eta_p^2 = 0.74$, but not of depicted pupil size, $F(2, 292) = 1.59$, $p = 0.205$, $\eta_p^2 = 0.01$. There was no significant 'number of eyes' × 'depicted pupil size' interaction, $F(4, 584) = 0.74$, $p = 0.567$, $\eta_p^2 = 0.01$.

Fig. 10 shows that $LD_{[1,9]}$ was highest for the stimulus with one eye and a large pupil. This finding can again be explained by the fact that, when there was only one eye, this was where participants looked. Fig. 11 shows a scatter plot of $LD_{[1,9]}$ averaged across participants and $PC_{[1,9]}$

Table 6
Means (standard deviations in parentheses) of pupil diameter change ($PC_{[1,9]}$, %), for the schematic eyes in Experiment 2 ($N = 147$).

Stimulus	Small pupils	Medium pupils	Large pupils
1 eye	0.54 (5.58)	1.74 (5.94)	2.74 (5.91)
2 eyes	0.51 (5.93)	0.93 (5.66)	1.23 (5.88)
3 eyes	−0.28 (5.34)	0.40 (5.63)	0.78 (5.32)

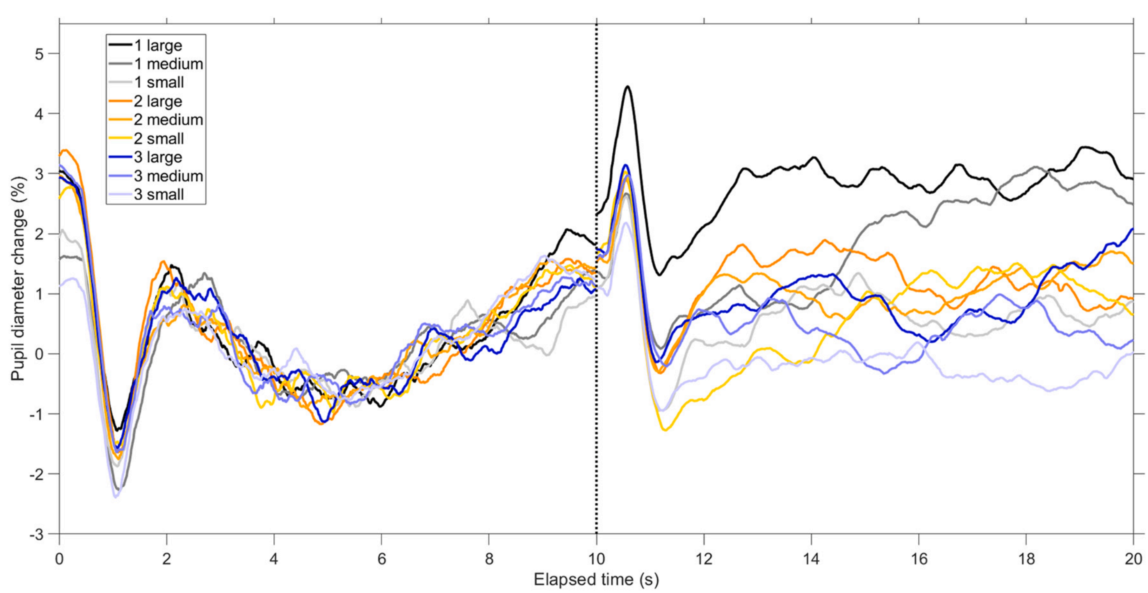


Fig. 9. Mean pupil diameter change (PC_t) for the schematic eyes in Experiment 2 with respect to the preceding control slide. A positive value indicates pupil dilation; a negative value indicates pupil constriction. The dotted vertical line indicates the moment of transition from the control slide to the stimulus slide.

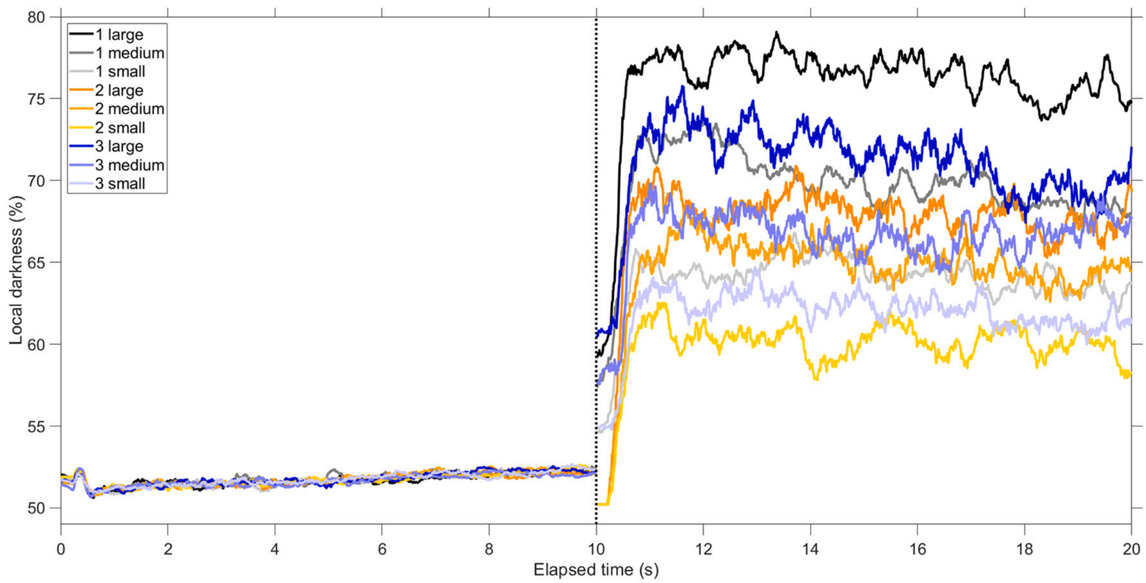


Fig. 10. Mean local darkness (LD_t) for the schematic eyes in Experiment 2 and the preceding control slide. The dotted vertical line indicates the moment of transition from the control slide to the stimulus slide.

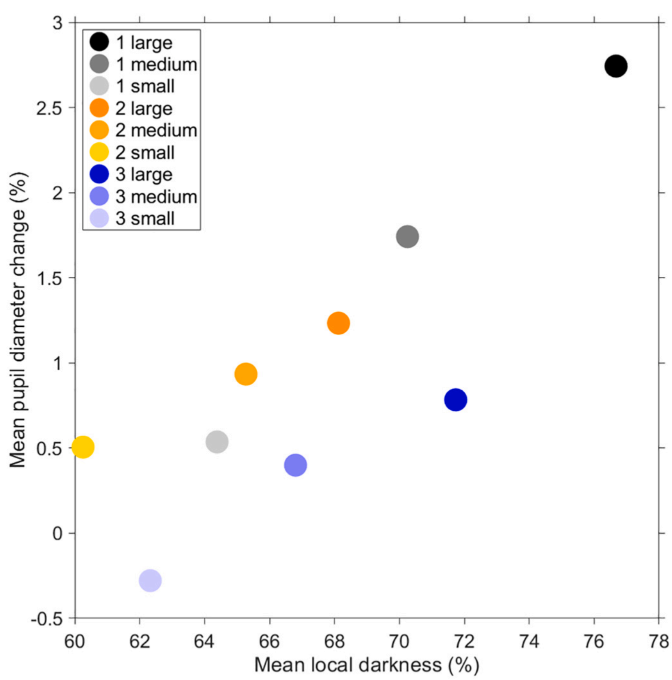


Fig. 11. Mean local darkness $LD_{[1,9]}$ averaged across participants versus mean pupil diameter change $PC_{[1,9]}$ averaged across participants, for the nine schematic eyes in Experiment 2.

averaged across participants. The strong correlation (r = 0.83, p = 0.006, n = 9 stimuli) indicates that local darkness is predictive of pupil diameter change.

3.5. Replication of Study 4 (Hess, 1975b): Western (Experiment 2)

3.5.1. Analyses examining whether Hess's results replicate

Our preregistration stated that support for Hess's hypothesis "*will be obtained if (1) d increases between 16.5 s and 57.0 s (i.e., while the man tries to escape), and (2) d decreases between 57.0 s and 73.4 s (i.e., man is caught and subdued till the moment when the scene starts fading)*". Here, 16.5 s is the moment the man is recognized, 57.0 s is the moment he is pulled off the horse, and 73.4 s is the when the scene starts fading out. Fig. 12 shows the mean PC_t and the d between the PC_t of male and female participants while watching the Western video clip. Consistent with Hess, we found that d increased between 16.5 s and 57.0 s from 0.03% to 1.21% and decreased between 57.0 s to 73.4 s from 1.21% to −1.99%. However, the increase and decrease were not as gradual as in Hess's data (Fig. 1). For example, around 40 s, d was −3.6%, which is inconsistent with Fig. 1.

3.5.2. Additional analyses

A repeated-measures ANOVA of pupil diameter change, with time (pupil diameter at 16.5, 57.0, and 73.4 s) as within-subject factor and gender as between-subjects factor showed a significant effect of time, $F(2, 290) = 65.7$, $p < 0.001$, $\eta_p^2 = 0.31$, but no significant effect of gender, $F(1, 145) = 0.03$, $p = 0.864$, $\eta_p^2 = 0.00$, and no significant time × gender interaction, $F(2, 290) = 2.25$, $p = 0.107$, $\eta_p^2 = 0.02$.

Fig. 13 shows that there are strong fluctuations in PC_t. We attempted to understand these fluctuations by examining the correlations with darkness levels. Fig. 13 shows the pupil diameter change PC_t together with the local darkness LD_t and global darkness GD_t as a function of elapsed time. There was moderate congruence between LD_t and mean PC_t of the video frames ($r = 0.40$, $n = 1877$). The correlation between global darkness GD_t and mean PC_t was of similar magnitude, $r = 0.48$ ($n = 1877$). In other words, the observed pupil diameter can be explained, in part, by the darkness of the video frame.

3.6. Replication of Study 5 (Polt and Hess, 1968): Visually presented words (Experiment 2)

3.6.1. Analyses examining whether Hess's results replicate

Fig. 14 shows the mean PC_t for the 12 words, whereas Table 7 shows the mean $PC_{[1,9]}$ values. A two-way repeated-measures ANOVA of $PC_{[1,9]}$ showed no significant effect of valence, $F(1,146) = 0.14$, $p = 0.713$, $\eta_p^2 = 0.00$, a significant effect of arousal, $F(1,146) = 5.27$, $p = 0.023$, $\eta_p^2 = 0.03$, and no significant valence × arousal interaction, $F(1,146) = 1.88$, $p = 0.172$, $\eta_p^2 = 0.01$. We also performed a one-way repeated-measures ANOVA with the four words of Polt and Hess (1968) as a within-subject factor, showing a significant effect, $F(3,438) = 7.48$, $p < 0.001$, $\eta_p^2 = 0.05$.

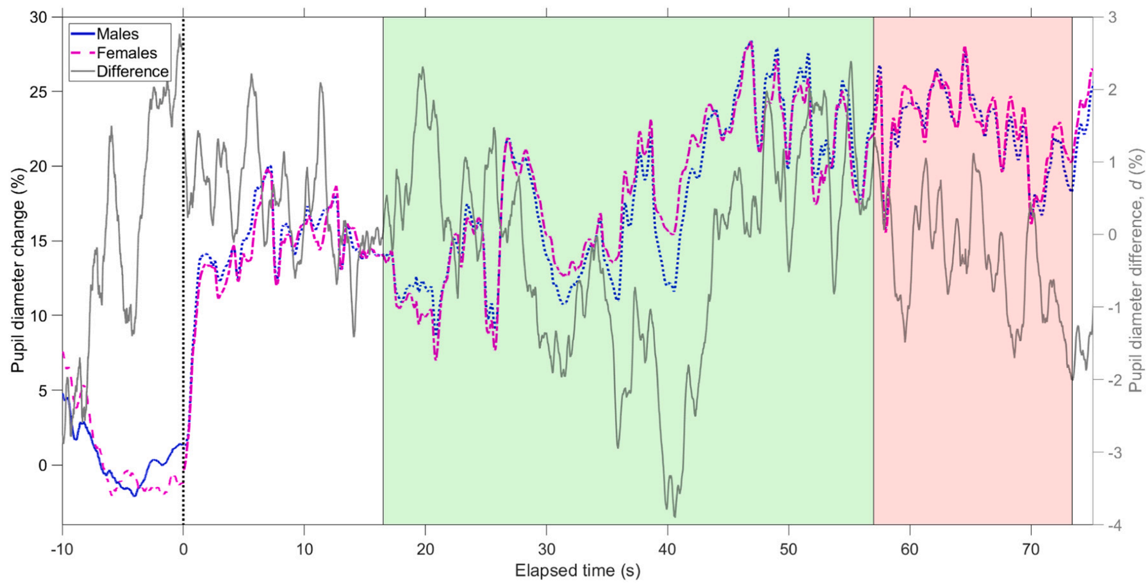


Fig. 12. Mean pupil diameter change (PC_t) of males and females for the Western in Experiment 2 with respect to the preceding control slide. A positive value indicates pupil dilation; a negative value indicates pupil constriction. Also shown is the difference (d) between the mean pupil diameter change (PC_t) of male and female participants. The dotted vertical line indicates the moment of transition from the control slide to the stimulus slide. The green and red backgrounds represent the periods the hero of the series tried to escape and was caught, respectively.

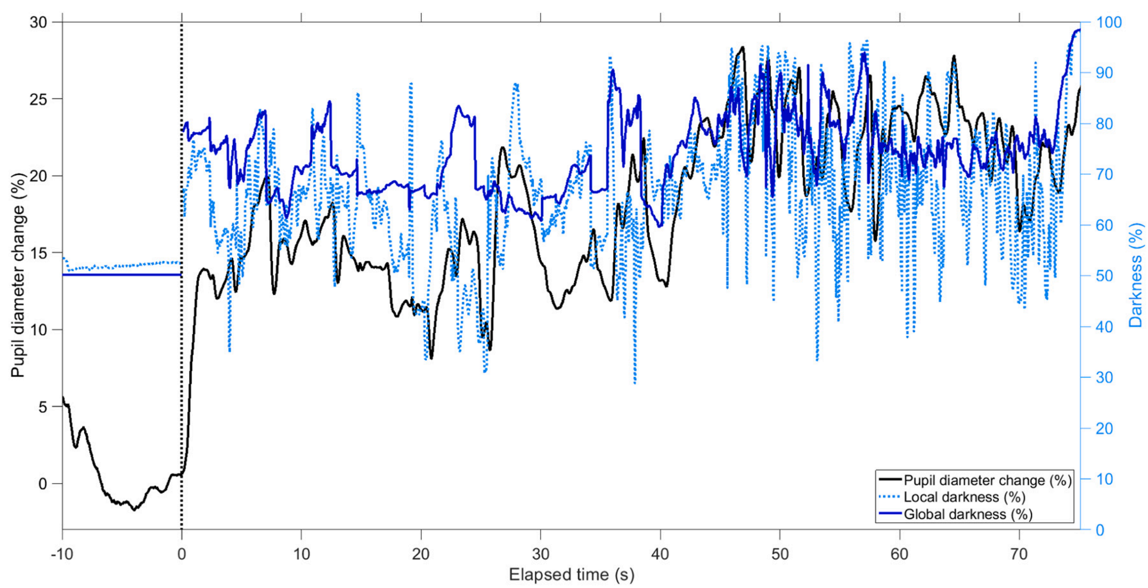


Fig. 13. Mean pupil diameter change (PC_t) for the Western in Experiment 2 with respect to the preceding control slide, local darkness (LD_t), and global darkness. The dotted vertical line indicates the moment of transition from the control slide to the stimulus slide.

3.6.2. Additional analyses

In the above repeated-measures ANOVAs, the $PC_{[1,9]}$ was averaged between the two words per category, as documented in the preregistration. This averaging may have masked word-specific effects on pupil diameter. As an additional non-preregistered test, we performed a one-way repeated-measures ANOVA with all 12 words as a factor to investigate word-specific effects, which might have been masked by averaging across words. Results showed a significant effect, $F(11,1606) = 5.55$, $p < 0.001$, $\eta_p^2 = 0.04$. Pairwise comparisons showed that 'nude' yielded a significantly larger $PC_{[1,9]}$ than all other words, except 'flay' and 'flirt'. Furthermore, 'flirt' yielded a significantly larger $PC_{[1,9]}$ than 'fragment', 'harmonica', 'hostile', and 'party'. The other word pairs were not statistically significantly different from each other.

4. Discussion

We replicated five studies of Eckhard Hess using a combined total of 329 participants. Hess used a slide projector for presenting the stimuli, whereas we used a computer monitor. Furthermore, Hess recorded pupil diameter twice a second, which may not be sufficient for capturing rapid reflexive responses. We captured eye movements and pupil diameter at a high frequency of 2000 Hz.

4.1. Luminance control and other validity threats in Hess's research

Our findings indicate that luminance has a strong effect on the results. A slide change of the projector used by Hess induces a 1-s period of increased darkness and corresponding pupil dilation. Moreover,

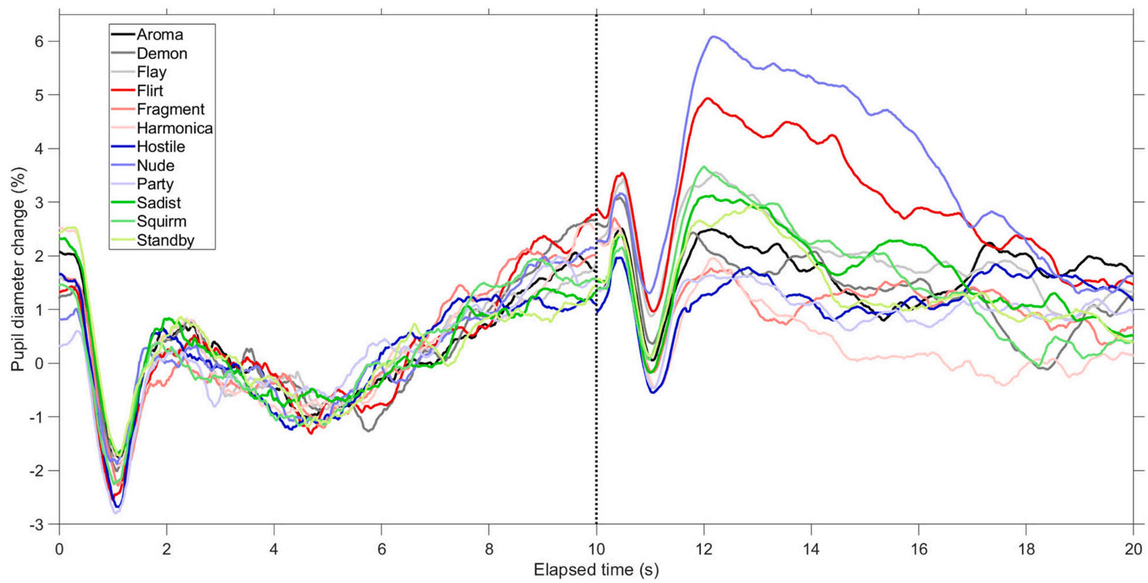


Fig. 14. Mean pupil diameter change (PC_t) for the words in Experiment 2 with respect to the preceding control slide. A positive value indicates pupil dilation; a negative value indicates pupil constriction. The dotted vertical line indicates the moment of transition from the control slide to the stimulus slide.

Table 7
Means (standard deviations in parentheses) of pupil diameter change ($PC_{[1,9]}$, %) for the words in Experiment 2. The pupil diameter changes for the four valence/arousal categories (average of two words per category) are shown in bold.

Stimulus	$PC_{[1,9]}$
Polt and Hess (1968)	
Flay	2.10 (6.12)
Hostile	1.23 (5.53)
Nude	4.02 (6.38)
Squirm	1.75 (6.25)
High valence-high arousal	**2.12 (4.56)**
Flirt	3.21 (5.77)
Party	1.03 (5.73)
Low valence-high arousal	**1.60 (4.06)**
Sadist	1.89 (5.20)
Demon	1.31 (5.33)
High valence-low arousal	**1.03 (4.16)**
Aroma	1.66 (5.59)
Harmonica	0.40 (5.24)
Low valence-low arousal	**1.34 (4.66)**
Fragment	1.16 (5.37)
Standby	1.53 (6.29)

Experiment 1 showed a strong reflexive constriction upon presenting a new image, a finding that is consistent with other pupillometry literature (e.g., Aboyoun and Dabbs, 1998; Bradley and Lang, 2015; Bradley et al., 2017; Snowden et al., 2019) and which we could attribute in part to 'local darkness', defined as the mean grayscale level of the point on the screen where participants looked. Using a method similar to ours, Bradley et al. (2017) found that local darkness had only a small influence on pupil diameter. Our analyses showed strong correlations between local darkness and pupil diameter change, possibly because we used a more accurate eye-tracker than Bradley et al. We further showed that presenting a pure white or black background causes mean constrictions and dilations in pupil diameter as large as 30% (Appendix J). Collectively, our findings suggest that it is essential to control for local darkness, such as by employing line drawings. Although the gender differences in pupil dilation were not explained by local darkness, it seems likely that the schematic eyes results (Study 3) are attributed to luminance effects, as discussed below.

In addition to the suboptimal equipment used by Hess, other issues were that the material we retrieved from the archive revealed that Hess and Polt selectively presented their results (Appendices I, N), that there appeared to be only quick or no peer review (Appendix R), and that Hess's research involved a conflict of interest as he had ties to a marketing company (Hess, 1975b; Krugman, 2013; Krugman, 1964a, 1964b; Rice, 1974; Sponsor, 1964; Van Bortel, 1968; West, 1962; see Appendix S, for details).

Nuijten et al. (2018) explained that "*if a result cannot be successfully reproduced, the original result is not reliable … raising the question of why one would invest additional resources in any replication*". Based on raw pupil diameter data retrieved from the archive, we could reproduce the gender differences reported in Hess and Polt (1960), although not perfectly so (Appendix N). Based on the previous observations, we concluded that Hess's works are valuable for their ideas and hypotheses but not for the empirical results. Accordingly, we decided to deviate from a direct replication by employing modern means of luminance control and extra stimuli.

The results of the replications of the five studies of Hess are summarized as follows:

4.2. *Replication of Study 1 (Hess and Polt, 1960): Images of five themes*

Hess and Polt (1960) found gender differences in pupil dilation depending on the image theme. We found that line drawings of nude females and nude males evoked a pupil dilation compared to neutral images. However, we did not obtain support for the hypothesis of gender differences in pupil dilation: Out of the 15 significance tests performed in Experiments 1 and 2, two showed a statistically significant gender difference: one with a direction consistent with and the other opposite to Hess and Polt. These gender differences in pupil diameter change were not explained by gender differences in viewing behavior and local darkness.

Recent large-sample studies have yielded a mixed picture about gender differences in pupil dilation in response to sexually arousing images: Some have found that pupil dilation is consistent with the sexual orientation of the participant (Attard-Johnson and Bindemann, 2017; Attard-Johnson et al., 2017; Finke et al., 2017; Rieger et al., 2015; Watts et al., 2017), whereas others have not found such an effect (Aboyoun and Dabbs, 1998; Scott et al., 1967; Snowden et al., 2019). Our results are in line with Snowden et al. (2019), who reported that the pupils dilate to sexual imagery but that the dilation does not relate to a person's gender. It remains to be investigated why these studies gave discrepant results. The degree of arousal may be an explanation: more substantial gender differences may be expected for more explicit material. On the other hand, Attard-Johnson and Bindemann (2017) reported that "*pupillary responses provide a sex-specific measure, but are not sensitive to sexually explicit content*". Context and instructions provided to participants could be another moderating factor. Snowden et al. (2019) suggested that it would be interesting to examine whether asking the participants to reflect on the sexual appeal of the images evokes a different pupillary response compared to passive viewing. It should be noted that participants did not necessarily get sexually aroused in our study. It is also possible that other mechanisms, such as embarrassment, nervousness, or the experimenter's style (e.g., Chapman et al., 1969), may have caused activation of the autonomic nervous system and hence pupil dilation.

In our study, we used images from the 1960s together with modern equivalents. Future replication research could examine the impact of social and cultural change on the pupillary response. Greenfield (2017) argued that sociodemographic and cultural change could explain why some findings might not replicate. For example, he presented a failed replication of gender differences in identifying sexual intent, which could be because "*female sexuality becomes more similar to masculine sexuality*" (p. 768).

Although gender differences in pupil diameter were small, we found substantial gender differences in viewing behavior, where males were more likely than females to look at the nude female's breast. These findings are consistent with Hewig et al. (2008), who used images of casually dressed male and female models and found that men gazed longer than women at the female breast area, and with Nummenmaa et al. (2012), who reported similar results for nude stimuli.

In summary, in our experiments, the pupils proved to be responsive to images of sexually arousing nature as compared to stimuli of other themes, but gender differences in pupil diameter change were not systematic.

4.3. Replication of Study 2 (Hess and Polt, 1964): Multiplications

Hess and Polt (1964) showed a positive association between the difficulty level of the multiplication and the degree of pupil dilation. We found that this relationship holds when examining the data in the way done by Hess and Polt: when assessing the maximum pupil dilation or the pupil dilation in the 2.5-s period before the participant provided an answer. However, these two indexes are biased because they depend on the length of the measurement period. That is, given the fluctuating nature of pupil diameter, the longer the calculation time, the higher the opportunity for reaching a high maximum pupil diameter, and the smaller the likelihood that the 2.5-s period includes the pupil diameter before dilation. When assessing the pupil diameter at a particular moment (i.e., 3 s after the multiplication presentation), the easier multiplications yielded a *larger* dilation. We conclude that the answer to the question of whether more difficult problems yield larger pupil diameter change is dependent on the measure that is used. In summary, our results indicate that easy calculations yield a burst of pupil dilation, followed by a recovery period. These findings resemble Van der Meer et al. (2010), who showed that high-IQ participants exhibited a shorter-lasting yet higher-amplitude pupil dilation than average-IQ participants. Van der Meer et al. argued that high-IQ individuals allocate more resources to the problem-solving task.

Our findings call for a reinterpretation of many pupillometry findings in the literature. For example, Ahern and Beatty (1979), Klingner et al. (2008), and Marquart and De Winter (2015) had participants solve multiplications within a fixed time budget and showed that more difficult multiplications yielded a larger dilation averaged over that time budget. These findings, which appear to run counter to our present observations, can be explained by the fact that more difficult problems take longer to solve, resulting in a longer period of dilation, not necessarily a larger dilation.

Solving easy (e.g., single-digit) multiplications involves retrieval from long-term memory, whereas solving a complex multiplication may involve additional processes such as decomposition of the problem into simpler ones/tens, retrieving the answers for the simple calculations from long-term memory, storing answers in short-term memory, and adding the partial results (Reys et al., 1995; Seitz and Schumann-Hengsteler, 2000; Tronsky, 2005). The short burst of pupil dilation for easy calculations could relate not only to a high amount of mental resources allocated to short-lasting tasks but also to retrieval effort from long-term memory and emotional arousal (e.g., the stress of meeting expectations, embarrassment if failing).

Our sample consisted of students at a technical university. Hess and Polt (1964) deemed their sample of "*above average in intelligence*", where "*one held a Ph.D. degree, two were at an advanced graduate level, one held a B.A. degree, and one was an undergraduate research assistant in the psychology department of this university*" (p. 1190). Hess and Polt did not report how long it took per participant to solve the multiplications, except that these times were "*anywhere from 3 to 30 seconds*" (p. 1191). It is possible that our sample of engineering students solved the multiplications faster than the participants in Hess and Polt. Future research is needed to examine the generalizability of the present findings to other samples. Asking the participants afterwards about the solution strategies they employed could be insightful regarding the type of mental processes employed and how these strategies associate with pupil response. Research has shown that skilled and unskilled mental calculators employ different strategies: unskilled calculators tend to follow strategies similar to those used for written right-to-left computation, whereas skilled ones use a variety of strategies, including recall of large products and summation of intermediate results into a single product (Hope and Sherrill, 1987).

4.4. Replication of Study 3 (Hess, 1975a): Schematic eyes

Hess (1975a) claimed that the pupils have an important role in communication, as the pupils of human observers respond to the pupil diameter of other people's eyes. Hess found this pupil mimicry effect when participants were presented with two schematic eyes, but not for images containing one or three schematic eyes. We found a statistically significant mimicry effect for one-eyed stimuli and not for two or three-eyed ones.

We further found that participants' eye movements were strongly dependent on how many schematic eyes were shown: When presented with only one schematic eye, participants stared at that eye, whereas for two of three schematic eyes, they scanned back and forth between the eyes. Hess and Goodwin (1974) argued that their findings could not be caused by the amount of darkness in the image: "*a hypothesis that pupil responses should be larger toward schematic eyespots with large 'pupils' because of the greater amount of dark area, particularly in the case of the triple eyespots, did not receive support*" (p. 219). Our analyses, however, indicate that 'local darkness', an index calculated based on where participants looked, provides a plausible explanation for our pupil diameter values. This observation is consistent with Derksen et al. (2018), who, based on several experiments with luminance-controlled and luminance-not-controlled stimuli of static and dynamic pupils of various sizes, concluded that the pupil mimicry phenomenon is due to luminance and participants' attention shift towards the eye region.

4.5. Replication of Study 4 (Hess, 1975b): Western

Hess (1975b) presented gender differences in pupil dilation for participants watching a specific scene from an episode of a Western TV series; he found that the pupils of males dilated more than those of females when the male hero of the series was trying to escape the attacking crowd. In our replication, we did not find this gender-specific pattern. Interestingly, Hess and Goodwin (1974) presented the same pupil diameter data as Hess (1975b), yet concluded that "*the men and women had essentially similar pupil responses*" (p. 213), which suggests that the gender differences in Hess (1975b) were presented selectively.

A limitation of our replication of the Western study is that the video we showed was of brief duration, whereas Hess (1975b) showed the full 30-min episode. Also, Hess and Goodwin (1974) applied a global luminance control technique using a photocell that scanned the film just before it entered the projector. The photocell determined the overall luminance of each film section and opened or shut down a lens diaphragm on the projector lens. We recommend further research using longer-lasting videos and luminance control to examine arousal and interest effects.

In summary, we did not find support for Hess's hypothesis of gender differences in pupil diameter when viewing a video of a male trying to escape an attacking crowd. What we did find is large fluctuations in pupil diameter while participants were watching the video. These fluctuations were consistent for males and females and could be explained, in part, by changes in local and global darkness. In future pupillometry studies, instead of using local and global darkness indexes, more precise predictors of the pupillary light reflex could be considered. We see potential in using a two-dimensional function that weights the screen luminance based on the pupillary sensitivity as a function of retinal eccentricity.

4.6. Replication of Study 5 (Polt and Hess, 1968): Visually presented words

Polt and Hess (1968) suggested that words rich in arousal, such as the word 'nude', cause pupil dilation, whereas threatening words cause pupil constriction. Our findings are consistent with this hypothesis: just as the line drawings of nudes caused pupil dilation, so did the presentation of the words 'nude' and 'flirt' cause a larger pupil dilation than neutral stimuli. These effects were found for words that may be regarded as sexually arousing, and not for the other words that were pre-registered as having high arousal scores ('hostile', 'squirm', 'party', 'sadist', 'demon').

Bayer et al. (2011) found that arousing words were associated with a slightly *smaller* pupil diameter than non-arousing words, which the authors attributed to arousing words being more easily recognized and therefore associated with a lower cognitive load. However, Bayer et al. did not distinguish between sexual and non-sexual words. Future research may be needed to determine what types of words of arousing nature evoke pupil dilation.

We found no significant difference in pupil response between words that scored high and low in valence, a finding that is consistent with Paivio and Simpson (1966), Peavler and McLaughlin (1967), and Siegle et al. (2001). Similarly, Henderson et al. (2018) investigated pupil response to brief scripts and found pupil dilation for both pleasant and unpleasant emotionally arousing scripts.

5. Conclusions

Table 8 provides an overview of the methods and results of the five studies of Hess and our replications. Overall, our replications confirm Hess's findings in that pupils dilate in response to mental demands (with

Table 8
Overview of the methods and results in the original studies of Hess and our replications.

		Our replication
Study 1: Hess and Polt (1960)		
Participants	4 males, 2 females	129 males, 53 females (Experiment 1) 102 males, 45 females (Experiment 2)
Stimuli	5 images of themes	10 images of themes (Experiment 1) 10 line drawings of themes (Experiment 2)
Results	"*a clear sexual dichotomy in regard to the interest value of the pictures, with no overlap between sexes*" (p. 350)	**Failure to replicate,** as only 2 of the 15 statistical tests showed a significant gender difference in pupil dilation (one consistent with and the other in the opposite direction to Hess and Polt, 1960). Line drawings of nudes caused pupil dilation. Additional analyses showed gender differences in viewing behavior. Furthermore, local darkness of images was predictive of participants' pupil dilation.
Study 2: Hess and Polt (1964)		
Participants	4 males, 1 female	129 males, 53 females
Stimuli	4 multiplications	12 multiplications
Results	Larger pupil dilation in the 2.5 s before providing an answer for difficult multiplications (e.g., 16 × 23) as compared to easy multiplications (e.g., 7 × 8)	**Successful replication.** However, a nuance is provided: we found larger dilation for easier multiplications and more prolonged dilation for more difficult multiplications.
Study 3: Hess (1975a)		
Participants	10 males, 10 females	102 males, 45 females
Stimuli	9 images of schematic eyes	9 images of schematic eyes
Results	Pupillary mimicry for two-eyed images, not for one-eyed and three-eyed images	**Failure to replicate,** as we found statistically significant pupillary mimicry for one-eyed images, not for two-eyed and three-eyed images. Pupillary mimicry could be explained by local darkness (i.e., pupillary light reflex).
Study 4: Hess (1975b)		
Participants	50 males, 50 females	102 males, 45 females
Stimuli	30-min episode of TV series	75-s video clip from the same episode
Results	Pupils of males dilated more than those of females when the male hero of the series was trying to escape the attacking crowd	**Different overall pattern** than that described in Hess (1975b): Sharply fluctuating pupil diameter of males and females, in part explained by global and local darkness.
Study 5: Polt and Hess (1968)		
Participants	9 males, 6 females	102 males, 45 females
Stimuli	4 words	12 words
Results	Some dilation for arousing words such as flay and nude	**Successful replication:** Arousing words caused greater dilation than non-arousing words. No significant effects for word valence. There were word-specific effects, with the word 'nude' causing dilation.

the nuance that easier multiplication yielded a shorter burst of stronger dilation) and stimuli (line drawings and words) of sexually arousing nature, whereas Hess's hypotheses regarding pupil mimicry and gender differences in pupil dilation did not replicate.

Finally, several methodological factors need to be discussed. First, in our experiment, while the stimuli within each study were presented in random order, the studies were presented in a fixed order. This approach seemed reasonable because each study involved separate hypotheses. For future research, the studies could be randomized.

Second, in all our analyses, we used the percentage pupil diameter change with respect to the preceding control slide as a dependent variable, consistent with our preregistration and all pupillometry works of Hess. Recent research has shown that a baseline correction in millimeters (i.e., subtractive correction) is physiologically more sensible than a percentage-difference baseline correction (i.e., divisive correction; Mathôt et al., 2018; Reilly et al., 2019). However, for our experiment, it hardly matters whether subtractive or divisive baseline correction is used (see Appendix O, showing correlations of 0.99 between the results of these two approaches).

Third, our control slide, the design of which was based on Hess (e.g., Hess, 1965), may have been suboptimal because it required eye movements. Because eye movements may affect (the measurement of) pupil size, it may have been better to use a control slide with a single crosshair instead.

Fourth, except for the multiplications where we applied a 2.5-s baseline period to allow recovery from the previous trial (cf. Fig. 7), we used an 8-s baseline period, in agreement with our preregistration of Hess's procedures. Other studies used considerably shorter baseline periods of 200 ms, 500 ms, or 1000 ms (e.g., Mathôt et al., 2018; often accompanied by relatively short inter-trial intervals), which may be beneficial for obtaining a baseline value that is not contaminated by long-term trends and carryover effects from the previous trial. On the other hand, short baseline periods may be problematic because pupil diameter shows strong variability (also called pupillary hippus, unrest, or 'noise', see Stark, 1959). Given the highly fluctuating nature of pupil diameter, a longer baseline period can be expected to cancel out noise better, resulting in higher statistical power and a more statistically reliable estimate of pupil diameter change, as illustrated through extra analyses in Appendix P. In summary, it seems that, provided that the mean pupil diameter has stabilized from the previous trial, longer baseline periods are preferred.

Fifth, viewing angle may interact with pupil diameter, a problem known as the pupil foreshortening effect (Hayes and Petrov, 2016). Analysis of the foreshortening effect can be found in Appendix Q. In the present study, we did not correct the pupil diameter for viewing angle, because our stimuli were presented relatively centrally on the screen.

Supplementary Material

Raw data, scripts, stimuli, questionnaires, demonstration videos of the experiments, videos with gaze overlay for the Western and schematic eyes, and videos about the workings of the projector used by Hess are available online: at https://doi.org/10.4121/14134874.v2. The appendices below contain extra analyses and information on Hess's work retrieved from the Drs. Nicholas and Dorothy Cummings Center of the History of Psychology.

Acknowledgments

This work was supported by the Netherlands Organization for Scientific Research under the Replication Studies Program (Grant number 401.16.083). We thank BSc students Stan Otte, Sander van Overbeeke, and Irene Schmidt for constructing the replica of Hess's pupil apparatus. We thank Yke Bauke Eisma for being one of the experimenters. We are grateful to the Drs. Nicholas and Dorothy Cummings Center for the History of Psychology for their support.

Appendix A. Archive

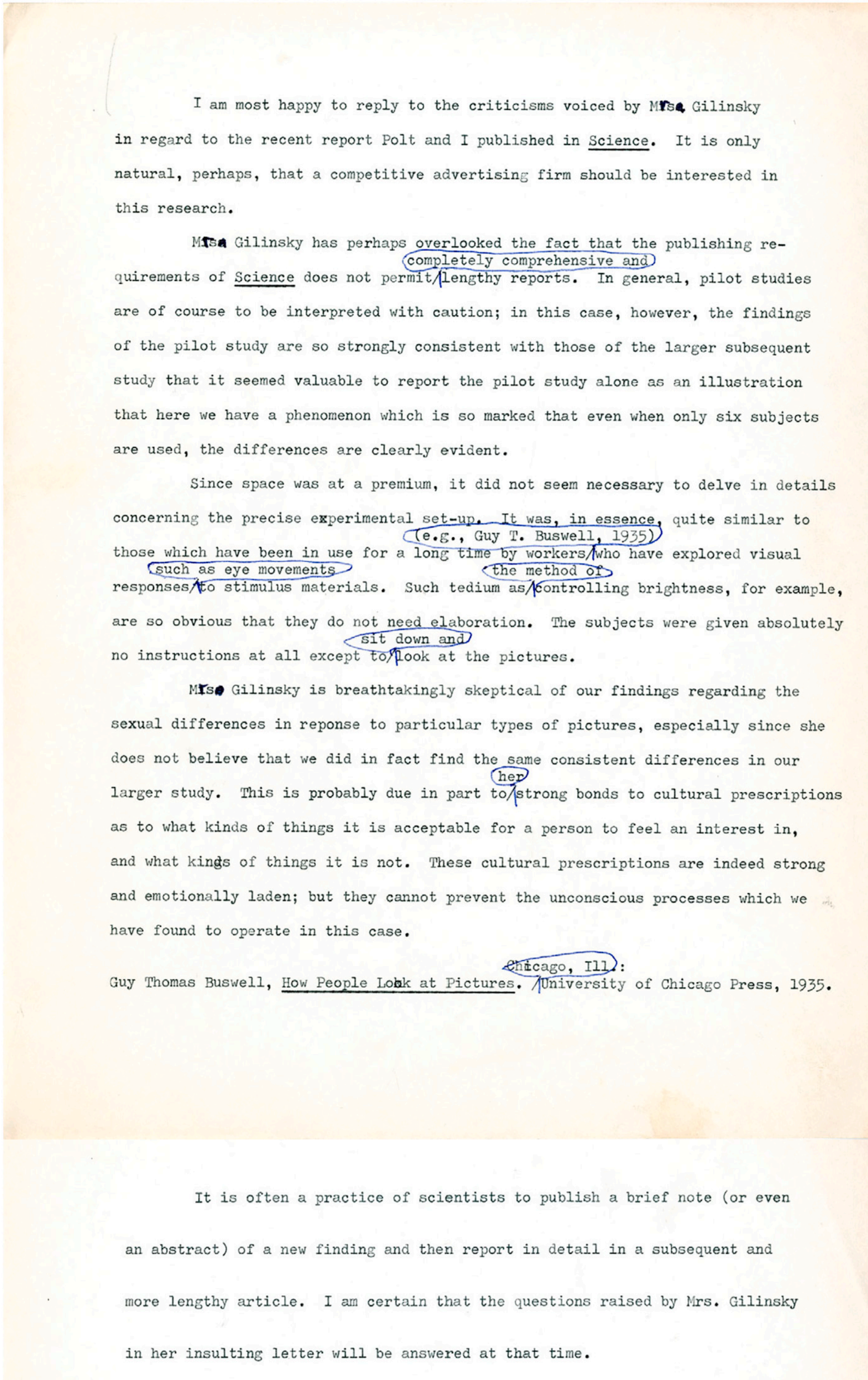
The last author visited the Drs. Nicholas and Dorothy Cummings Center of the History of Psychology, at the University of Akron, Ohio, twice (15–18 August 2017, 22 January–1 February 2018). This archive is home of the Archives of the History of American Psychology, where collections of several psychologists are located. In this archive, there are 48 boxes containing material of Eckhard Hess (Fig. A1). The boxes contain reports, proposals, outlines of presentations, datasheets, notes, correspondence, and photographs of stimuli and equipment. Slides, audiotapes, and film tapes are available in additional boxes. According to the archive staff, the original labeling of the folders and organization of the folders in boxes has been preserved. We inspected the entire collection. In Box *M4138*, a folder labeled *EARLY Pupil Research* contains information associated with the study of Hess and Polt (1960).

Fig. A1. Left. The Hess Collection (only half of the boxes are visible). Middle. Box *M4138*, which contains information about the study of Hess and Polt (1960). Right. The label of the folder with data related to Hess and Polt (1960).

Appendix B. Criticisms of the works of Hess

Hess's pupillometry research has received several criticisms, which can be categorized as follows:

- **Luminance of visual stimuli**. Hess and Polt (1960) stated: "*Brightness was kept relatively constant to rule out an effect of illumination on the size of the pupil*" (p. 350), but did not provide details. In an unpublished letter to *Science* in 1960 criticizing Hess and Polt, Gilinsky asked how "brightness" was kept constant (Box *M4140*, Folder *SCIENCE*). In a draft of a reply to Gilinsky's letter in 1960, Hess wrote: "*Such tedium as the method of controlling brightness, for example, are so obvious that they do not need elaboration*" (Fig. B1). Hess (1965) provided some information about applying luminance control: "*First we show a control slide that is carefully matched in overall brightness to the stimulus slide that will follow it*" (p. 46). Hess (1972), on the other hand, referred to Hess and Polt (1960) as follows: "*Our first published experiment (Hess and Polt, 1960), carried out before we had developed adequate techniques to control brightness…*" (pp. 496–497). In the archive, we found pupil data corresponding to Hess and Polt (1960), with numerical corrections for brightness (see Fig. N3), but were unable to retrieve information about how these corrections were computed and whether they were used in Hess's published works. Goldwater (1972) noted that visual stimuli are problematic in pupillometry research: "*It is difficult to escape the conclusion that visual stimulation is inappropriate in this type of pupillometric research*" (p. 344), whereas Loewenfeld and Lowenstein (1993) noted: "*Anyone familiar with the low threshold of the pupillary light reflex knows, of course, that it is impossible to shift from one picture to a recognizably different one without the likelihood of a pupillary change*" (p. 667). As explained in the Introduction of our paper, differences in luminance between locations within the same image are a possible confounder of gender differences in pupil diameter (Janisse, 1977).
- **Physiological plausibility of a bidirectional pupil response**. The plausibility of a bidirectional pupillary response has been questioned. Janisse (1973) regarded "*intensity, not valance (sic), as the major variable effecting* (*sic*) *the extent of pupillary change. In well controlled experiments, this change has consistently been dilation*" (p. 323), and Loewenfeld and Lowenstein (1993) pointed out: "*Hess's own descriptions varied somewhat with time, from early claims that strong positive feelings evoked* '*extreme dilation*' *and strong negative ones 'extreme constriction' to later statements that bidirectional changes occurred only in some subjects, and only to some pictures*" (p. 667). "*Extreme constriction*" is indeed mentioned in Hess (1965, p. 50), whereas Hess (1975b) argued that pupil constriction is "*an extremely individualistic matter*" (p. 44). In later years, Hess acknowledged that pupil constriction as a response to psychological effects might not be a robust phenomenon: "*The apparent psychopupil constriction indicative of negative affects may in fact be an experimental artifact produced by utilization of particular visual stimuli*" (Hess and Petrovich, 1987, p. 343). Loewenfeld and Lowenstein (1993) noted: "*Now, about 25 years since Hess's first publications, what has been accomplished by all this expenditure of work and time? Nothing, really. It has been shown over and over again that what could not be, according to the anatomic and physiologic properties of the iris system, really was not: emotional stimuli and all other sensory and psychologic stimuli—with the exception of light, and of stimuli that alter the eye's near point of vision—do not constrict the pupil but dilate it*" (p. 667; emphasis as in the original).
- **Sample size.** Hess and Polt (1960) used a small number of participants who viewed each stimulus only once. Hess and Polt argued: "*We purposely report the data for the small sample used in our first study to indicate the type of results obtainable with this technique with a minimum number of subjects*" (p. 350). The use of small sample sizes by Hess has been extensively criticized. Scott et al. (1967) pointed out: "*Hess's results are surprising because most autonomic variables display an amount of spontaneous variability which would make the assessment of interest patterns impossible for groups as small as those used by Hess*" (p. 433). Similarly, Woodmansee (1966) argued in a commentary paper: "*Pupillary diameter can be expected change at least 1% from second to second and as much as 10%–20% over a period of several seconds. Test-retest reliability is generally about .30 in single-trial designs used in studying psychosensory phenomena. With reliability this low, the need for caution in interpretation of findings is obvious*" (p. 134). Zuckerman (1971) commented: "*Parenthetically, it is amazing how the labeling of an experiment as 'pilot' has so little effect in inhibiting the tendency to play up the results. Generalizations about pupillographic sex differences based on these two females and four males have been widely promulgated despite the fact that the author has not yet published an extended study based on an adequate number of subjects*" (p. 318). Similarly, in her letter to *Science* in 1960, Gilinsky wrote: "*To use only two female and four male subjects to represent the sexes on a task in which individual differences are usually large suggests a lack of elementary scientific caution. To argue further that these presumed sex differences are valid indices of differences in 'interest value' is breathtakingly naive*". Hess wrote in his unpublished reply: "*Ms Gilinsky is breathtakingly skeptical of our findings regarding the sexual differences in response to particular types of pictures, especially since she does not believe that we did in fact find the same consistent differences in our larger study. This is probably due to her strong bonds to cultural prescriptions as to what kinds of things it is acceptable for a person to feel interested in, and what kinds of things it is not*" (Fig. B1). Hess mentioned several times that he had replicated the Hess and Polt (1960) study with larger sample sizes. Specifically, in Hess and Polt (1960), it is already reported that "*Further studies, in which we utilized similar materials and more subjects, gave essentially the same results*" (p. 350). Similarly, in his response to Gilinsky, Hess argued that "*we did in fact find the same consistent differences in our larger study*". Zuckerman (1971) mentioned: "*In a personal communication* (*July 17, 1969*) *Hess stated that the Hess and Polt (1960) study has been* '*consistently replicated*' '*with a few thousand subjects.*' *In a second communication* (*August 5, 1969*) *Hess said that he has 'personally run several hundred subjects*' *and found similar results*" (p. 318). Hess (1972) wrote: "*Even though a very small number of subjects was used in this first study, the results have been more than reconfirmed by further unpublished studies of at least 45 subjects, which showed an extremely reliable result for the subjects retested after the interval of a day*" (p. 497). We identified part of the data of one replication study (see Appendix N), but not its processed results, nor the other replications mentioned by Hess. While a large number of subsequent pupillometry studies were conducted (see Appendix S), we could not find evidence in the archive that Hess replicated Hess and Polt (1960).



I am most happy to reply to the criticisms voiced by Mrs. Gilinsky in regard to the recent report Polt and I published in Science. It is only natural, perhaps, that a competitive advertising firm should be interested in this research.

Mrs. Gilinsky has perhaps overlooked the fact that the publishing requirements of Science does not permit completely comprehensive and lengthy reports. In general, pilot studies are of course to be interpreted with caution; in this case, however, the findings of the pilot study are so strongly consistent with those of the larger subsequent study that it seemed valuable to report the pilot study alone as an illustration that here we have a phenomenon which is so marked that even when only six subjects are used, the differences are clearly evident.

Since space was at a premium, it did not seem necessary to delve in details concerning the precise experimental set-up. It was, in essence, quite similar to those which have been in use for a long time by workers (e.g., Guy T. Buswell, 1935) who have explored visual responses such as eye movements to stimulus materials. Such tedium as the method of controlling brightness, for example, are so obvious that they do not need elaboration. The subjects were given absolutely no instructions at all except to sit down and look at the pictures.

Mrs. Gilinsky is breathtakingly skeptical of our findings regarding the sexual differences in reponse to particular types of pictures, especially since she does not believe that we did in fact find the same consistent differences in our larger study. This is probably due in part to her strong bonds to cultural prescriptions as to what kinds of things it is acceptable for a person to feel an interest in, and what kinds of things it is not. These cultural prescriptions are indeed strong and emotionally laden; but they cannot prevent the unconscious processes which we have found to operate in this case.

Guy Thomas Buswell, How People Look at Pictures. Chicago, Ill.: University of Chicago Press, 1935.

It is often a practice of scientists to publish a brief note (or even an abstract) of a new finding and then report in detail in a subsequent and more lengthy article. I am certain that the questions raised by Mrs. Gilinsky in her insulting letter will be answered at that time.

Fig. B1. Draft of a response by Hess to Gilinsky's letter to *Science* in 1960. Source: Box *M4138*, unlabeled folder.

Appendix C. Apparatus

The first published descriptions of the apparatus appeared in Hess and Polt (1964), Hess (1965), and Hess et al. (1965). Hess (1972) provided a detailed overview of the setup, including dimensions and specifications regarding illumination. The apparatus consisted of a box with a viewing aperture at one edge and a screen at the other (Hess, 1972). Stimuli were projected using a Bell and Howell Slider Master projector (Hess, 1972). Inside the box, a lamp illuminated the eyes. The participant's eye was reflected by a mirror towards a 16-mm Bolex camera (Hess, 1972, 1975b; Hess and Polt, 1966; Polt and Hess, 1968) or an Arriflex camera (Hess and Polt, 1964), with a Kilar lens (Box *M4138*, folder *EARLY Pupil Research*), a Kilfitt lens (Hess and Polt, 1964), or a macro-Yvar lens (Hess and Polt, 1966; Hess, 1972). The participant's left eye was recorded (Box *M4138*, folder *EARLY Pupil Research*); an exception is Hess and Polt (1966), where the right eye was measured. Two frames per second were recorded (Hess, 1972; Hess and Polt, 1964; Hess et al., 1965; Polt and Hess, 1968), with an exposure time of 0.25 s (Hess and Polt, 1964). Fig. C1 shows images of pupil apparatuses retrieved from the archive.

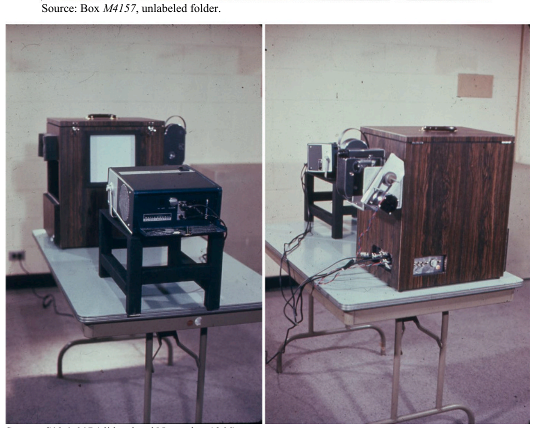
The equipment used by Hess has evolved over the years. In a description of the experimental procedure located in the *EARLY Pupil Research* folder, there is no mention of a box: "*The subject was seated at a table placed directly before the screen. With his face enclosed by a headholder, his eyes were 18 1/2 inches from the screen and centered on the middle of the screen*". Moreover, while a small box is presented in published works (e.g., length of 2½ feet = 76.2 cm in Hess et al., 1965; 24 in. = 61.0 cm in Polt and Hess, 1968; 68.6 cm in Hess, 1972), a large box was used in early research: "*Much as one would build a boat in his basement, without thought of later removal from that basement, we built this apparatus in one of our experimental rooms at the University of Chicago. When the need developed to run subjects in other places, high schools, hospitals, etc., it soon became apparent that we would need a portable machine. The result is the Hess Pupil Response Apparatus. It is easily transported, and be set up at any location with a table, chair and electrical outlet, in a matter of minutes*" (Box *M4144*, Folder *PUPIL TALK OUTLINE - April 1965*). This information is consistent with photographs from the archive, with at least two variations of larger pupillometry boxes and a viewing aperture at the side of the box (Fig. C2). Hess and Polt (1964) referred to a distance of 1.45 m between the head holder and the screen, which could refer to the larger box or no box.

Initially, a 150-w light bulb was used to illuminate the participant's eye (Box *M4138*, folder *EARLY Pupil Research*). In later years, Hess used a 100-w (Hess and Polt, 1964; Hess et al., 1965) or 25-w (Hess and Polt, 1966; Hess, 1972) infrared light bulb. Hess (1972) explained the reason for this change in illumination source: "*Originally, I had used standard negative film* (*Eastman Royal Pan film, ASA 800*) *to record pupil behavior but found it difficult to measure subjects who had dark eyes, because of the lack of contrast between the pupil and the iris. The infrared film produces excellent pictures of any eye*" (p. 505).

Hess and Polt (1960) mentioned that the pupil size was measured by projecting the film using a Percepto-Scope (Perceptual Development Laboratories, St. Louis, MO). From the archive, we retrieved that the model of the Percepto-Scope used was 5102-1 or 5102-2 (Box *M4150*, Folder *M N O P*). The pupil diameter was measured with a ruler (Fig. C3).

Source: Box *M4139*, Folder: *MIRRORS OF THE MIND (Dr. Hess Article)*.

Source: Box *M4157*, unlabeled folder.



Source: *S19.1-017* (slides dated November 1966).

Fig. C1. Measurement setup with a slide projector and a pupillometry box equipped with a lamp, a mirror, a camera, and a rear projection screen.

Fig. C2. Two versions of large pupillometry boxes with viewing aperture at the side of the box.
Source: Left: Box *M4157*, Folder *Cat and Apparatus*. Right: Box *M4167*, Folder *NEW APPARATUS*.

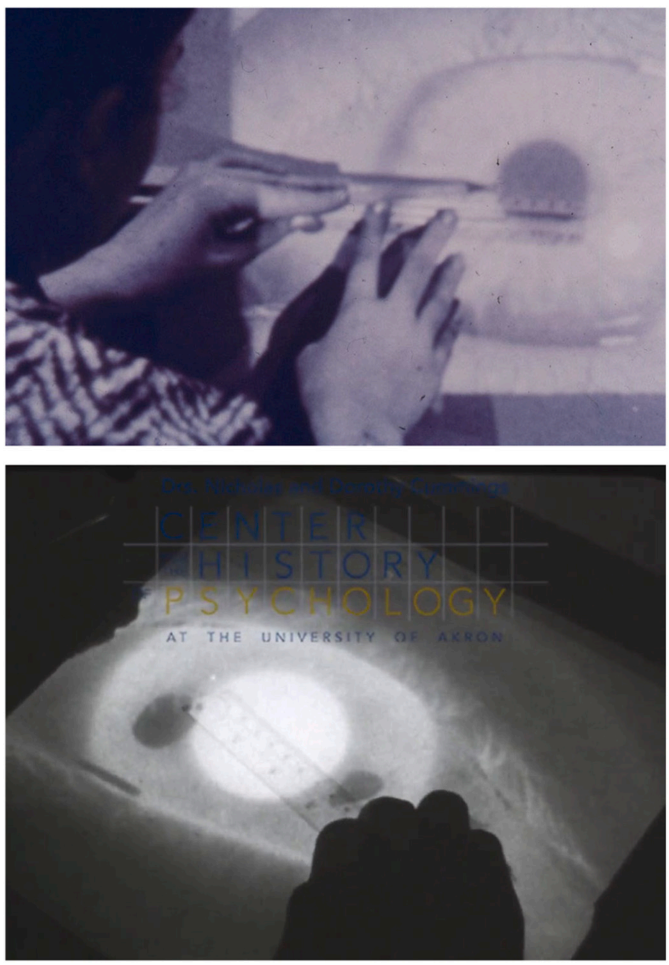


Fig. C3. Manual measurement of pupil size with a millimeter ruler.
Sources: Top: *S15-013*. Bottom: Screenshot from a BBC Horizons documentary (Taylor, 1966), retrieved from the Drs. Nicholas and Dorothy Cummings Center of the History of Psychology, at the University of Akron, Ohio.

Appendix D. Slide change in the Bell and Howell 935 Slide-Master

Mechanical function of the Bell and Howell 935 Slide-Master

We acquired a Bell and Howell 935 Slide Master, the projector used in Hess's research. The slides are stored in a supply tray. When pressing a pushbutton, an actuator drives the shutter in front of the light and pushes the slide towards the supply tray (Fig. D1). After the slide has returned to the supply tray, the supply tray moves one position forward or backward, and a new slide is picked and pulled to the projection location (Hall, 1959). During a slide change, the light from the projector lamp is obscured by the shutter. Also, a mechanical sound is produced.

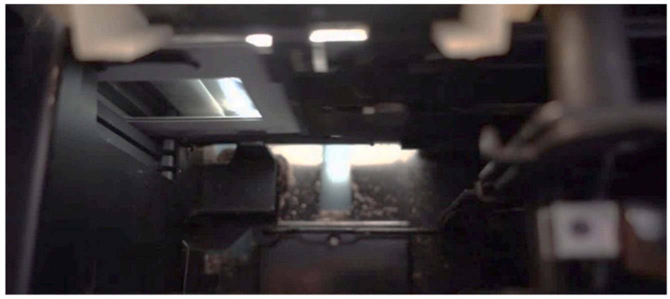


Fig. D1. View from inside the projector, with shutter pushing a slide from the projection location towards the supply tray.

Replica of Hess's pupil apparatus

We built a replica of Hess's pupil apparatus (Fig. D2). For our replica, we relied on Hess (1972), which offers a comprehensive description of the equipment. Accordingly, a box with a length of 686 mm, a width of 381 mm, a height of front panel 305 mm, and a height of back panel 405 mm was fabricated. A rear projection screen of 240 × 150 mm was used, as in Hess and Polt (1964), instead of the 305 × 305 mm screen reported in Hess (1972). An oval viewing aperture with a height of 130 mm and a width of 150 mm was created in the front panel. The front panel was covered with foam for comfortable positioning of the participant's head. A halogen lamp (370 lm, 2800 Kelvin) with a dimmer was placed inside of the box. A 50-mm wide and 75-mm high mirror inside the box reflected the participant's left eye on a high-speed video camera (Sony RX100V) located at the side of the box. The distance between the projector lens and the projection screen was 700 mm.

Luminance during a slide change: methods

We conducted four measurement series to understand the effect of a slide change on luminance. The supply tray was loaded with 40 slide holders, 24 of which contained identical control slides (nine numbers on a gray background, grayscale level 50% or 127 on a scale from 0 to 255; Fig. D3); the remainder of the holders were empty. The digital slides were transferred to 35-mm photographic Kodak film with a Polaroid 8000 film recorder. The slide change was controlled manually by pressing the pushbutton of the projector.

The following measurement series were conducted:

(1) Videos of 24 slide changes were recorded at 1000 Hz through the lens of the projector by placing the camera right in front of the lens (see Fig. D4 for the experimental configuration and Fig. D7, top left, for the corresponding camera view). These measurements were conducted in batches of three slide changes, as the recording time of the camera at this frequency was limited to 4 s.
(2) Video and sound of a continuous sequence of 24 slide changes were recorded at 50 Hz with the camera positioned in front of the viewing aperture of the box and pointing towards the projection screen (see Fig. D5 for the experimental configuration and Fig. D7, top right, for the corresponding camera view).
(3) The luminance of the projection screen (defined as the amount of light reflected from a surface) was measured during a sequence of 24 slide changes using a luminance meter (Konica Minolta LS-150) positioned in front of the viewing aperture and pointing towards the middle of the projection screen (see Fig. D6 for the experimental configuration and Fig. D7, top right, for the corresponding view from the location of the luminance meter – that is, the same as in Measurement series 2).
(4) Videos of 24 pushbutton presses were recorded at 1000 Hz, together with the projection of the slide on the screen, in batches of three slide changes, with the camera positioned in front of the viewing aperture of the box (see Fig. D5 for the experimental configuration and Fig. D7, bottom, for a corresponding camera view).

The room was lit with natural light (Fig. D2). The illuminance (defined as the amount of light that falls on a surface) at the viewing aperture when a control slide was projected on the rear screen was between 755 and 800 lx (measured with a Konica Minolta T-10MA illuminance meter). The 1000-Hz recordings were without audio, whereas the 50 Hz recordings included audio. Video recordings are available in the Supplementary Material.

Fig. D2. Replica of Hess's pupil apparatus. The top lid of the box was closed during the measurements.

Fig. D3. Slide used in all four measurement series in this appendix.

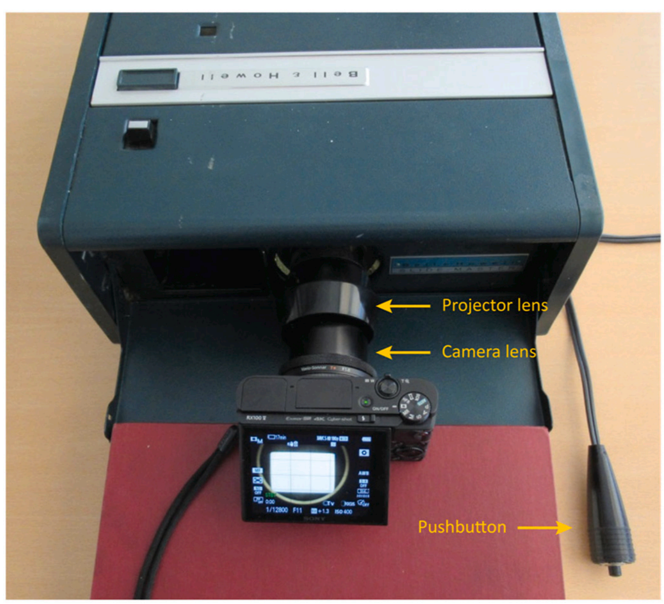


Fig. D4. Measurement configuration for recording the slide change through the projector lens (Measurement series 1).

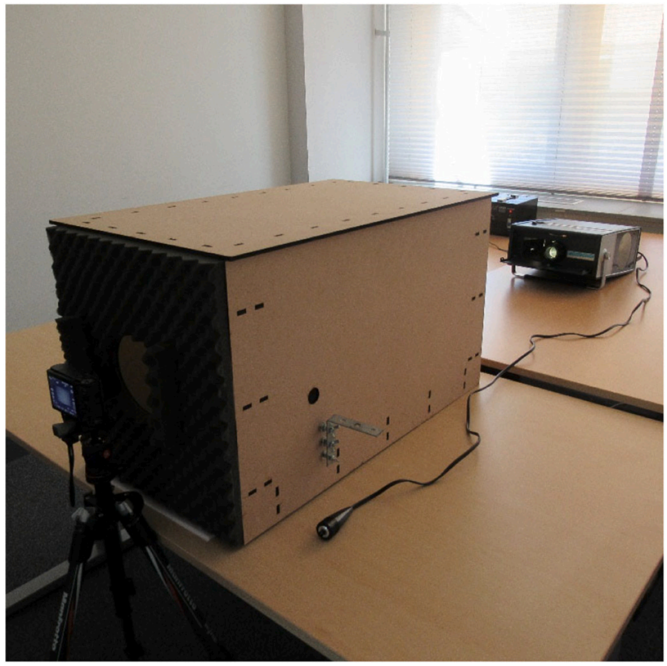


Fig. D5. Measurement configuration for recording the slide change from the viewing aperture (Measurement series 2 and 4).

Fig. D6. Measurement configuration for recording luminance of the rear projection screen (Measurement series 3).

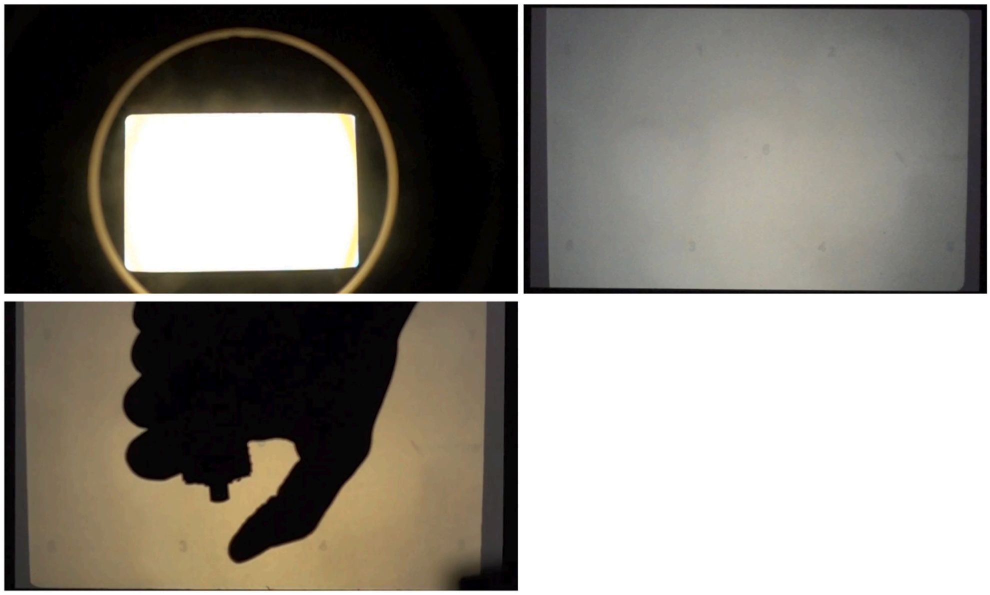


Fig. D7. View from the location of the camera for Measurement series 1 (top left), 2 (top right), and 4 (bottom).

The video frames were exported to .jpg images and read in MATLAB. The images were converted to grayscale, and then to black and white using a threshold value of 230 on a scale from 0 (black) to 255 (white). For each image, we calculated the number of pixels being white, where 100% is the maximum number of white pixels observed. We also calculated the 'change value', defined as the number of pixels being different from the frame 10 ms ago, with 100% being the maximum value observed. 10 ms was used because there was a mild 100 Hz flicker caused by the projector lamp operating at the AC utility frequency of 50 Hz. The change value represents the speed with which the slide was moving.

Luminance during a slide change: results

Fig. D8 (top) shows an example of change value during a slide change, measured with the camera pointed towards the lens (Measurement series 1). Fig. D8 (bottom) shows the luminance on a scale from 0% to 100% (Measurement series 3). A slide change lasted on average 1237 ms, of which 646 ms was entirely dark.

Fig. D9 shows the luminance of the projection screen as measured from the front of the viewing aperture (Measurement series 2). The luminance values were between 1054 and 1175 cd/m^2 when the slide was on the projection location and between 72 cd/m^2 and 86 cd/m^2 during the periods of darkness.

Fig. D10 combines the information of all four measurement series and shows a timeline of a slide change, including luminance and sound production. It can be seen that slide changes yielded a 1-s period of darkness (about 650 ms of full darkness and about 200 ms of partial darkness before and after).

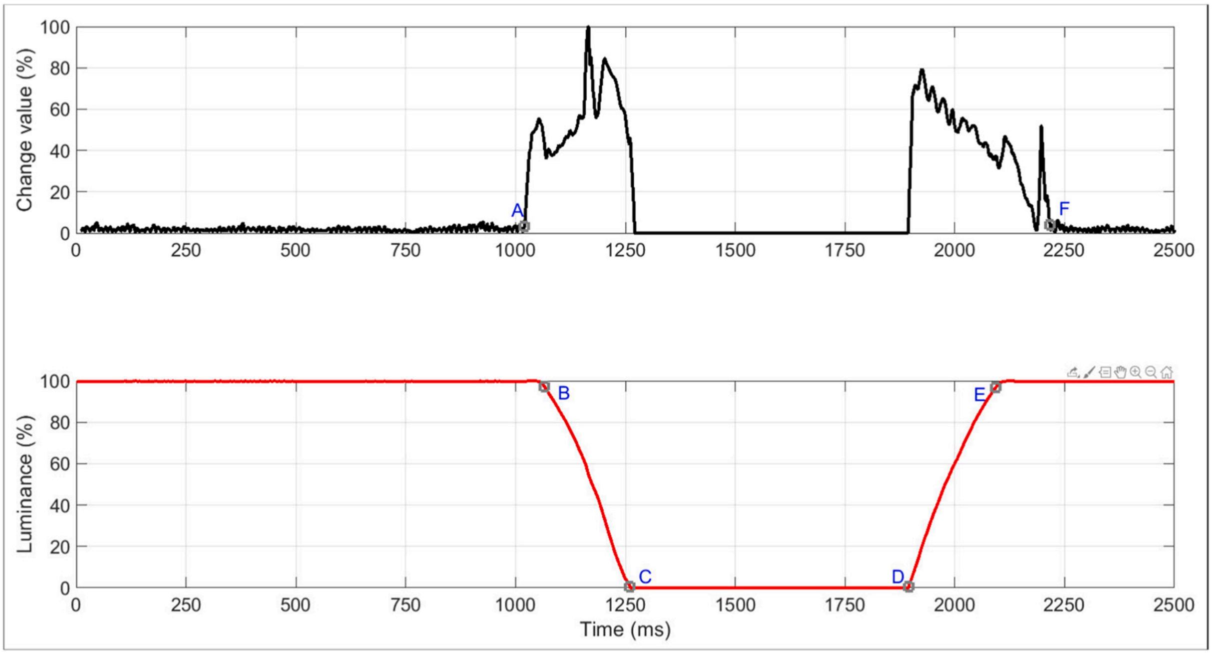


Fig. D8. Change value (%) and luminance (%) and during a slide change. For this example, the full darkness interval was 635 ms, and the total slide change time was 1197 ms. The following intervals can be identified: A–B: Slide is moving away from the projection location and is still fully visible. B–C: Slide is moving away from the projection location and is partially visible (partial darkness). C–D: Slide is not visible (full darkness). D–E: New slide is moving towards the projection location and is partially visible (partial darkness). E–F: New slide is moving towards the projection location and is fully visible.

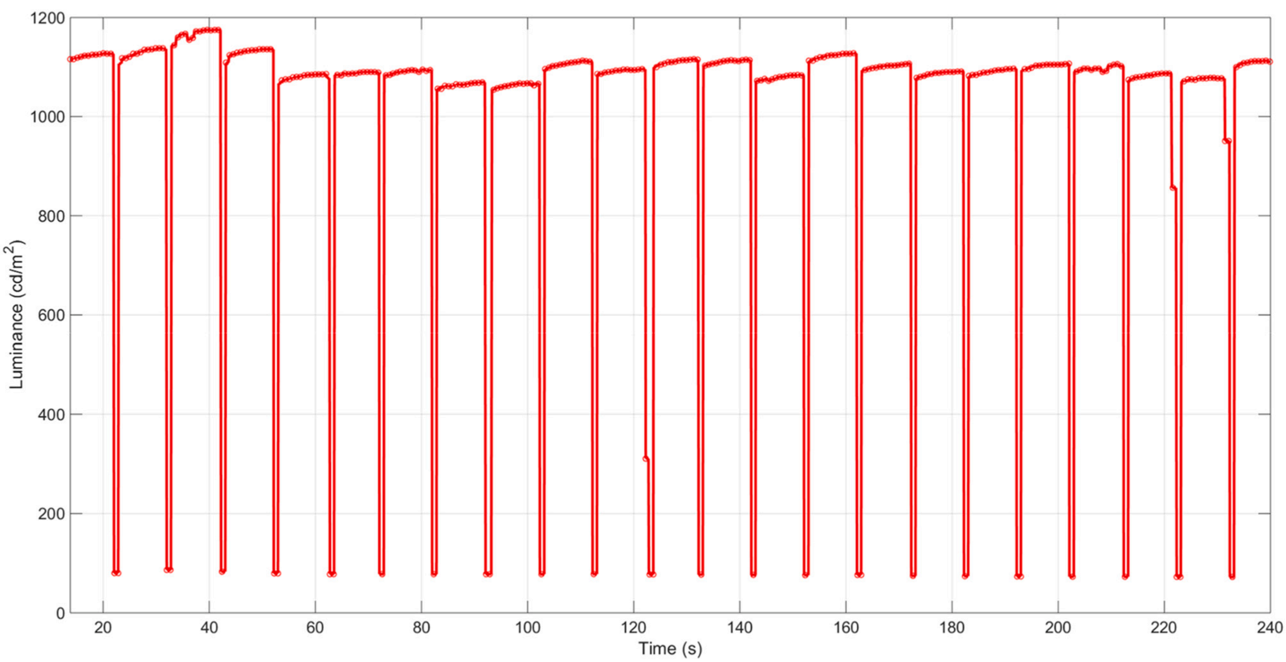


Fig. D9. Luminance of the projection screen as measured in cd/m^2 from the front of the viewing aperture (Measurement series 2) as a function of time.

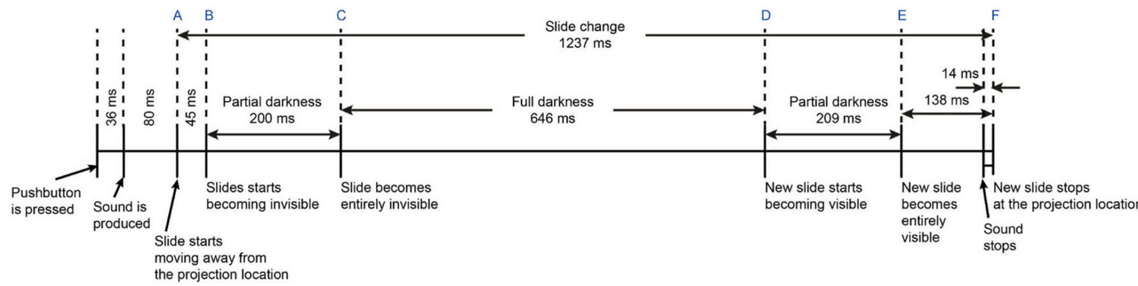


Fig. D10. Timeline of a slide change. A–B: Slide is moving away from the projection location and is still fully visible. B–C: Slide is moving away from the projection location and is partially visible (partial darkness). C–D: Slide is not visible (full darkness). D–E: New slide is moving towards the projection location and is partially visible (partial darkness). E–F: New slide is moving towards the projection location and is fully visible.

Appendix E. Slide change effects on pupil diameter

Measurement series 1: eight multiplication problem trials of twelve participants

We used the replica pupil apparatus to conduct a replication of Hess and Polt (1964). In brief, fifteen participants (14 male, 1 female; mean age: 22.5 years, $SD = 2.2$) were each asked to solve eight multiplication problems (four of which were taken from Hess and Polt, 1964) shown on the projector screen of the apparatus. The slides were printed on transparency film with a Ricoh Aficio MP C3001 laser printer.

Each slide with a multiplication problem was preceded by a control slide depicting an x shown for 7 s. There was no time limit for solving the multiplications. In the analysis presented here, we focus on the pupil light response during the slide change. Three participants were excluded because of poor data quality, leaving 12 participants for further analysis (all male; mean age: 22.8 years, $SD = 2.3$).

The pupil diameter was recorded at 50 Hz. A higher sampling rate was not possible for the required recording time. Moreover, the image quality at high sampling rates is low, which would have inhibited a proper image analysis of the pupil measurements. Considering that the pupil diameter changes are low-frequency, a sampling rate of 50 Hz was deemed sufficient.

Each frame was extracted from the videos, and the resulting images were cropped in MATLAB around the left eye of the participant, the red channel was extracted, converted to binary values, and using the MATLAB function *imfindcircles*, the participant's pupil in each frame was identified. Pupil diameter values during blinks were linearly interpolated from 1 frame before to 1 frame after the blink.

Fig. E1 shows the pupil diameter change (%) as a function of the time from the onset of the slide change. A slide change from a stimulus to a control slide occurred at around 0 s, and a slide change from a control slide to a stimulus slide occurred at 10 s. It can be seen that, during a slide change, the pupil dilates, then constricts. The slight increase observed during the stimulus slide is due to mental effort associated with solving the multiplications.

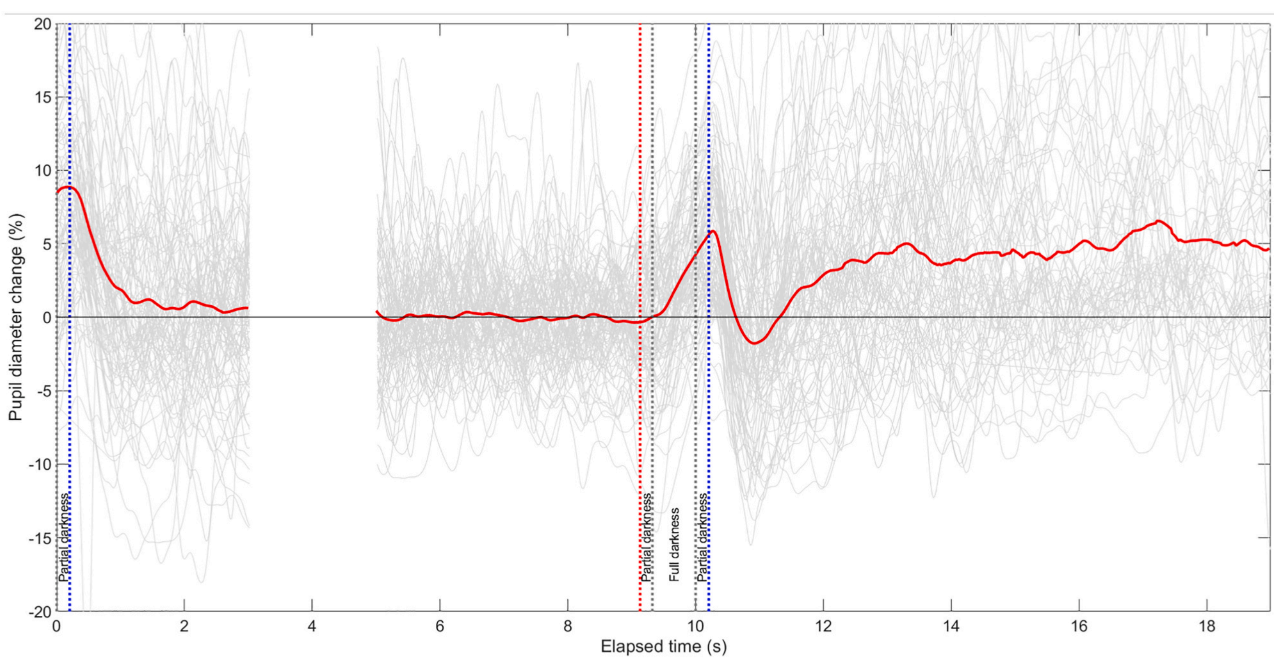


Fig. E1. Pupil diameter change (%) as a function of time calculated using data generated with our replica of Hess's pupil apparatus (12 participants × 8 trials). A positive value indicates pupil dilation; a negative value indicates pupil constriction. The solid red line represents the mean of 96 time series (12 participants × 8 trials). The light gray lines represent the 96 individual time series. The gray dotted vertical line indicated the moment of transition from a stimulus slide to a control slide at 0 s and from a control slide to a stimulus slide at 10 s. The gap between 3 and 5 s is because the control slide was shown for only 7 s. The red dotted vertical line indicates the onset of partial darkness (point B in Figs. D8 & D10), and the blue dotted vertical line indicates the end of partial darkness (point E in Figs. D8 & D10). The period between the two gray vertical dotted lines defines the period of full darkness (interval C–D in Fig. D8).

Measurement series 2: twenty-three time series of one participant

Measurement series 1 was conducted with slides printed with a laser printer leading to some visual inhomogeneity. Moreover, the slide background was black. In Measurement series 2, we used the same control slides as in Appendix D, which were more homogenous thanks to their production via a film recorder. Moreover, the slide background was gray, and thus more similar to the control slides used by Hess (Appendix G). Using the replica pupil apparatus, the pupil diameter of a single participant was recorded at 50 Hz for 46 slide changes (Fig. E2).

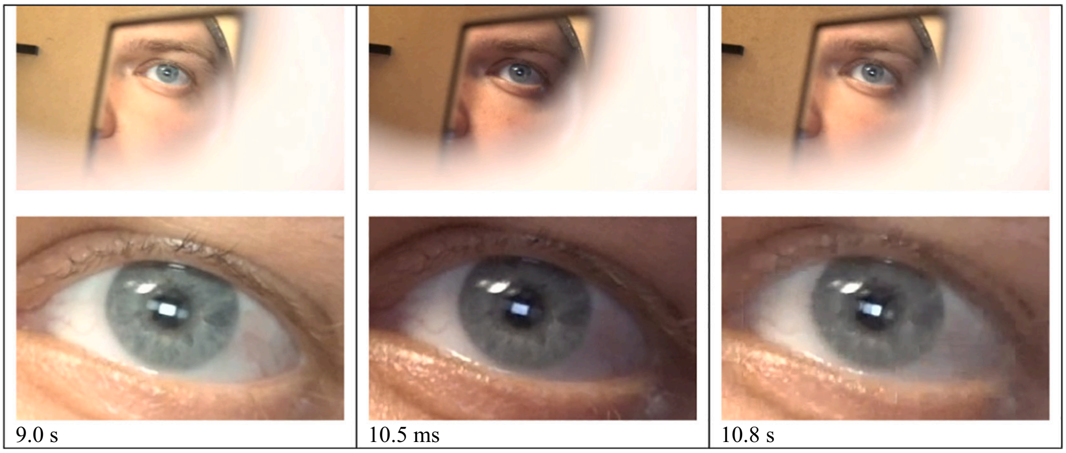


Fig. E2. Pupil during a slide change. At 9.0 s, the slide is stationary on the projection screen. At 10.5 s, the slide has entirely moved away from the projection screen (full darkness). At 10.8 s, the new slide has started becoming visible (partial darkness).

The image and data processing was conducted as in Measurement series 1. Fig. E3 shows the pupil diameter change (%) as a function of the time from the onset of the slide change. It can be seen that during a slide change, the pupil dilates, then constricts. The peak pupil diameter change is smaller than in Fig. E2, likely because the slides in Measurement series 1 were darker than the slides in Measurement series 2, making the difference in luminance between the slides and the shutter that was visible at the projection location during a slide change in the former case smaller.

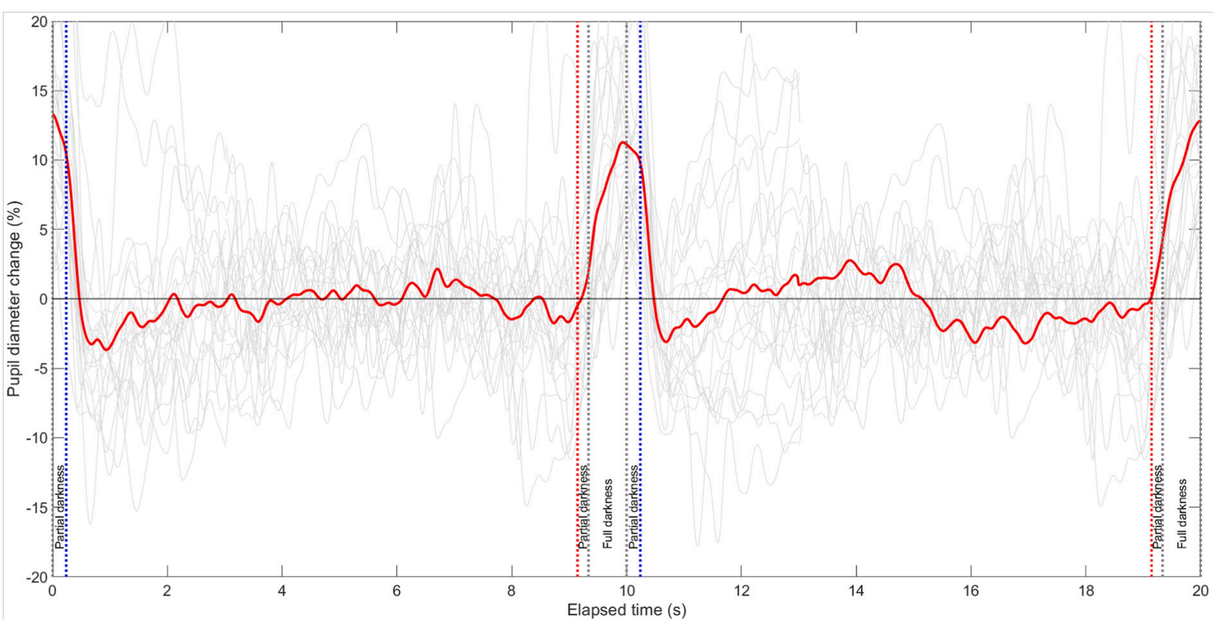


Fig. E3. Pupil diameter change (%) as a function of time calculated using data generated with our replica of Hess's pupil apparatus (1 participant × 23 trials). A positive value indicates pupil dilation; a negative value indicates pupil constriction. A slide change occurred at around 0 s, 10 s, and 20 s. The solid red line represents the mean of 23 time series (46 measurements presented in pairs) of the one participant. The red, blue, and gray vertical lines are as in Fig. E1.

Appendix F. Participants in Hess and Polt (1960)

Hess and Polt (1960) reported that the participants were "*one single female, one married female, three single males, and one married male. Neither of the married subjects had children*" (p. 350). We retrieved details of the participants from scoresheets and notes (Table F1).

Table F1
The six participants from Hess and Polt (1960).

No	Initials	Presumed relation to Hess	Gender	Age	Marital status
1	EK	Graduate student of Hess, co-author of two papers	Male	30	Married
2	IK	Graduate student in biopsychology	Male	22	Unmarried
3	GK	Bachelor student in psychology	Male	20	Unmarried
4	RR		Male	Early 20s	Unmarried
5	GL	Research assistant	Female	Late 20s or early 30s	Unmarried
6	AT	Graduate student in clinical psychology	Female	24	Married

Source: Box *M4138*, Folder Early Pupil Research.

Appendix G. Control slides

Several types of control slides were retrieved from the archive, some with an x in the middle and others with five numbers, some in portrait and others in landscape format (Fig. G1). Hess and Polt (1960) stated that their control slide concerned a "*10-second presentation of the test pattern*" (p. 350). Subsequent publications refer to a slide with five numbers (see Hess, 1965 for an image such a slide, and Hess, 1972, 1975b for textual descriptions). We retrieved a series of 12 identical slides with five numbers; the series was coded as *B-x*, with *x* being an odd number (Fig. G2), which could correspond to the B-series of the stimuli (see Appendix I).

For Experiments 1 and 2, we used a control slide consisting of the numbers 1 to 9, presented in a black outline of 2-pixel thickness, in Mangal font with a height of 44 pixels (0.8°) (Fig. G3).

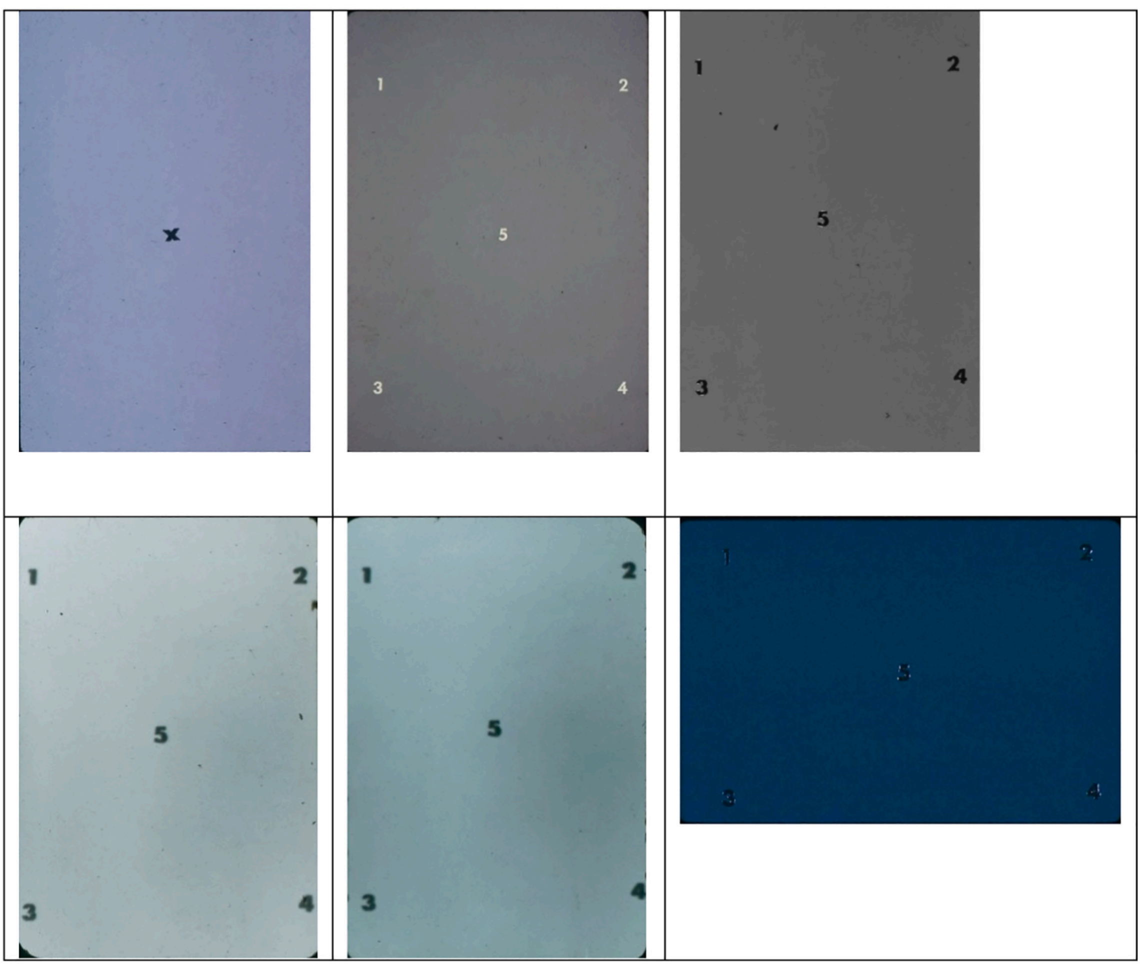


Fig. G1. Examples of control slides.
Source: *S19.1-020* and *S19.1-021*.

Fig. G2. Control slides coded as *B-x*, with *x* being an odd number. The slides were identical.
Source: *S19.1-021*.

Fig. G3. Control slide used in Experiments 1 and 2.

Appendix H. Stimuli used in Experiments 1 and 2, Images of five themes

In the archive, we found slides from a presentation with images and the same pupil size data as Hess and Polt (1960) (Fig. H1). We used these images in Experiment 1 (Fig. H2). Fig. H2 also shows the modern equivalents used in the experiment and their sources.

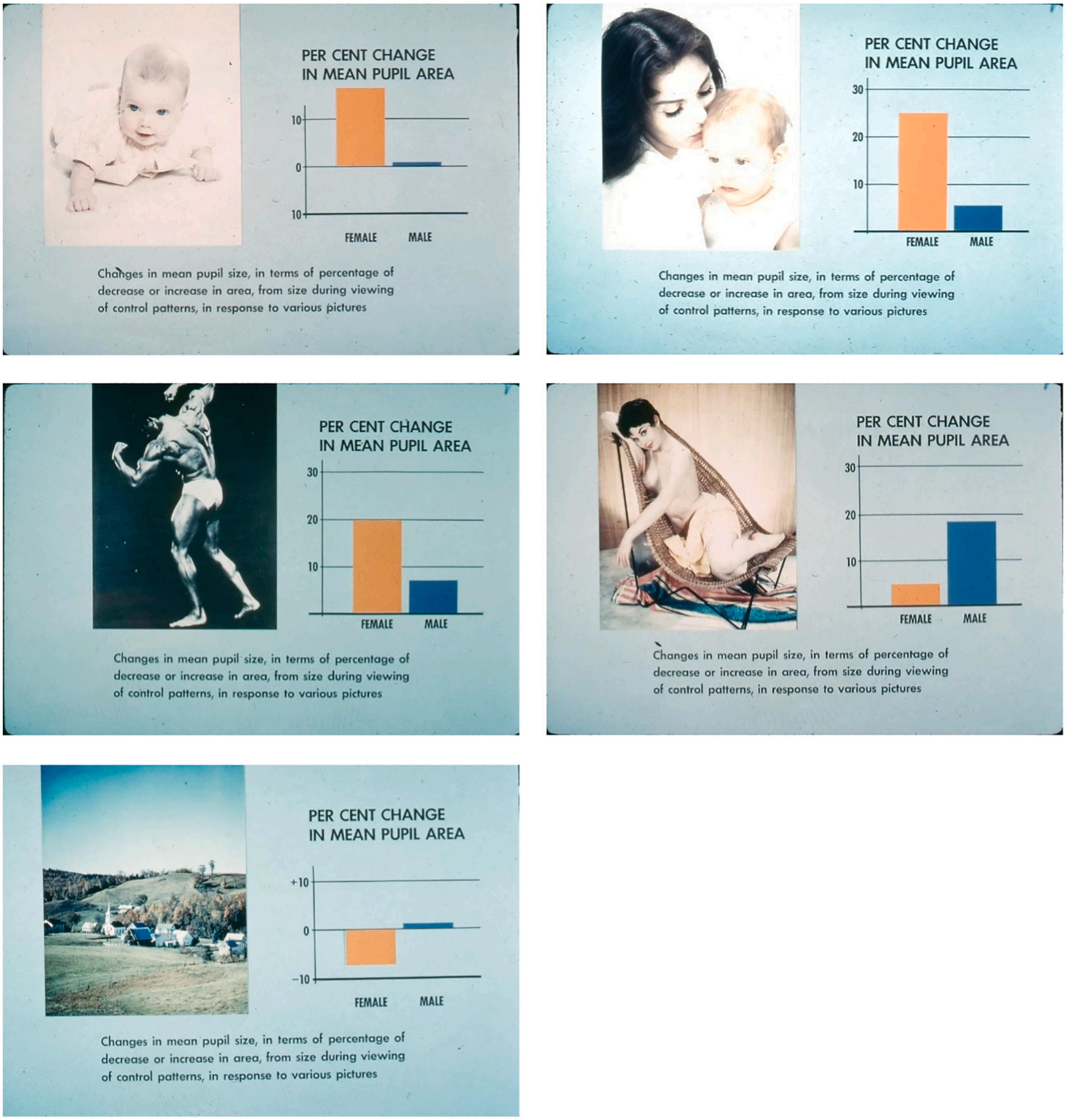


Fig. H1. Slides from a presentation dated October 1962 with pupil data corresponding to Hess and Polt (1960). The images at the left side of the slides were used in Experiment 1 (see Fig. H2).
Source: *S17-008*.

Fig. H2. Images of five themes used in Experiment 1. Top: Images from a presentation by Hess in 1962 as retrieved from the Drs. Nicholas and Dorothy Cummings Center of the History of Psychology, at the University of Akron, Ohio (Source: *S17-008*). Left to right: Baby, Mother and baby, Male, Female, Landscape. Bottom: Modern equivalents.

The sources of the modern equivalents are as follows:

- Baby: dolgachov (photographer). (n.d.). *Bright picture of crawling baby boy in diaper* [photograph] (Image ID: 3348561). Retrieved from https://www.123rf.com/photo_3348561_bright-picture-of-crawling-baby-boy-in-diaper.html
- Mother and baby: linavita (photographer). (n.d.). *A mother with a small child* [photograph] (Stock photo ID: 381120277). Retrieved from https://www.shutterstock.com/nl/image-photo/mother-small-child-381120277
- Male: Ivanov, Vadim (photographer). (n.d.). *Muscled male model posing in studio* [photograph] (Stock photo ID: 91330259). Retrieved from https://www.shutterstock.com/image-photo/muscled-male-model-posing-studio-91330259
- Female: Ollyy (photographer). (n.d.). *Beautiful naked woman sitting on an old chair in an empty room* [photograph]. (Stock photo ID: 100851019). Retrieved from https://www.shutterstock.com/image-photo/beautiful-naked-woman-sitting-on-old-100851019
- Landscape: Mirvav (photographer). (n.d.). *Picturesque village in the South Bohemian* [photograph] (Stock photo ID: 61052128). Retrieved from https://www.shutterstock.com/image-photo/picturesque-village-south-bohemian-61052128

All images in Experiment 1 were converted to grayscale and processed to have the same mean grayscale level for all images and a similar standard deviation of the grayscale level between the original and modern version of each image (Table H1).

Table H1
Means (standard deviations in parentheses) of the percentage grayscale level of the 2,073,600 pixels of the images used in Experiment 1, on a scale from 0% (black) to 100% (white).

Stimulus	Percentage grayscale level
Baby – Modern	49.86 (3.78)
Baby – Original	49.72 (3.67)
Female – Modern	49.71 (17.64)
Female – Original	49.84 (18.49)
Landscape – Modern	49.81 (13.40)
Landscape – Original	49.80 (13.26)
Male – Modern	49.84 (11.79)
Male – Original	49.87 (11.45)
Mother and baby – Modern	49.80 (19.02)
Mother and baby – Original	49.81 (18.84)

In Experiment 2, line drawings were used instead of images (Fig. H3).

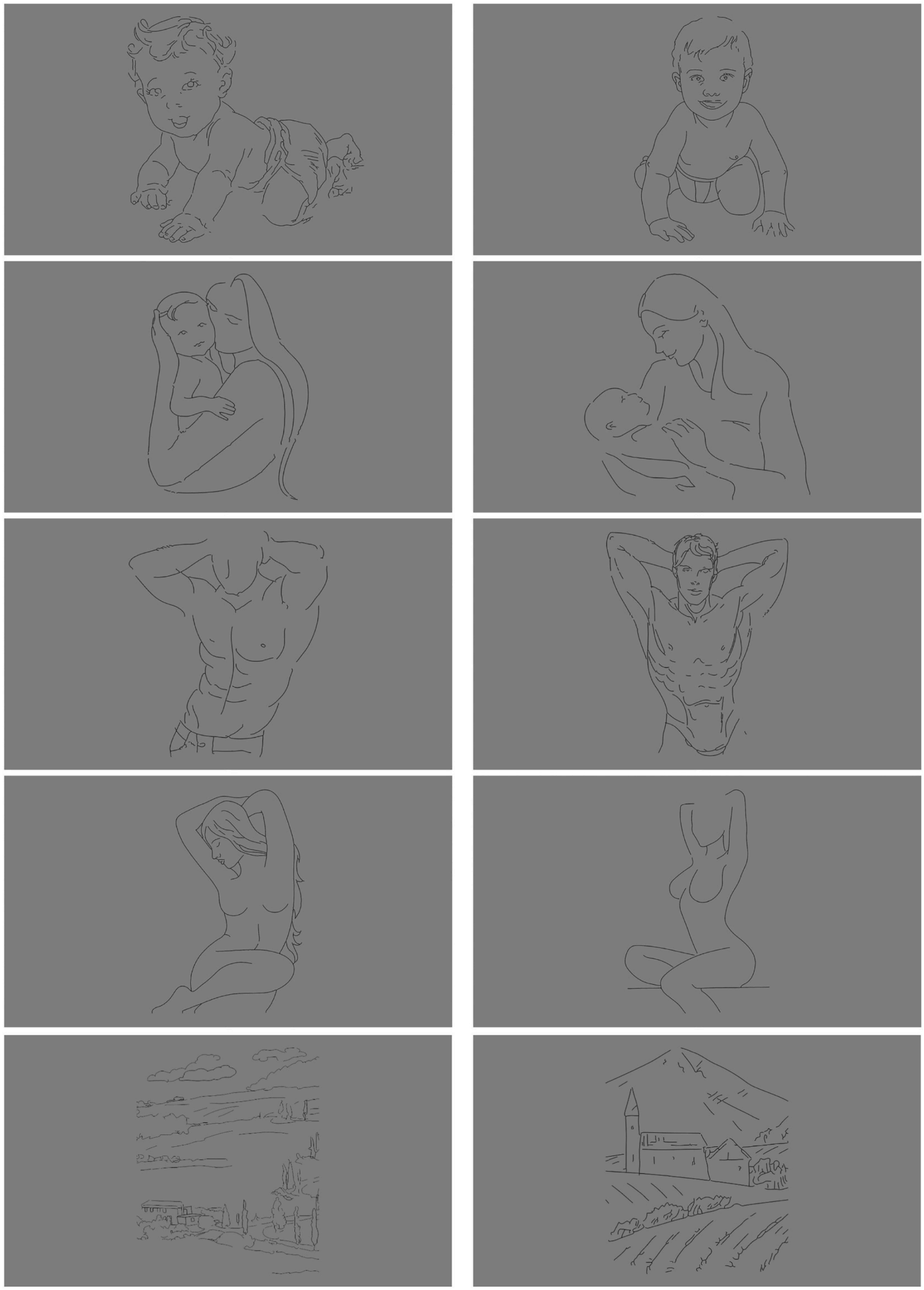


Fig. H3. Line drawings used in Experiment 2. Per row, from left to right: Baby 1 & 2, Mother and baby 1 & 2, Male 1 & 2, Female 1 & 2, Landscape 1 & 2.

The sources of the line drawings in Experiment 2 are as follows:

- Baby 1: RetroClipArt (Illustrator/Vector artist). (n.d.). *Crawling baby* [vector]. (Stock vector ID: 56756374). Retrieved from https://www.shutterstock.com/image-vector/crawling-baby-retro-clip-art-56756374
- Baby 2: Pop Path (2017, January 8). How to draw a baby laughing [blog]. Retrieved from https://web.archive.org/web/20190407204540/http://poppath.com/how-to-draw-a-baby-laughing/
- Mother and baby 1: CloudyStock (Illustrator). (n.d.). *Woman with a child. Logo of a young mother with a baby in her hands. Black and white illustration of a mother hugging her baby. Logo family. Tattoo* [vector] (Stock vector ID: 795743269). Retrieved from https://www.shutterstock.com/image-vector/woman-child-logo-young-mother-baby-795743269
- Mother and baby 2: ValeriSerg (Illustrator/Vector artist). (n.d.). *Mommy holding baby. Mom and baby in the room with window. Happy family. Black and white vector sketch. Simple drawing* [vector] (Stock vector ID: 751075258). Retrieved from https://www.shutterstock.com/image-vector/mommy-holding-baby-mom-room-window-751075258
- Male 1: Irina_QQQ (Illustrator/Vector artist). (n.d.). *Art sketched portrait of young sexy muscular powerful man in pose* [vector] (Stock vector ID: 280065848). Retrieved from https://www.shutterstock.com/image-vector/art-sketched-portrait-young-sexy-muscular-280065848
- Male 2: profartshop (Illustrator/Vector artist). (n.d.). *Sexy male body art* [vector] (Stock vector ID: 674022838). Retrieved from https://www.shutterstock.com/image-vector/sexy-male-body-art-674022838
- Female 1: Grama, Elena (Illustrator/Vector artist). (n.d.). *Silhouette of a beautiful naked woman* [Illustration] (Stock illustration ID: 114010270). Retrieved from https://www.shutterstock.com/image-illustration/silhouette-beautiful-naked-woman-114010270
- Female 2: Trawczynski, Marek (Illustrator/Vector artist). (n.d.). *Nude woman sitting* [vector] (Stock vector ID: 177959498). Retrieved from https://www.shutterstock.com/nl/image-vector/nude-woman-sitting-vector-illustration-177959498
- Landscape 1: gaudenzi, silvia (Illustrator/Vector artist). (n.d.). *Tuscan landscape in black and white* [Illustration] (Stock illustration ID: 666393622) https://www.shutterstock.com/image-illustration/tuscan-landscape-black-white-666393622
- Landscape 2: bioraven (Illustrator). (n.d.). *Vector hand-drawn village houses sketch and nature* [vector] (Stock vector ID: 310063286). Retrieved from https://www.shutterstock.com/image-vector/vector-hand-drawn-village-houses-sketch-310063286

Appendix I. Stimuli used by Hess and Polt (1960)

We identified a handwritten draft of a part of the Hess and Polt (1960) paper (Fig. I1). This draft contained code names of five slides (baby B-28, mother & baby C-22, nude man C-20, nude woman C-26, landscape C-12) with corresponding pupil change data for male and female participants consistent with Hess and Polt (1960).

Further inspection of the archive revealed that the five slides belonged to two slide series (coded as B-series and C-series), each series consisting of 30 slides. The even-numbered slides were stimuli slides, and the odd-numbered slides were control slides. We retrieved descriptions of the slide content of these series (Table I1). We were unable to retrieve photos or slides with the stimuli of the B-series, but we retrieved photos and/or slides matching the descriptions for the stimuli of the C-series (Table I1).

Hess and Polt (1960) did not mention that the five slides were part of a more extensive series. However, in a subsequent review paper, Hess stated that "*the sequence of control and stimulus is repeated about 10 or 12 times a sitting*" (Hess, 1965, p. 46).

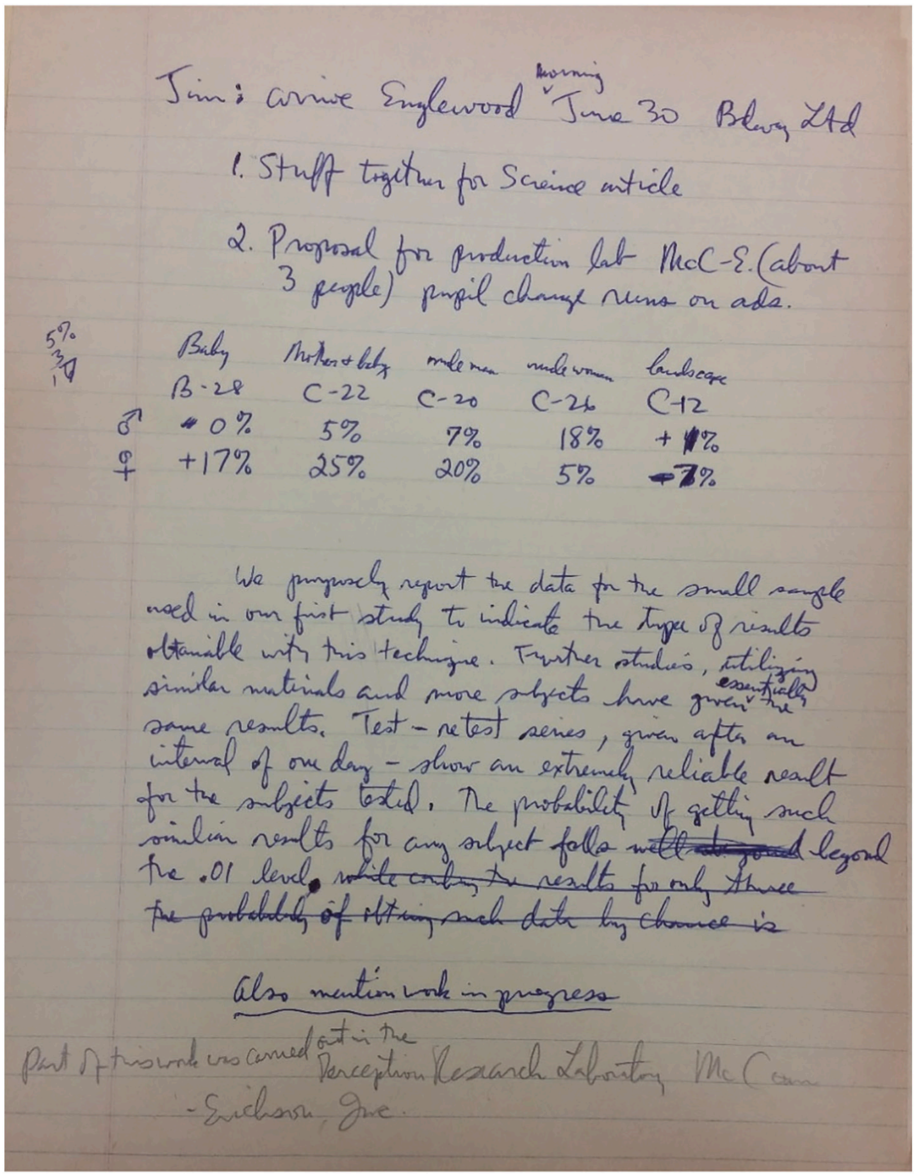
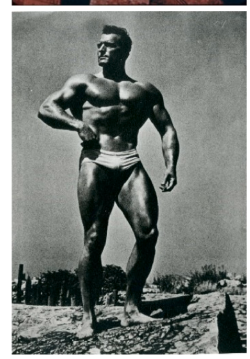
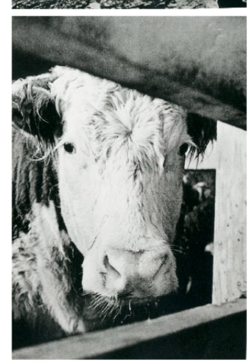


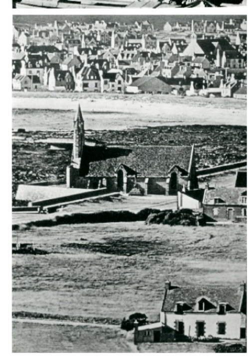
Fig. I1. Handwritten draft with data corresponding to Hess and Polt (1960).
Source: Box *M4138*, Folder *Early Pupil Research.*

Table I1
Descriptions of stimuli in the B- and C-series and presumed stimuli of the C-series. Slides B-28, C-12, C-20, C-22, and C-26, which were used in Hess and Polt (1960), are indicated in boldface.

Slides: B-series, 'light'	Slides: C-series, 'dark'	C-series stimuli	Source (bw = black-and-white, c = color)
2. Scene of bay, with barren hills left foreground and center.	2. Scene with brown and reddish brown stone buildings and towers. Street lower right.		*V108-3* (bw) *S15-16* (c) *S19.1-17* (c) (dated November 1966)
4. Ajax. Blue can left center, sink center, slogan top, copy below.	4. Front view of muscular man in brief bathing trunks. Fills center almost top to bottom.		*V108-F3* (bw) *S15-16* (bw) *S19.1-17* (bw) (dated November 1966)
6. Ajax. red can left center, sink centr (*sic*), copy below.	6. Front view of face of young steer. Almost fills.		*V108-3* (bw) *S15-13* (bw)
8. Ajax. Blue can. Can larger, left center, slogan top, copy below.	8. Profile of attractive girl. Looking up. Breast partially exposed above gown, lower left.		*V108-3* (bw)

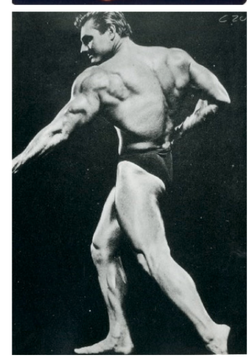
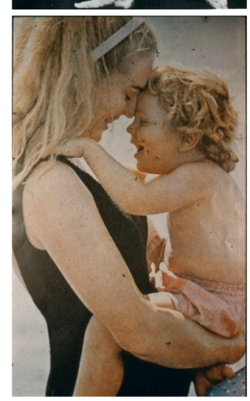
(*continued on next page*)

Table I1 (*continued*)

Slides: B-series, 'light'	Slides: C-series, 'dark'	C-series stimuli	Source (bw = black-and-white, c = color)
10. Beagle puppy left center, kitten right center.	10. Very young girl sitting and holding glass of orange juice. Blond. Face up-center.		*V108-3* (bw)
12. Satura sheen. head and face center, bottle lower right.	**12. Bay scene. Boats foreground and up-center. Buildings toward top, with sky above.**		*V108-3* (bw)
14. Dorothy Gray. Satura lipsitck (*sic*). Face in mirror center left. Lipstick center right, copy below.	14. Rural scene. House lower right. Group of buildings around church across center. Village across top.		*V108-3* (bw) *S15-13* (bw)
16. Dorothy Gray. "Apple on a stick". Phrase across lower center. Face above. Hand holding stick with apple center left. Copy below.	16. Girl. Bare from waist up. Breast center.		*V108-3* (bw)

(*continued on next page*)

Table I1 (*continued*)

Slides: B-series, 'light'	Slides: C-series, 'dark'	C-series stimuli	Source (bw = black-and-white, c = color)
18. Carnation ad. Baby's face almost fills. Can lower left.	18. Heads of lovers. Evidently reclining. Girl lower center, man upper center.		*V108-3* (bw) Box *M4180*, Folder *Cats and Photos* (c)
20. USP ad. Three mounds of potash. Copy Below.	**20. Side view, muscular young man. Almost fills center, top to bottom.**		*V108-3* (bw) *S15-13* (bw)
22. USP ad. Three linear cross designs. Copy below.	**22. Side view of mother holding young child whose legs are around her waist. Mother's face is upper left. Child's face is upper center.**		*V108-3* (bw) Box *M4180*, Folder *Cats and Photos* (c) *S15-16* (c)
24. USP ad. Three piles of pothash (*sic*), still being poured. Copy below.	24. Nude under water. Prominent breasts fill center.		*V108-3* (bw)

(*continued on next page*)

Table I1 (*continued*)

Slides: B-series, 'light'	Slides: C-series, 'dark'	C-series stimuli	Source (bw = black-and-white, c = color)
26. USP ad. Three piles potash, being poured from containers which are in view. Copy below.	**26. Nude holding garment to right. Face up-center. Breast below.**		*V108-3* (bw) *S15-13* (bw)
28. Ivory ad. Baby center. Box of Ivory Snow lower right. Copy lower left and center.	28. Girl, scantily clothed, sitting on white skin rug. Fills center.		*V108-3* (bw) *S15-16* (c) *S19.1-17* (bw)
30. Puppy left center. Kitten right center.	30. Girl, side view, kneeling, nude from waist up. Most of breasts (center) concealed by fur piece.		*V108-3* (bw)

Note. It is unclear whether the stimuli used in Hess and Polt (1960) were in color. The stimuli in the B-series were likely in color (as a blue vs. red can is mentioned in the descriptions of B4 vs. B6). In the archive, we found some stimuli in black-and-white, others in color, and some in both.
Folder *V108-3* is located in Box *M4180*. The odd-numbered slides were the control slides.
Slides were stored in boxes separated from the 48 ones mentioned in Appendix A. Source locations starting with '*S*' refer to slide boxes.
The slide descriptions were retrieved from the archive (two copies; source: Box *M4146*, Folder *EYE MOVEMENT DATA* and Box *M4170*, Folder *STIMULI - for Pupil Research Word Series: Slides-Description Slides-Series B-C-D-E-F-G*).

Appendix J. Experimental setup, examples of stimuli, and extra stimuli and results in Experiments 1 & 2

Fig. J1 shows the experimental setup used in Experiment 2. The setup in Experiment 1 was the same, but the experiment took place in a different room.

Fig. J1. The setup of Experiment 2.

Replication of Study 2 (Hess and Polt, 1964) (Experiment 1)

Fig. J2. An example of a multiplication slide used in Replication of Study 2 (Hess and Polt, 1964) in Experiment 1.

Replication of Study 3 (Hess, 1975a) (Experiment 2)

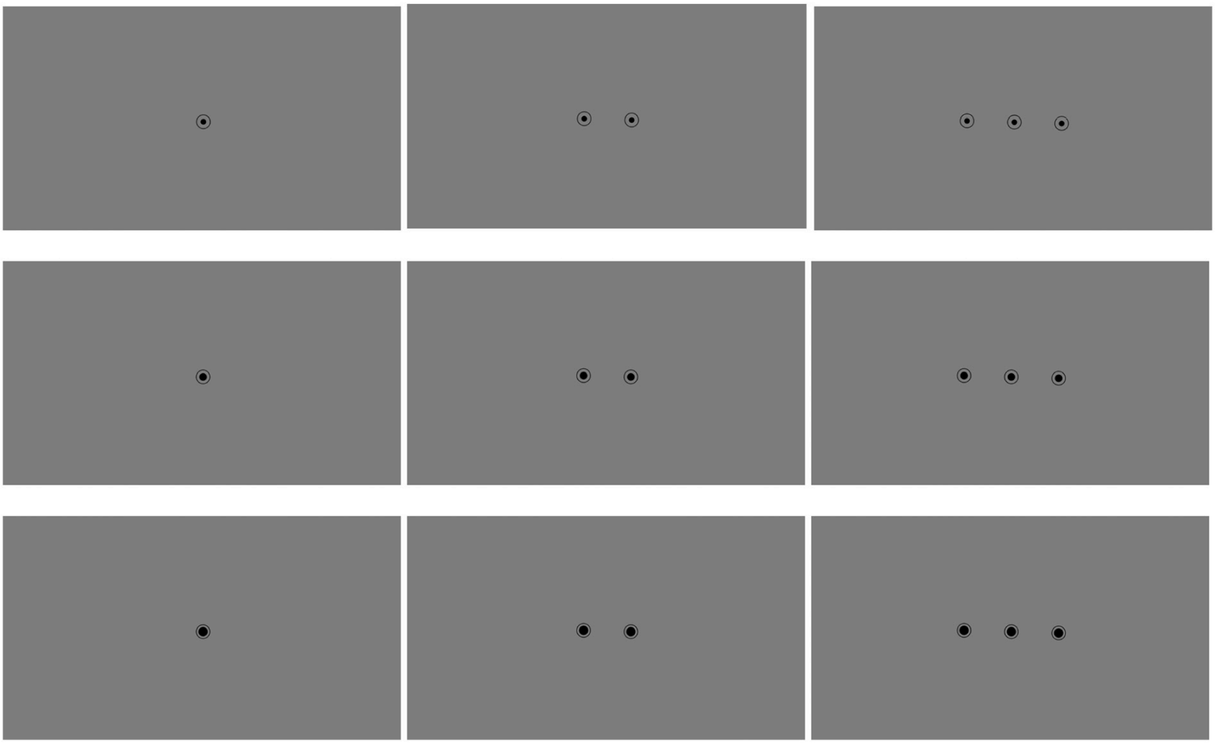


Fig. J3. Stimuli with schematic eyes used in Replication of Study 3 (Hess, 1975a) in Experiment 2.

Replication of Study 4 (Hess, 1975b) (Experiment 2)

Fig. J4. Video frame from the Western video clip used in Replication of Study 4 (Hess, 1975b) in Experiment 2.

Replication of Study 5 (Polt and Hess, 1968) (Experiment 2)

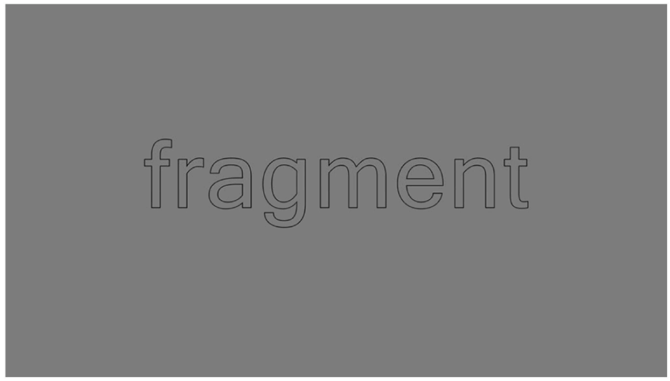


Fig. J5. Example of a word stimulus used in Replication of Study 5 (Polt and Hess, 1968) in Experiment 2.

Extra stimuli and results: grayscale images (Experiment 1)

At the end of Experiment 1, a series of grayscale images were shown to determine the maximum possible pupillary effects due to screen luminance. More specifically, eleven grayscale images were presented in the following order of grayscale levels (corresponding 8-bit values in parentheses) for all participants: 100% (255), 78% (200), 59% (150), 39% (100), 20% (50), 0% (0), 20%, 39%, 59%, 78%, 100%.

The sequence of grayscale images showed that screen luminance has substantial effects on pupil diameter change, causing constrictions and dilations up to 30%, as shown in Fig. J6. The dynamics are asymmetric: dilation is a slow process. Constriction, on the other hand, occurs rapidly, starting after about 0.3 s and lasting for about 1 s. After constriction, so-called pupillary escape occurs (Loewenfeld and Lowenstein, 1993).

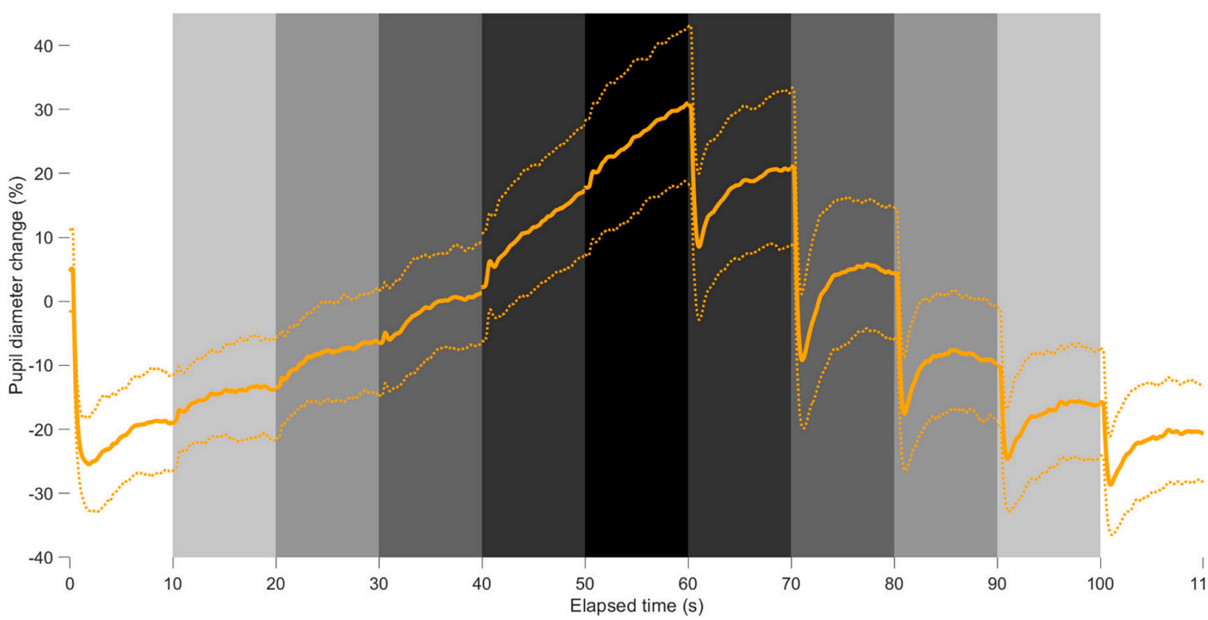


Fig. J6. Mean pupil diameter change (PC_t) for the grayscale images in Experiment 1. A positive value indicates pupil dilation; a negative value indicates pupil constriction. The thin dotted lines represent the mean ± 1 standard deviation. The grayscale images were not preceded by a control slide. The pupil diameter change is expressed with respect to the mean pupil diameter for the ten control slides belonging to the images of the five themes.

Extra stimuli and results: schematic pupils (Experiment 2)

At the beginning of the study with the schematic pupils (Replication of Study 3, Experiment 2), a drawing of a happy face and a drawing of an angry face were shown (Fig. J7). These drawings were presented in Hess (1973c, 1975a, 1975b), Hess and Goodwin (1974), and Hess and Petrovich (1987), in an experiment in which participants were asked to draw pupils with a size that best fitted the expression of each face (and see Kret, 2018 for a recent replication). We redrew the faces from letter-size printouts retrieved from the archive. The heights of the redrawn sad and happy face were 906 and 908 pixels, respectively, and the line thickness was proportional to the original drawings. In our experiment, no hypotheses are associated with the faces; we used the drawings without pupils as an introduction to the study of the schematic eyes.

Fig. J7. Stimuli with faces used at the beginning of Replication of Study 4 in Experiment 2.

The results showed that the angry face evoked a higher pupil diameter change than the happy face, as depicted in Fig. J8. Because we did not have any hypothesis for the faces, no statistical tests are reported.

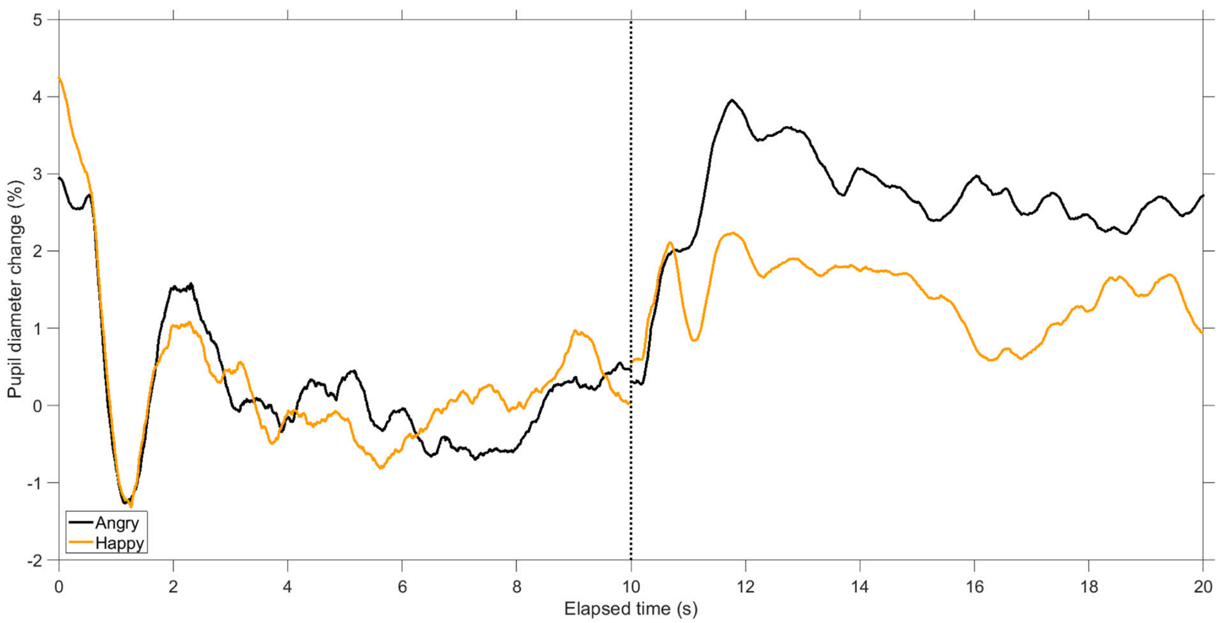


Fig. J8. Mean pupil diameter change (PC_t) for the angry and happy face in Experiment 2 with respect to the preceding control slide. A positive value indicates pupil dilation; a negative value indicates pupil constriction. The dotted vertical line indicates the moment of transition from the control slide to the stimulus slide.

Appendix K. Hess's instructions to participants

Hess and Polt (1960) did not mention what instructions were given to the participants. This omission was also pointed out by Woodmansee Jr. (1965): "*Unfortunately, Hess has never spelled out adequately the methodological details of his recent experiments*" (p. 10), and by Gilinsky in her letter to the *Science* editors. Hess's response to Gilinsky was: "*The subjects were given absolutely no instructions at all except to sit down and look at the pictures*" (Fig. B1). Hess (1972) stated: "*The subject is brought to the apparatus and given instructions as follows:* '*We would like to have you place your head so that you can comfortably see the numbers on this slide which we are showing. In a few moments we will show you a number of pictures. Each picture will be preceded by a control slide just like this one. Please look at each control slide when it appears by following the numbers 1, 2, 3, 4, and 5. The experimenter will pace you at first, and then you will follow the same procedure for each control slide. When the pictures come on, you of course look where you please. Do not look into the light or at the wall of the apparatus, etc. The entire run will take only a few minutes so that even if you are not too comfortable after the session begins, please try to keep your head in the exact position into which the experimenter has helped you to place it*'" (p. 229). In the archive, we found similar instructions, one dated 1963 (Box *M4143*, unlabeled folder) (Fig. K1) and another in an undated document (Box *M4138*, Folder *Portable Pupil Apparatus*) (Fig. K2). In summary, according to the material collected from the archive, participants did not receive specific instructions other than to look at the numbers of the control slide and to look where they wanted on the stimulus slide.

itself. In this way we kn w for certain when the subject run is completed. The instructions to the subject should be as f llows: " We would like to have you place your head so that you can confortably see the numbers on this slide which we are showing. In a few moments we will show you a number of pictures. Each picture will be preceded by a control slide just like this one. Please l ok at each control slide when it appears by following the numbers 1, 2, 3, 4, and 5. The experimenter will pace you at first, and then you will follow the same procedure f r each control slide. When the pictures come on, you of course look where you please. D not look anywhere except at the screen during this session - that is, do n t look into the light or look at the wall of the apparatus, etc. The entire run will nly take a few minutes, so that even if you are not too comfortable after the session begins, please try to keep your head in the exact position into which the experimenter has helped you place it ". That is the extent of the instructions.

Fig. K1. Task instruction for participants, dated 1963. Box *M4143*, unlabeled folder.

If you look xxxthex through the opening (have subject look) you'll see a series of numbers 1-2-3-4-5. Every other slide that you see will be this series of numbers. Each time they come on the screen, I wxxld wxxkxx would like to have you look at them in order. One, two three, four and then keep looking at the five until the slide changes. The slides that don't have the numbers on them, you can look at any way you want to.

Fig. K2. Task instruction for participants, undated. Box *M4138*, Folder *Portable Pupil Apparatus*.

Appendix L. Heatmaps of 'Female' stimuli and gender differences in local darkness, Images of the five themes (Experiments 1 and 2)

Fig. L1 shows the heatmaps of the eye-gaze coordinates of female versus male participants when viewing the two 'Female' stimuli in Experiment 1. Fig. L2 shows the corresponding data for Experiment 2.

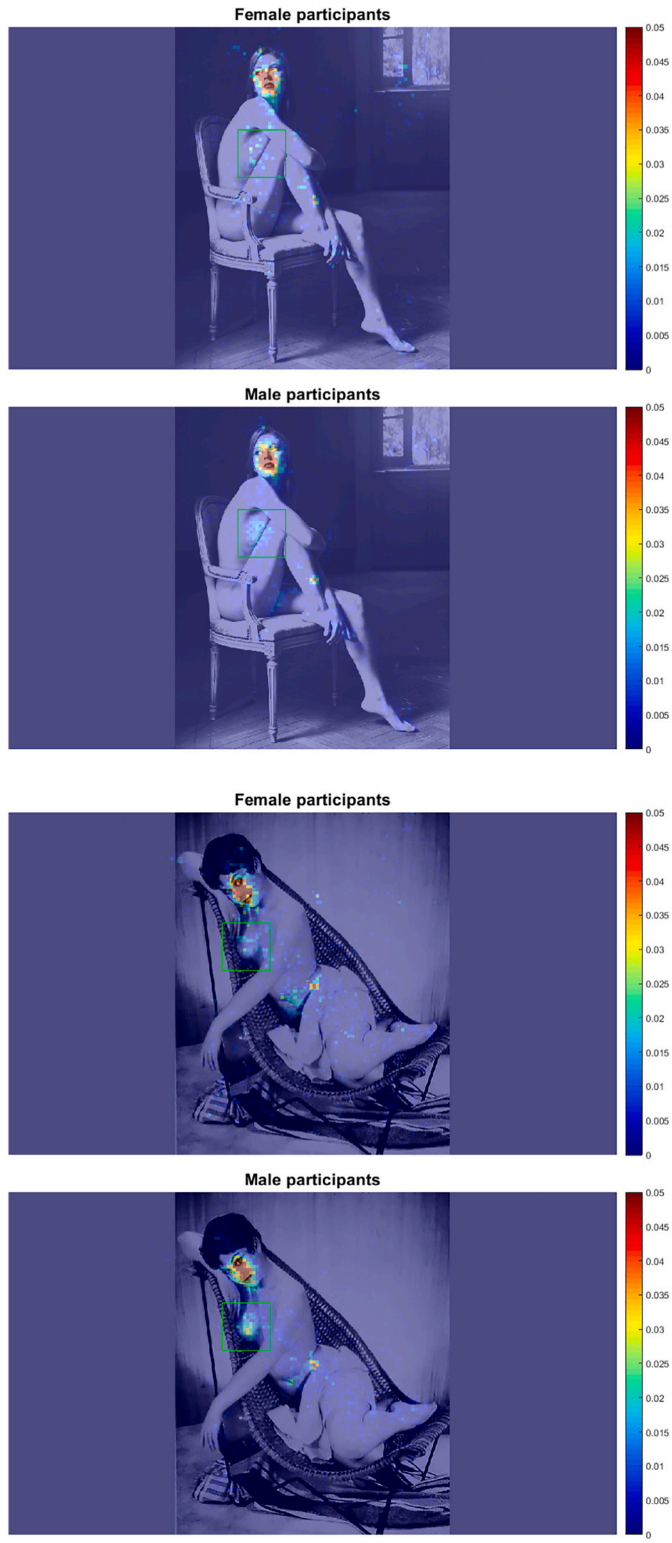


Fig. L1. Heatmaps of the eye-gaze coordinates for the 'Female – Modern' and 'Female – Original' images for female and male participants in Experiment 1. The 1920 × 1080-pixel (32.5 × 18.6°) image was divided into 10 × 10-pixel (0.2 × 0.2°) squares. The color-coding represents the number of seconds the eyes were looking at that square, averaged over the participants. The sum of all pixels equals the viewing time of 10 s. The area of interest is a 150 × 150-pixel (2.6 × 2.6°) square.

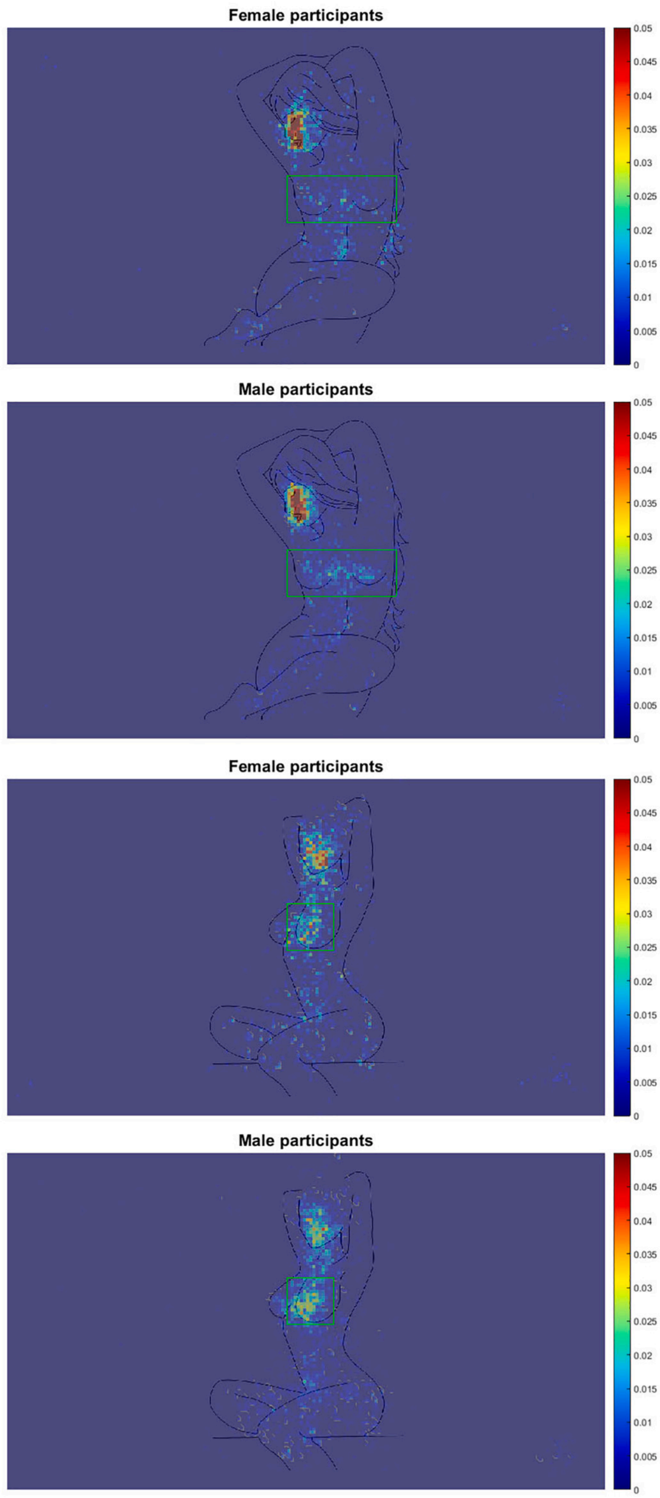


Fig. L2. Heatmaps of the eye-gaze coordinates the 'Female 1' and 'Female 2' line drawings for female and male participants in Experiment 2. The 1920 × 1080-pixel (33.3 × 19.1°) image was divided into 10 × 10-pixel (0.2 × 0.2°) squares. The color-coding represents the number of seconds the eyes were looking at that square, averaged over the participants. The sum of all pixels equals the viewing time of 10 s. The area of interest is a 350 × 150-pixel (6.2 × 2.7°) rectangle for the 'Female 1' drawing and a 150 × 150-pixel (2.7 × 2.7°) square for the 'Female 2' drawing.

Table L1 shows the mean local darkness for the images of five themes in Experiment 1 between 11 s and 19 s ($LD_{[1,9]}$) for males and females separately, as well as the results of statistical comparisons of local darkness between males and females.

Table L1
Means (standard deviations in parentheses) of mean local darkness between 11 s and 19 s ($LD_{[1,9]}$, %), and results of independent-samples *t*-tests, for the images of five themes in Experiment 1.

Stimulus	Females	Males	*t*(180)	Cohen's *d*	*p*
Baby – Modern	57.83 (2.97)	57.60 (3.04)	0.47	0.08	0.641
Baby – Original	59.12 (4.25)	57.75 (4.31)	1.95	0.32	0.053
Female – Modern	26.76 (8.38)	27.72 (8.68)	−0.68	−0.11	0.497
Female – Original	48.14 (7.80)	46.83 (7.32)	1.08	0.18	0.283
Landscape – Modern	55.54 (4.79)	56.18 (6.10)	−0.68	−0.11	0.497
Landscape – Original	54.60 (4.54)	56.99 (5.18)	−2.92	−0.48	**0.004**
Male – Modern	34.49 (6.68)	32.66 (7.61)	1.53	0.25	0.129
Male – Original	36.09 (7.18)	31.75 (6.95)	3.79	0.62	**<0.001**
Mother and baby – Modern	51.29 (5.34)	51.34 (5.73)	−0.06	−0.01	0.955
Mother and baby – Original	46.70 (7.27)	47.80 (7.18)	−0.93	−0.15	0.351

Note. Statistically significant *p*-values are indicated in boldface.

Appendix M. Additional results for the twelve multiplications in Experiment 1

Table M1 provides the results for the twelve multiplications in Experiment 1, for the 65 participants with complete data (i.e., by selecting only the 'slow' participants who did not press the spacebar within 3 s). The effects for $PC_{[ans\text{-}2.5,ans]}$ and PC_{ans} are not significant anymore, which can be explained by the fact that at least 3 s of data are included for all 65 participants; hence the pupil diameter data are less susceptible to the artifacts described in our paper.

Table M1
Means (standard deviations in parentheses) for four measures of pupil diameter change (%), and results of tests of within-subject linear contrasts, for the 65 participants with available pupil-diameter data at 3 s for all 12 calculations.

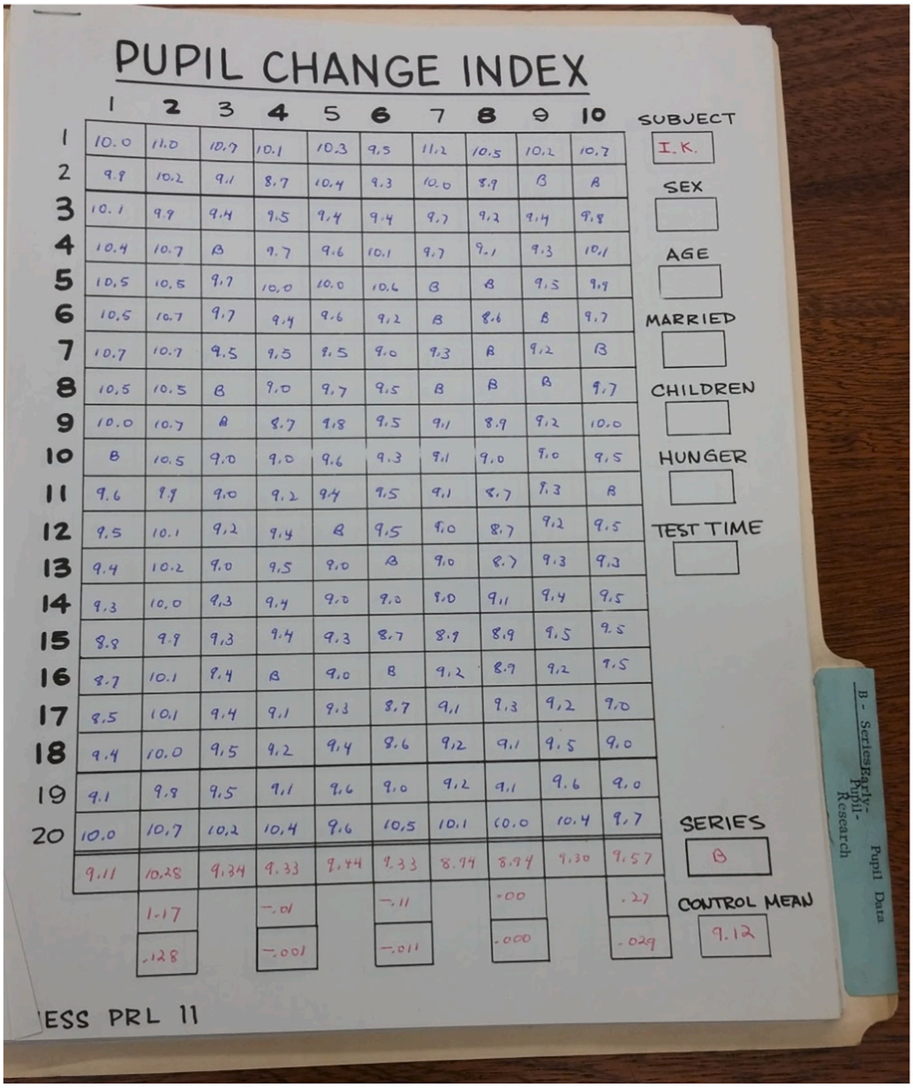
Multiplication	$PC_{[ans\text{-}2.5,ans]}$	PC_{max}	PC_{ans}	PC_3
9 × 8	9.92 (8.15)	15.90 (8.34)	10.97 (8.89)	10.70 (7.65)
6 × 7	10.67 (8.16)	16.40 (8.48)	10.47 (9.64)	9.72 (7.03)
7 × 8	12.13 (7.55)	17.39 (7.87)	12.52 (7.93)	9.91 (6.63)
6 × 16	8.99 (7.17)	14.98 (8.35)	10.33 (8.52)	8.57 (6.56)
8 × 13	11.09 (7.69)	16.52 (8.03)	11.36 (7.82)	8.39 (7.35)
7 × 14	11.36 (5.18)	17.29 (5.80)	11.50 (6.08)	7.92 (5.38)
9 × 17	12.60 (7.77)	18.02 (8.15)	12.28 (7.68)	8.62 (6.59)
12 × 14	10.33 (7.66)	15.93 (7.98)	10.84 (7.27)	7.55 (6.15)
13 × 14	11.69 (7.30)	18.34 (8.54)	11.43 (8.27)	8.22 (6.13)
15 × 17	11.51 (8.60)	19.10 (8.39)	10.46 (7.99)	7.38 (5.72)
16 × 18	11.76 (7.14)	18.54 (7.29)	11.39 (8.04)	7.35 (5.51)
16 × 23	11.90 (8.40)	18.83 (8.29)	11.13 (9.20)	6.72 (6.50)
Tests of within-subject linear contrasts (7 × 8, 8 × 13, 13 × 14, 16 × 23)	$F(1,64) = 0.00$, $p = 0.984$, $\eta_p^2 = 0.00$	$F(1,64) = 2.20$, $p = 0.143$, $\eta_p^2 = 0.03$	$F(1,64) = 0.92$, $p = 0.341$, $\eta_p^2 = 0.01$	$F(1,64) = 10.8$, $p = 0.002$, $\eta_p^2 = 0.14$
Tests of within-subject linear contrasts (all 12 multiplications)	$F(1,64) = 2.65$, $p = 0.108$, $\eta_p^2 = 0.04$	$F(1,64) = 10.5$, $p = 0.002$, $\eta_p^2 = 0.14$	$F(1,64) = 0.00$, $p = 0.990$, $\eta_p^2 = 0.00$	$F(1,64) = 20.2$, $p < 0.001$, $\eta_p^2 = 0.24$

Appendix N. Reproduction of theresults in Hess and Polt (1960)

We retrieved the raw pupil diameter data per slide and frame from the B-series. These include the 6 participants from Hess and Polt (1960) and 10 other participants (7 females and 3 males), 2 of which (one male, one female) were tested twice. The latter 10 participants belong to a measurement series called "*Run May 3 and May 4*" (Box *M4138*, folder *EARLY Pupil Research*).

An example of a datasheet for one participant (IK) for slides 1–10 out of 30 is shown in Fig. N1. The columns indicate slide numbers, and the rows indicate frame numbers. Odd-numbered slides are the control slides, and even-numbered slides are the stimulus slides. A recording was taken every 0.5 s. The last row shows the mean pupil diameter of frames 3–18. The *B*s listed in cells represent blinks: "*Approximately 15 percent of the frames could not be scored because of blinking and eye movement which caused a blurring of the pupil*" (Box *M4146*, Folder *EYE MOVEMENT DATA*). The datasheets also show the difference between the average pupil diameter for the stimulus slide and the average pupil diameter for the previous control slide, and the proportion difference. The pupil diameter is expressed in arbitrary units, presumably because the pupil diameter was measured using a Percepto-Scope, which magnified the pupil with an unknown magnification factor (about 20 times: Hess et al., 1965; or 30 times: Box *M4138*, folder *EARLY Pupil Research*).

We typed down all 10,800 numbers (i.e., (16 participants + 2 repetitions) × 30 slides × 20 frames). For the C-series, only the mean pupil diameter data per slide per participant were available in the archive (Fig. N2).

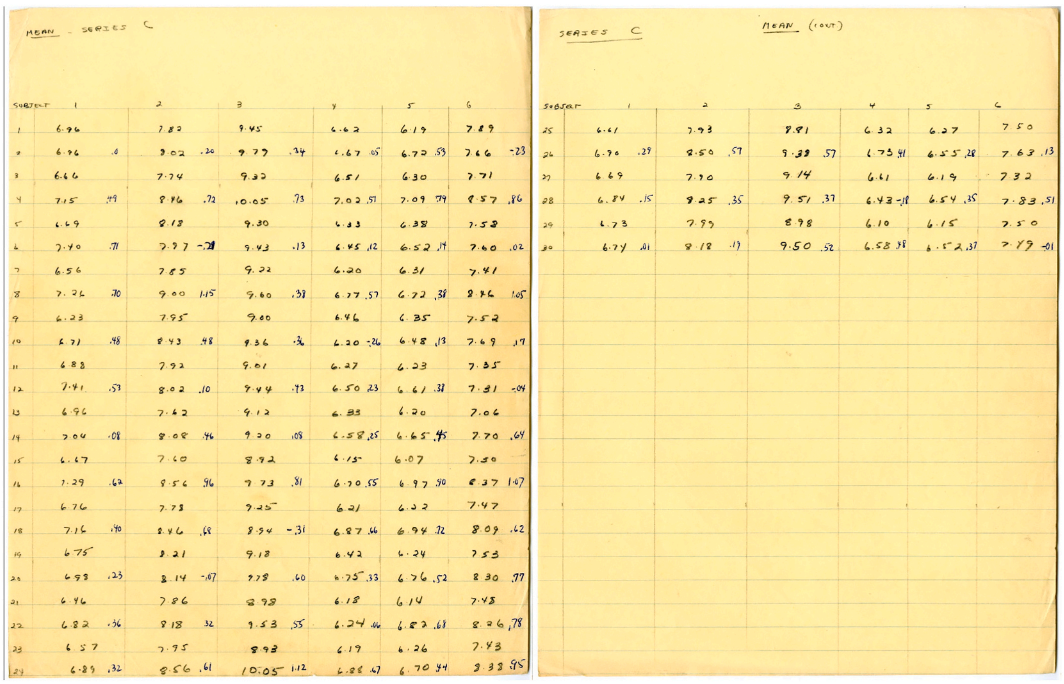


PUPIL CHANGE INDEX

	1	2	3	4	5	6	7	8	9	10
1	10.0	11.0	10.7	10.1	10.3	9.5	11.2	10.5	10.2	10.7
2	9.9	10.2	9.1	8.7	10.4	9.3	10.0	8.9	B	B
3	10.1	9.9	9.4	9.5	9.4	9.4	9.7	9.2	9.4	9.8
4	10.4	10.7	B	9.7	9.6	10.1	9.7	9.1	9.3	10.1
5	10.5	10.5	9.7	10.0	10.0	10.6	B	B	9.5	9.9
6	10.5	10.7	9.7	9.4	9.6	9.2	B	8.6	B	9.7
7	10.7	10.7	9.5	9.5	9.5	9.0	9.3	B	9.2	B
8	10.5	10.5	B	9.0	9.7	9.5	B	B	B	9.7
9	10.0	10.7	B	8.7	9.8	9.5	9.1	8.9	9.2	10.0
10	B	10.5	9.0	9.0	9.6	9.3	9.1	9.0	9.0	9.5
11	9.6	9.9	9.0	9.2	9.4	9.5	9.1	8.7	9.3	B
12	9.5	10.1	9.2	9.4	B	9.5	9.0	8.7	9.2	9.5
13	9.4	10.2	9.0	9.5	9.0	B	9.0	8.7	9.3	9.3
14	9.3	10.0	9.3	9.4	9.0	9.0	9.0	9.11	9.4	9.5
15	8.8	9.9	9.3	9.4	9.3	8.7	8.9	8.9	9.5	9.5
16	8.7	10.1	9.4	B	9.0	B	9.2	8.9	9.2	9.5
17	8.5	10.1	9.4	9.1	9.3	8.7	9.1	9.3	9.2	9.0
18	9.4	10.0	9.5	9.2	9.4	8.6	9.2	9.1	9.5	9.0
19	9.1	9.8	9.5	9.1	9.6	9.0	9.2	9.1	9.6	9.0
20	10.0	10.7	10.2	10.4	9.6	10.5	10.1	10.0	10.4	9.7
	9.11	10.28	9.34	9.33	9.44	9.33	8.94	8.94	9.30	9.57
		1.17		-.01		-.11		.00		.27
		.128		-.001		-.011		.000		.029

SUBJECT: I.K.
SEX
AGE
MARRIED
CHILDREN
HUNGER
TEST TIME
SERIES: B
CONTROL MEAN: 9.12
ESS PRL 11
B - Series Early-Pupil-Research Pupil Data

Fig. N1. Datasheet with pupil diameter data of one participant (IK) for slides 1–10 out of 30.
Source: Box *M4146*, Folder *EYE MOVEMENT DATA*.



MEAN - SERIES C

SUBJECT	1	2	3	4	5	6
1	6.96	7.82	9.45	6.62	6.19	7.89
2	6.96 .0	8.02 .20	9.79 .34	6.67 .05	6.72 .53	7.66 -.23
3	6.66	7.74	9.32	6.51	6.30	7.71
4	7.15 .49	8.46 .72	10.05 .73	7.02 .51	7.09 .79	8.57 .86
5	6.69	8.18	9.30	6.33	6.38	7.58
6	7.40 .71	7.97 -.21	9.43 .13	6.45 .12	6.52 .14	7.60 .02
7	6.56	7.85	9.22	6.20	6.31	7.41
8	7.26 .70	9.00 1.15	9.60 .38	6.77 .57	6.72 .38	8.46 1.05
9	6.23	7.95	9.00	6.46	6.35	7.52
10	6.71 .48	8.43 .48	9.36 .36	6.20 -.26	6.48 .13	7.69 .17
11	6.88	7.92	9.01	6.27	6.23	7.35
12	7.41 .53	8.02 .10	9.44 .43	6.50 .23	6.61 .38	7.31 -.04
13	6.96	7.62	9.12	6.33	6.20	7.06
14	7.04 .08	8.08 .46	9.20 .08	6.58 .25	6.65 .45	7.70 .64
15	6.67	7.60	8.92	6.15	6.07	7.50
16	7.29 .62	8.56 .96	9.73 .81	6.70 .55	6.97 .90	8.37 1.07
17	6.76	7.78	9.25	6.21	6.32	7.47
18	7.16 .40	8.46 .68	8.94 -.31	6.87 .66	6.94 .72	8.09 .62
19	6.75	8.21	9.18	6.42	6.24	7.53
20	6.98 .23	8.14 -.07	9.78 .60	6.75 .33	6.76 .52	8.30 .77
21	6.46	7.86	8.98	6.18	6.14	7.48
22	6.82 .36	8.18 .32	9.53 .55	6.24 .06	6.82 .68	8.26 .78
23	6.57	7.95	8.93	6.19	6.26	7.43
24	6.89 .32	8.56 .61	10.05 1.12	6.86 .67	6.70 .44	8.38 .95

SERIES C MEAN (CONT.)

SUBJECT	1	2	3	4	5	6
25	6.61	7.93	8.81	6.32	6.27	7.50
26	6.90 .29	8.50 .57	9.38 .57	6.73 .41	6.55 .28	7.63 .13
27	6.69	7.90	9.14	6.61	6.19	7.32
28	6.84 .15	8.25 .35	9.51 .37	6.43 -.18	6.54 .35	7.83 .51
29	6.73	7.99	8.98	6.10	6.15	7.50
30	6.74 .01	8.18 .19	9.50 .52	6.58 .48	6.52 .37	7.79 -.01

Fig. N2. Mean pupil diameter data of the C-series for the 6 participants in Hess and Polt (1960).
Source: Box *M4138*, Folder Early Pupil Research.

Table N1 shows the mean pupil diameter change with respect to the preceding control slide for males ($n = 4$) and females ($n = 2$) as obtained from the datasheets. The results match Hess's analyses that we found in the archive, see Fig. N3 for an example.

We were unable to determine how Hess and Polt (1960) computed the percentage area difference from their raw data. There are multiple ways of doing so, for example, by averaging at the aggregate level or at the level of individual participants. In the archive, all analyses we found were performed using pupil diameter instead of pupil area. Regardless, there is a strong correlation between the gender differences in pupil diameter as calculated based on the raw data from the archive (0.34, −0.15, 0.37, 0.41, −0.26) and the gender differences in pupil area reported in Hess and Polt (17%, −8%, 13%, 20%, −13%), Spearman's $\rho = 0.90$, Pearson's $r = 0.99$ ($n = 5$ stimuli) (Fig. N4). These findings indicate that the results in Hess and Polt (1960) match the raw data in the archive.

In Table N1, we provide color-coding for three thematic categories: (1) 'Babies and baby animals' in blue, (2) 'Nude men' in green, and (3) 'Nude women' in orange, based on Hess's taxonomy (see Fig. N3). An important observation is that there were multiple stimuli from the same category. For example, participants were shown six nude female stimuli (C8, C16, C24, C26, C28, C30), but only C26 was presented in Hess and Polt (1960). Stimuli C26 and C30 were the only nude female stimuli that showed a stronger dilation for males than for females.

In summary, using datasheets retrieved from the archive, we were able to reproduce the results presented in Hess and Polt (1960) with substantial congruence. However, it also became clear that the results for the 5 stimuli presented in Hess and Polt (1960) were part of a series of 30 stimuli.

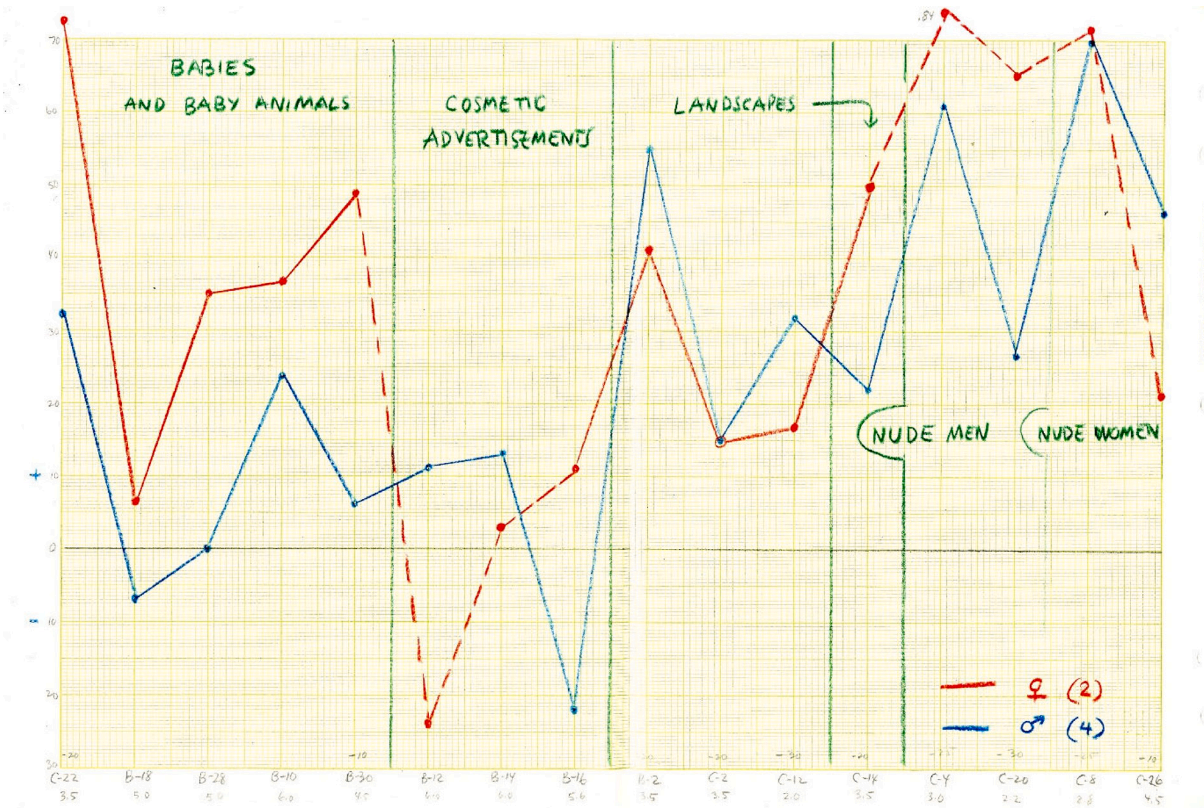


Fig. N3. Mean pupil diameter change with respect to the preceding control slide (a.u.) for selected stimuli clustered in thematic categories. A distinction is made between the mean of the four male participants (in blue) and the mean of the two female participants (in red). The labels on the horizontal axis correspond to the slide numbers of the B- and C-series. The numbers above these labels (e.g., −20, −10, …) refer to numerical corrections for brightness, which were not implemented in the graph.
Source: Box *M4166*, Folder *Early Pupil*.

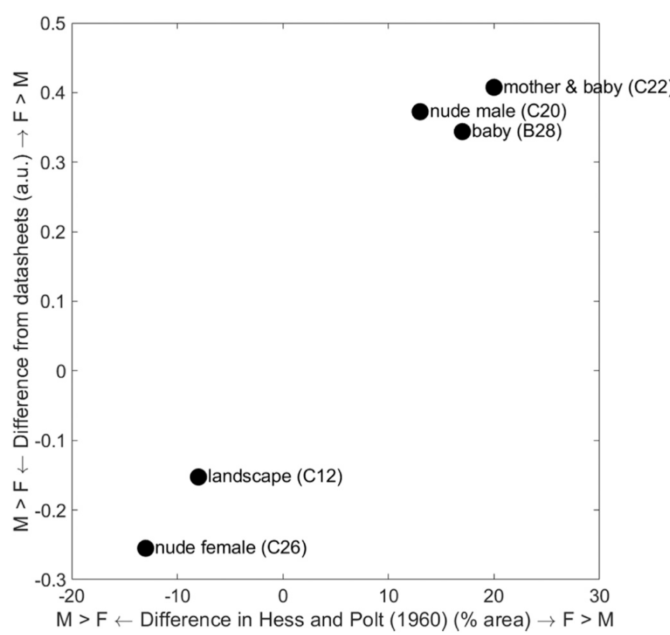


Fig. N4. Gender differences in pupil diameter as calculated from the datasheets versus gender differences in pupil area, as reported in Hess and Polt (1960). Note that Hess and Polt presented percentage values that were rounded to the nearest digit.

Table N1
Pupil diameter change values from the datasheets in the archive.

	Pupil diameter change with respect to control slide (a.u.), from datasheets		
Stimulus	**Males**	**Females**	**Difference**
B2	0.55	0.41	−0.14
B4	−0.27	0.09	0.35
B6	0.04	0.04	0.00
B8	0.03	0.23	0.20
B10	0.24	0.39	0.15
B12	0.07	−0.24	−0.30
B14	0.13	0.03	−0.10
B16	−0.22	0.11	0.32
B18	−0.07	0.06	0.12
B20	−0.18	0.08	0.26
B22	0.02	−0.23	−0.25
B24	0.10	0.06	−0.04
B26	0.09	0.21	0.12
B28	**0.00**	**0.35**	**0.34**
B30	0.06	0.49	0.43
C2	0.17	0.15	−0.02
C4	0.61	0.83	0.21
C6	0.19	0.08	−0.11
C8	0.70	0.73	0.03
C10	0.27	0.15	−0.11
C12	**0.32**	**0.17**	**−0.15**
C14	0.22	0.55	0.33
C16	0.74	0.98	0.25
C18	0.36	0.67	0.31
C20	**0.27**	**0.65**	**0.37**
C22	**0.32**	**0.73**	**0.41**
C24	0.69	0.70	0.01
C26	**0.46**	**0.21**	**−0.26**
C28	0.17	0.43	0.26
C30	0.30	0.18	−0.12

Note. Blue: babies and baby animals; green: nude men; orange: nude women.

Appendix O. Divisive versus subtractive baseline correction

For one of the studies, visually presented words of Experiment 2, we examined whether the results are affected by using subtractive instead of divisive baseline correction. The results for subtractive baseline correction, shown in Fig. O1, showed a pattern that is highly similar to the divisive baseline correction. A correlational analysis revealed a strong similarity between the pupil diameter change for the two baseline corrections (correlations around 0.99, see Fig. O2).

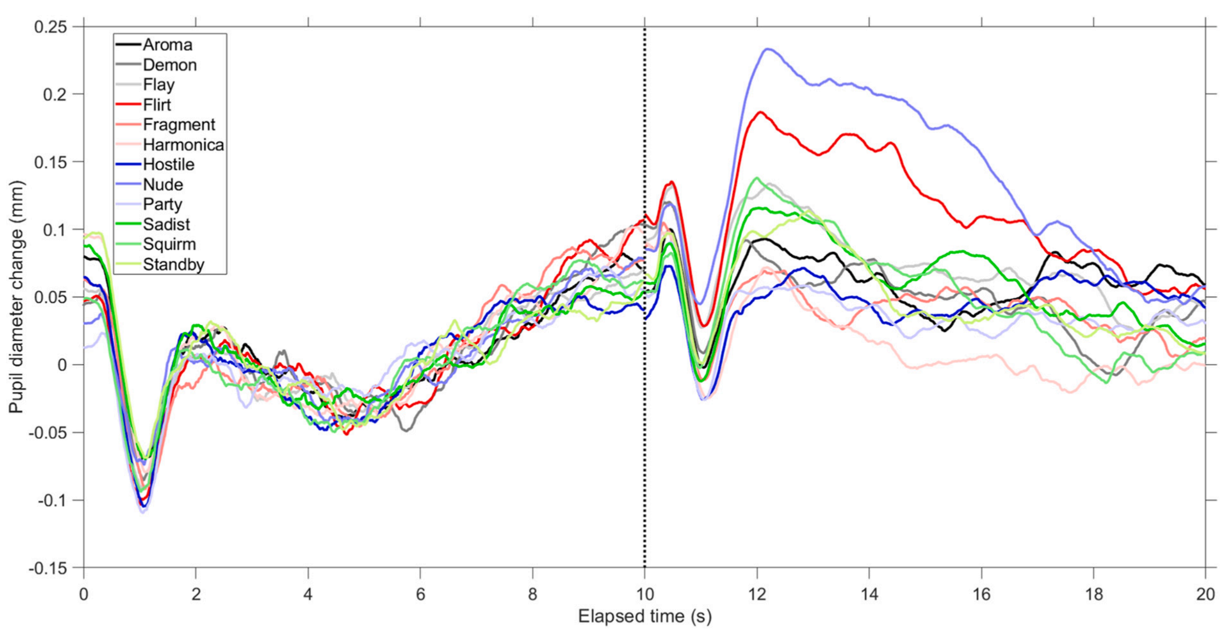


Fig. O1. Mean pupil diameter change (PC_t) for the words in Experiment 2 with respect to the preceding control slide. A positive value indicates pupil dilation; a negative value indicates pupil constriction. The dotted vertical line indicates the moment of transition from the control slide to the stimulus slide. Note that this figure contains the same results as Fig. 14, except that a subtractive instead of divisive baseline correction is used.

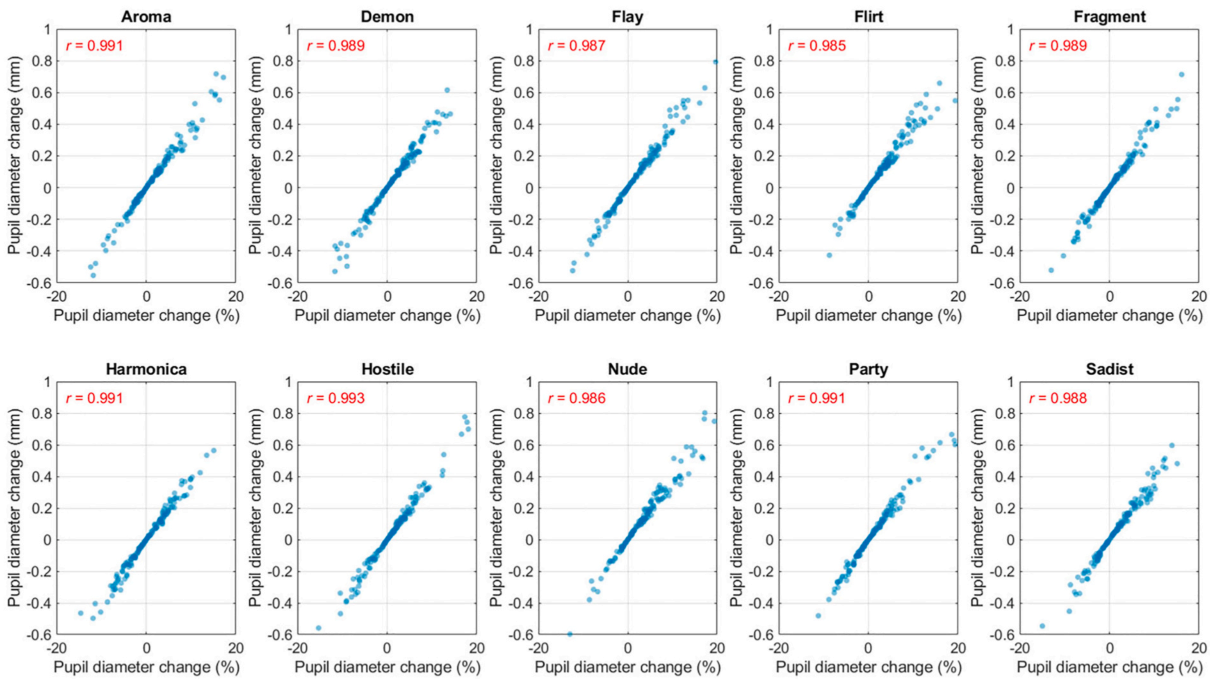


Fig. O2. Scatter plots of participants' ($N = 147$) $PC_{[1,9]}$ values for the ten visually presented words. The *x*- and *y*-axis show the pupil diameter change using divisive and subtractive baseline correction, respectively.

Appendix P. Length of baseline period

Except for the multiplications (which used a 2.5-s baseline period to exclude pupil recovery from the previous trial, cf. Fig. 7), our analyses used an 8-s period for normalizing the pupil diameter (see Eq. (1)). The recommended duration of the baseline period is a trade-off. If the baseline period is short, then too much noise may be captured. Pupil diameter is highly variable (i.e., pupillary hippus), and statistical power may be diminished if the baseline period is affected by this variability. On the other hand, if the baseline period is long, one may capture pupil diameter variance from the previous trial or other types of pupil diameter trends (e.g., due to learning effects, mood swings, variations in room lighting conditions) irrelevant to the research question. In our study, bias from the previous trial is ruled out because all stimuli within a study were presented in random order.

The choice of optimal length of the baseline period is a question that needs to be addressed empirically. We examined the effect of five baseline periods: 200, 1000, 2500, 5000, and 10,000 ms. More specifically, for one of the studies of Experiment 2 (Visually presented words), we divided the 10-s period of the control slide into fifty 200-ms periods, ten 1000-ms periods, four 2500-ms periods, two 5000-ms periods, or one 10,000-ms period. We then performed separate one-way repeated-measures ANOVAs with all 12 words as a factor. Fig. P1 shows the effect size, η_p^2. Two things can be noticed: (1) a longer baseline period results in a more robust η_p^2, and (2) a longer baseline period results in a higher η_p^2. These two effects can be explained because, at the individual level, pupil diameter is highly variable. In other words, a more statistically reliable estimate of the baseline pupil diameter is obtained when averaging across a longer time window. Higher reliability, in turn, can be expected to yield higher statistical power (e.g., Rushton et al., 1983).

Our observations are supported by a reliability analysis of the baseline pupil diameter. In this analysis, reliability is defined as the mean correlation of the participants' baseline pupil diameter in millimeters for all combinations of the 12 control slides. Thus, reliability is defined as the mean of 66 ($11 + 10 + 9 + 8 + 7 + 6 + 5 + 4 + 3 + 2 + 1$) correlation coefficients. The results in Fig. P2 show that pupil diameter reliability is higher for longer baseline periods. It can also be seen that pupil diameter is less reliable just after the previous stimulus slide.

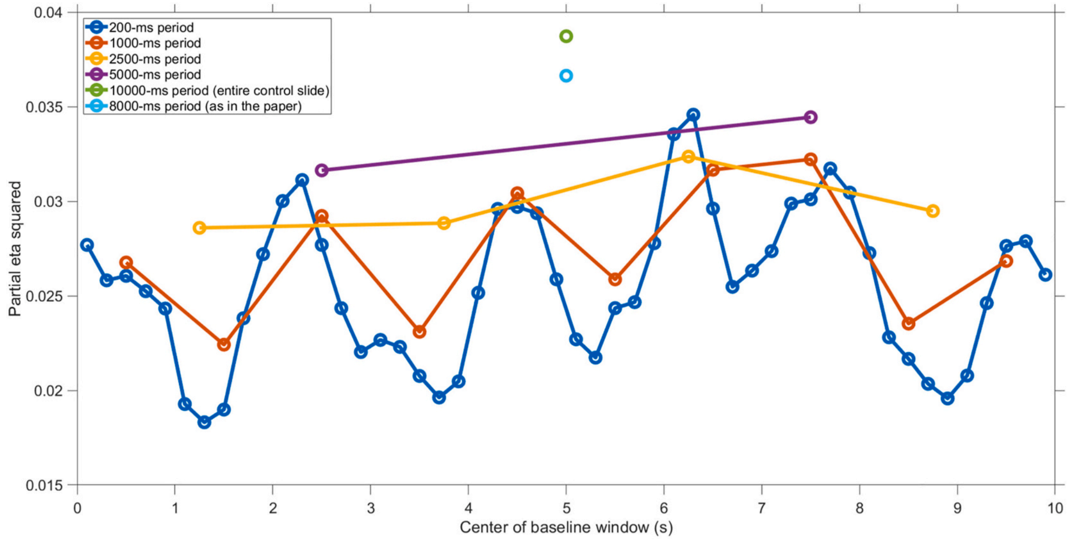


Fig. P1. Effect size η_p^2 as a function of the center of the baseline window for different lengths of the baseline period (200, 1000, 2500, 5000, 10,000 ms, as well as the 8000 ms used in the paper).

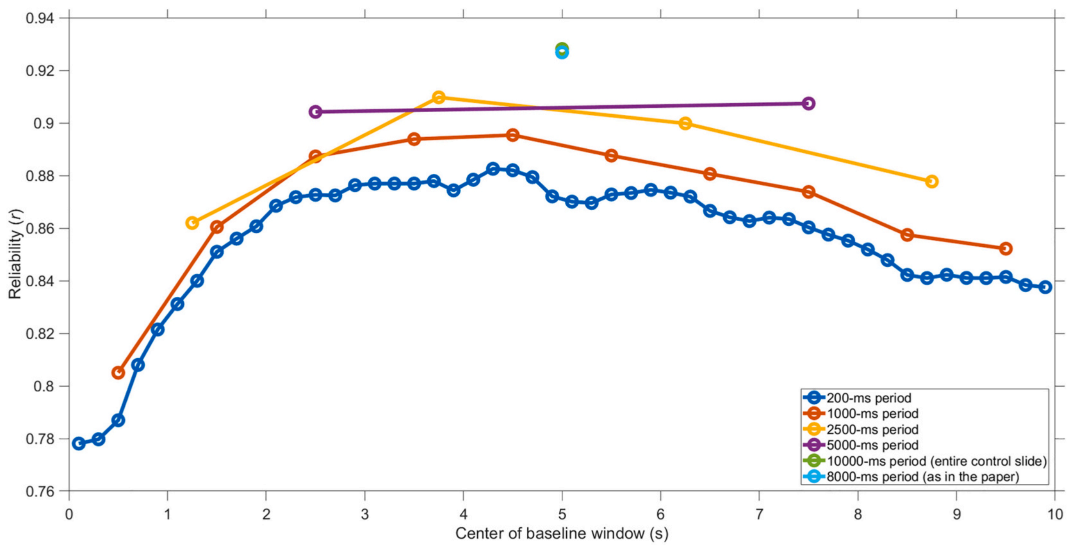


Fig. P2. Inter-trial reliability of baseline pupil diameter as a function of the center of the baseline window for different lengths of the baseline period (200, 1000, 2500, 5000, 10,000 ms, as well as the 8000 ms used in the paper).

Appendix Q. Effect of looking direction on pupil diameter

A possible validity threat in pupillometry research is the effect of viewing angle on the measurement of pupil size. The so-called 'pupil foreshortening error' (Gagl et al., 2011) refers to the fact that the pupil image becomes smaller and more elliptical when the participant gazes away from the camera.

We divided the screen into 100 × 100-pixel squares and assessed the pupil diameter data of all control slides per square. Pupil diameter data were only considered when at least 60 s of data were available, which was mostly around the nine digits on the control slides.

Fig. Q1 shows the number of available seconds per 100 × 100-pixel square, and Fig. Q2 shows the corresponding mean pupil diameter. It can be seen that pupil diameter was strongly affected when looking at the edges of the screen. The smallest pupil diameter was measured at the right top corner of the screen (3.78 mm), whereas the largest pupil diameter was recorded at the bottom of the screen (4.23 mm), reflecting a difference of about 10%. In our experiment, no stimuli were presented near the edges of the screen, and so we did not correct the pupil diameter for viewing angle, as was done, for example, by Hayes and Petrov (2016).

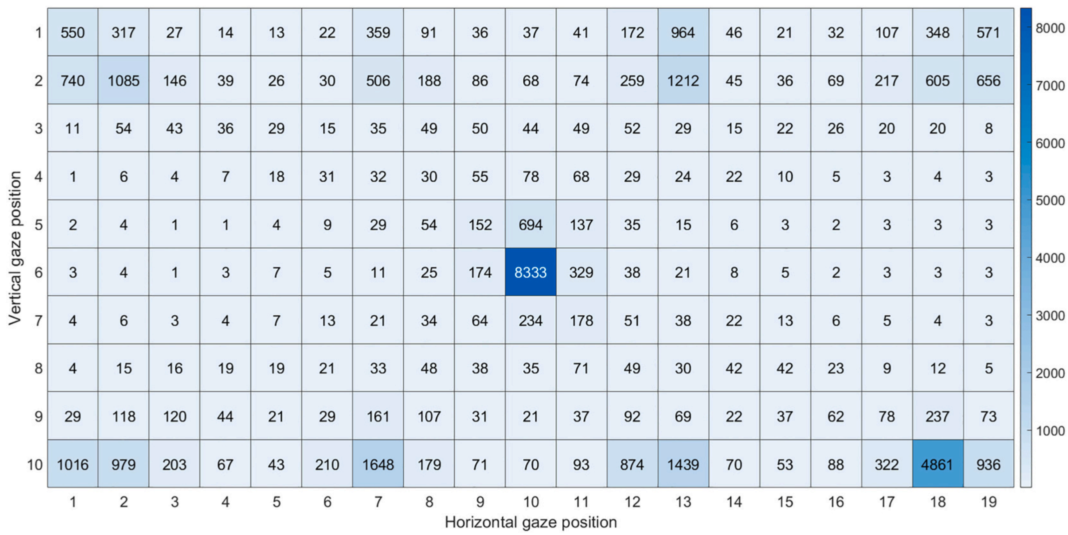


Fig. Q1. Number of seconds of available data for all control slides of all participants of Experiment 2 combined. The screen was divided into 100 × 100 squares.

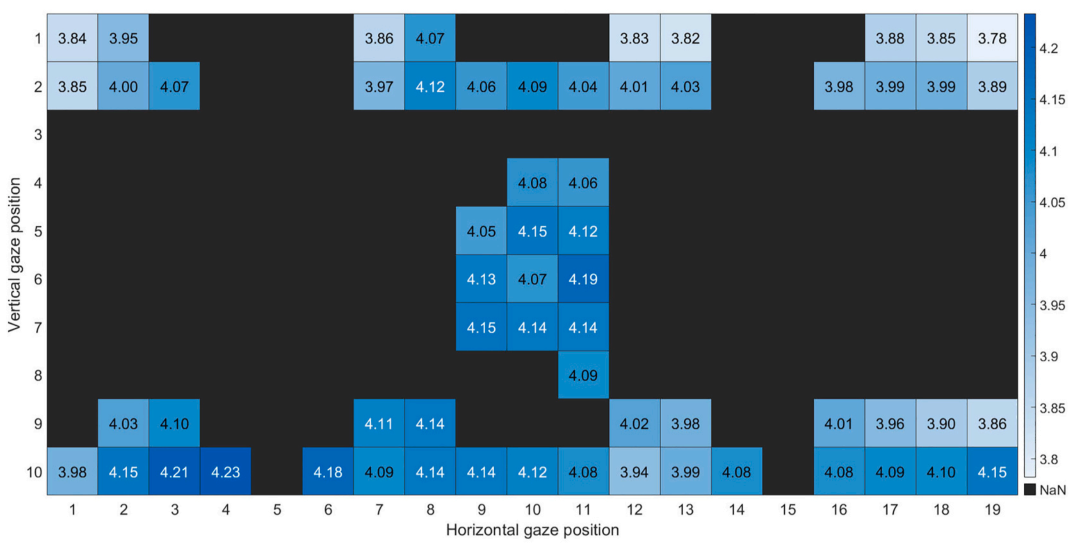


Fig. Q2. Mean pupil diameter for all control slides of all participants of Experiment 2 combined. The screen was divided into 100 × 100 squares. The pupil diameter is only shown for the squares for which at least 60 s of data were available.

Appendix R. Writing and submission

We retrieved the submission letter to *Science*, dated July 8, 1960: "*Enclosed is the manuscript about which I phoned you yesterday relative to the eye pupil studies and experiments currently being conducted at the University of Chicago. We sincerely trust you will find the report of interest and suitable for publication in an early issue of SCIENCE. We hope, too, that it can be treated with priority, if it is found to be acceptable. I am looking forward to seeing you in Chicago early next Fall*" (Box *M4167*, Folder *"Pupil Size as Related to Visual Stimuli" Original Paper submitted to SCIENCE* (*1960*)). We also retrieved the acceptance letter in the same folder: the paper was accepted 3 days later. The paper appeared in the issue of August 5, 1960.

This timeline of events is consistent with Hess's biography written 15 years later: "*I wasted no time getting to the laboratory, got out the manuscript, and mailed it off to Science. It was published within a matter of weeks, which is a tremendously short publication lag-the usual sequence being that one does a study, writes a paper, and then it may be in some instances a year, a year and a half, sometimes two years before the paper actually appears in print in a journal*" (Hess, 1975b, p. 16). Similarly, at a conference in 1973, Hess commented: "*That was in addition to getting the Science paper out, which I really wanted to get out to be on record. So Jim Polt and I put this thing together and send it in and got it published in a very short time, in a couple of weeks*" (transcript from the audio recording of a conference talk; Hess, 1973b).

Appendix S. Cooperation between Hess and Marplan

Publicly available information

Various publications provide indications that Hess had ties with an advertisement agency. Footnote 6 of Hess and Polt (1960) stated: "*Part of this work was carried out in the Perception Research Laboratory McCann-Erickson, Inc.*", whereas a footnote in Hess et al. (1965) reads: "*This research was supported in part by a grant from Social Sciences Research Committee of the University of Chicago and in part by Interpublic, New York*". West (1962) reported: “*the Marplan division set up its Perception Research Laboratory, and engaged Dr. Hess as consultant-director…. Two years ago, as part of its continuing program of basic research* (*in which the parent company has invested over $5 million during the last 17 years*)*, the Marplan division set up its Perception Research Laboratory, and engaged Dr. Hess as consultant-director. The unit was endowed with a grant to carry out basic research in the area of perception. No strings were attached; the direction of the research was left to Dr. Hess, working in concert with Russell Schneider, president of the Marplan division. Dr. Hess would be free to publish any scientific progress resulting from his research under the Interpublic grant. Interpublic would benefit from any commercial applications that might evolve under it. As consultant-director, Dr. Hess was to commute between Chicago and New York, supervising the work of a full-time professional staff in New York and coordinating research with another laboratory maintained at the university*” (p. 60). McCann-Erickson was an advertising agency and predecessor of the holding Interpublic Group founded in 1961. Marplan was the research laboratory of Interpublic. Van Bortel (1968), director of the Chicago office of Marplan, provided further information: “*Hess is also a consultant to MARPLAN and professional director of the Marplan Perception Research Laboratory, which is concerned with the commercial application of the techniques and procedures developed by Hess working under a grant for basic research sponsored by MARPLAN*” (p. 439).

Hess initially expressed reservations about the opportunity of working with Interpublic: “*I was asked, in the latter part of 1959, to help in setting up a perception laboratory for Interpublic, the second largest advertising and marketing organization in the United States. I did not agree to this without some soul searching.… there is no question in the minds of most academicians that they ‘know’ that they too, if only they were willing to prostitute themselves, could obtain a great deal of money from the advertising world or some other industrial organization*” (Hess, 1975b, pp. 159–160). Hess (1973b) offered further details about his cooperation with Interpublic: “*We also had obligations because all our research was funded by one organization; not the federal government, but Interpublic, which is a very large, I guess the second largest advertising company in the world. I was the director of their perception laboratory from 1959 to 1967, and we gathered a tremendous amount of data*”. In the same line, Herbert Krugman wrote in 1964: “*In 1960, Hess and Polt [1] reported finding a relationship between pupil dilation and the interest value of visual stimuli. Since then, over seventy studies utilizing measurement of changes in pupil diameter have been conducted by Marplan personnel on problems involving the evaluation of advertising materials, packages and products*” (H. E. Krugman, 1964a, p. 15). An even higher number of studies is mentioned in H. E. Krugman (1964b): “*In 1960, Hess and Polt reported finding a relationship between pupil dilation and the interest value of visual stimuli. Since then, over 100 studies utilizing measurement of changes in pupil diameter have been conducted by Marplan personnel*” (p. 27). Rice (1974) similarly commented: “*Within a few years Marplan was using the eye camera to gauge consumer reaction to everything from greeting cards to beer bottles to sterling-silver patterns. By the mid-1960s the company was pretesting magazine ads, package designs, TV pilot films, and TV commercials. At the peak of the boom, Marplan tested several commercials a week* (*at a cost of about $2,000 each*) *at field labs in shopping centers in Los Angeles, Chicago, New Jersey and Texas*” (p. 56). The magazine Sponsor (1964) illustrated the extent to which Hess’s pupil apparatus was used: “*There are portable eye cameras now in use in New York, Chicago, Los Angeles, Toronto, Mexico City, Sydney, London, San Paulo, Frankfurt, Johannesburg, and Tokyo*” (p. 28).

The ties with Marplan seem to have ended in the late 1960s. Krugman explained: “*In the late sixties a combination of reduced research budgets, controversy over the ‘directionality’ of the pupil, and changed personnel led to the gradual demise of pupil measurement at Marplan, although eye-tracking research lived on. My own view of this demise sees it more in terms of the trade secrecy which kept Dr. Hess’s elaborate and precise stimulus preparation procedures from being made available to others. Thus, when other enthusiasts attempted pupil measurement without adequate technology their results were bound to be contradictory with one another. This did not help the reputation of pupil research. Considering the unique financial investment which Interpublic made in such research it was understandable, however, that they should have sought competitive advantages and exclusive use of it*” (Krugman, 2013; p. 215). Hess (1975b) provided his view about the discontinuation of the cooperation with Interpublic: “*The careful controls which were possible in the laboratory situation apparently were not carried out. As a result of one study in which the outcome was not satisfactory to the client, Interpublic became disenchanted with the idea and Krugman recommended that this procedure no longer be used. I objected because it was obviously the most important way the pupil technique could be used to effectively determine and predict the advertising value of any material. I lost*” (pp. 189–190) and “*…most of my time, which was limited to a few days each month, was taken up in solving practical problems for the operation. Finally, a decision was made easier for me in terms of a way out when Mr. Harper left his position as chief of the operation in 1967. For me too, it was a good time to go*” (p. 188).

Information from the archive

Our search of the archive material confirmed that Hess had associations with Marplan and provided additional information on the scale of this cooperation.

Scale of cooperation

Based on the material we retrieved from the archive, we were able to confirm that the pupillometry activities at Marplan were of a large scale. We retrieved more than 90 reports of pupillometry studies conducted by Marplan between 1961 and 1969. These reports covered commercial products, including soft drinks, biscuits, beer, cereals, cake mixes, instant breakfast, underwear, toys, household insecticides, soap, pain relievers, fabric softeners, wood panels, gasoline, as well as TV series (see Table S1 in the Supplementary Material). It is not clear to what extent Hess was involved in these studies. In the archive, we retrieved pupil data, drafts, and progress reports on studies by Hess at/for Marplan. To illustrate, in a progress report on activities in November and December 1961 written by Hess with Chicago Perception Research Laboratory as affiliation, several studies are described: “*A study was carried out on Lucky Lager Beer. Thirty male subjects were tested with several perceptual and questionnaire techniques. Thirty female subjects were given perceptual tests to evaluate four Tidy House products….Relative to a new advertising campaign for Swift, a total of forty male and female subjects were tested in the laboratory….The most extensive study of the period involved sixteen Coca-Cola displays and three displays for other soft drinks…. Twenty male and twenty female were used in the study….For three of the above studies, the first and the last two, stimuli were prepared in the Chicago laboratory*” (Box *M4143*, Folder *PROGRESS RPT.*). We also retrieved correspondence from Hess to Marplan from December 1963, where he expresses his concerns about poor practices by operators running experiments at Chicago Marplan (Box *M4143*, Folder *PRESENTATION* (*COPY*)), indicating that he was monitoring the quality of the process.

Table S1
Overview of pupillometry studies conducted by Marplan, as retrieved from the archive.

Year	Title	Prepared for	Products/logos
1961	Proposal for Consumer Visual Research Study of Fountain Point of Purchase Advertising and Promotion Material	The Fountain Sales Department, The Coca-Cola Company	Drink dispenser
1962	An Exploratory Application of the Eye Camera to Point-of-Purchase Materials	The Fountain Sales Department, The Coca-Cola Company	Drink dispenser
1963	A Perceptual Pre-Test for Outdoor Posters	The Coca-Cola Company	Drink
1964	Taste Test Feasibility Project	The Coca-Cola Company	Drink
1964	Study of Consumers' Reactions to Different Promotional Ideas	The Coca-Cola Company	
1964	Perceptual Evaluation of Script and Print Versions of the Coke Logo	The Coca-Cola Company	Logo
1965	Study of the Relative Visibility of the "Floating Star" Sign	The Coca-Cola Company	Logo
1965	An Evaluation of Four Television Commercials "Brooks Robinson" "Parnelli Jones" "Soup and Sandwich" "Arnold Palmer"	The Coca-Cola Company	Logo
1965	Tachistoscopic Evaluation of Eight Cooler Facings	The Coca-Cola Company	Cooler facings
1965	Tachistoscopic Evaluation of Alternate Copy Formats	The Coca-Cola Company	Drink
1966	Perceptual Evaluation of Three Tab Commercials "Uninhibited" "Now Concept" "Olmstead"	The Coca-Cola Company	Dietetic drinks
1961	An Experimental Application of the Eye-Camera to Package Designs	The Nestle Company, Inc.	Cookie mix
1962	Proposal for a Survey of Attitudes Toward Alternative Premiums for Zip	The Nestle Company, Inc.	Syrup
1962	A Report on Advertising Research for the Nestle Company	The Nestle Company, Inc.	
1964	Summary of ASI Reports on the Ten Nabisco Television Commercials	Nabisco Biscuit Company	Biscuits
1965	Summary of ASI Reports on Fifteen Nabisco Television Commercials	Nabisco Biscuit Company	Biscuits
1962	Revised Proposal for Advertising Research on Del Monte Pineapple	McCann-Erickson, Inc.	Juice drink
1965	Perceptual Evaluation of the New Del Monte Whirly-Go-Round Label	California Packing Corporation	Juice drink
1966	Perceptual Evaluation of the New Brand Campaign For Del Monte	California Packing Corporation	Juice drink
1966	Perceptual Evaluation of the New General Line Campaign For Del Monte	California Packing Corporation	Juice drink
1964	Evaluation of New Orange Juice Package	The Minute Maid Company	Juice drink
1961	Proposal for Consumer Research on a Contemplated Package Revision	Lucky Lager Brewing Company	Beer
1962	Research on Proposed Package: 1. Interviews with Beer Drinkers 2. Perception Laboratory Tests	Lucky Lager Brewing Company	Beer
1965	Perceptual Evaluation of Eight Beer Cans	The Carling Brewing Company	Malt liquors, Stag beers
1966	Perceptual Evaluation of Two New Label Designs	Carling Brewing Limited	Beer
1966	Perceptual Evaluation of Three New Label Designs	Carling Brewing Limited	Beer
UN	Eye Camera Evaluation of Beer Label Designs	Heileman Brewing Co.	Beer
UN	Selecting a Package for Jaguar Malt Liquor		Malt liquor
1966	Perceptual Evaluation of Two Television Commercials Mayo "Lego Premium" Maltex "Professor Nutty"	The Fletcher Richards Company	Cereals
1966	Perceptual Evaluation of Two Television Commercials "Byrrh on the Rocks" "The Ballad of Snap-E-Tom"	The Fletcher Richards Company	Drinks
1964	A Perceptual Evaluation of Bread Wrappings: Ad-Seal-It Bands and End Labels	National Biscuit Company	Bread wrapping
1964	Addendum to A Perceptual Study of Bread Wrappings	National Biscuit Company	Bread wrapping
1965	Pastry Chef Cinnamon Coffee Cake Taste Test	Frozen Food Division National Biscuit Company	Cake mix
1965	An Evaluation of the Pastry Chef "What is Your Pleasure" Commercial	Frozen Food Division National Biscuit Company	Cake mix
1965	A Perceptual Evaluation of The Friskies "Meow" Commercial	The Carnation Company	Cat food
1965	Perceptual Evaluation of Six Leads for Carnation Instant Breakfast	The Carnation Company	Instant breakfast
1965	An Eye Camera Study of Two T.V. Commercials for "Carnation Evaporated Milk"	The Carnation Company	Evaporated milk
1965	A Perceptual Evaluation of Two Television Commercials "Oranges" and "Coffee"	The Carnation Company	Instant breakfast
1965	Perceptual Evaluation of Fifty-Two Television Commercials	The Interpublic Group of Companies	Carnation, coca-cola, esso, gillette
1966	Total Response Technique Evaluation of Television Commercials Carnation Instant Breakfast "Little Angel" "Good Morning World" "Family"	The Carnation Company	Instant breakfast
1966	An Evaluation of Three Television Commercials "Beach" "Race Track" "Hunter"	Promotion of the International Coffee Organization	Coffee
1964	Evaluation of Three Maxwell House Cans	The American Can Company	Cans
1962	Proposal for Package Research and Design	Modern Globe Underwear	Underwear
1963	A Perceptual Pre-Test of New Packaging	Modern Globe Sales, Inc.	Underwear
1964	Evaluation of Slips and Nightgowns	Warner Brothers Company	Underwear
1969	Total Response Technique Evaluation of Six Print Advertisements for Undergarments	The Warnaco Company	Underwear
1964	Evaluation of Four Toys	Multiple Products, Inc.	Toys
1964	A Perceptual Evaluation of Cans for Household Insecticides	The Geigy Chemical Corporation	Household insecticides
1964	Evaluation of New Labels for Spectracide and Sequestrene	Geigy Chemical Corporation	Lawn and garden care products
1964	A Perceptual Evaluation of Eleven Paper Plate Designs	The Dow Chemical Company	Paper plates
1965	An Evaluation of Three Television Commercials "Puzzled" "65 Products" "Runaway Cart"	The Borax Company	Detergent
1966	Perceptual Evaluation of Three Television Commercials "Fred, You're a Genius" Borateem "Two Products" Boraxo "Hand Clapping" Boraxo	U.S. Borax Corporation	Soap
1966	Perceptual Evaluation of Five Television Commercials "Fantasy" "Time Machine" "Bomb" "Good Housekeeping" "Black Light"	U.S. Borax Corporation	Soap
1964	Selection of a New Package for Smokers Drops*	The Warner Lambert Pharmaceutical Company	Smoker drops
1965	Pre-Sate Pilot Study	Warner-Chilcott Laboratory	Appetite suppressant
1966	An Evaluation of a Commercial for Alka-Seltzer Resolve – "Carousel Woman" Men vs. Women	Jack Tinker and Partners	Pain reliever

(*continued on next page*)

Table S1 (*continued*)

Year	Title	Prepared for	Products/logos
1966	Total Response Technique Evaluation of One Television Commercial Alka-Seltzer – “One minute to Five”	Jack Tinker and Partners	Pain reliever
1966	Total Response Technique Evaluation of One Television Commercial Focus – “Speedy”	Jack Tinker and Partners	Pain reliever
1968	Total Response Technique Evaluation of One Television Commercial Alka-Seltzer – “Getting Ready”	Jack Tinker and Partners	Pain reliever
1969	Total Response Technique Evaluation of One Television Commercial Vicks – “Nyquil”	Jack Tinker and Partners	Cold and flu medicine
1969	Total Response Technique Evaluation of One Television Commercial Carnation Slender – “Reminiscence”	Jack Tinker and Partners	Low-calorie diet aid
1969	Total Response Technique Evaluation of One Television Commercial Carnation Slender – “Zippers”	Jack Tinker and Partners	Low-calorie diet aid
1966	Perceptual Evaluation of Ten Tupperware Products	The Tupperware Company	Household containers
1967	New Service Presentation		
1967	Total Response Technique Evaluation of One Television Commercial Noxzema – “Girl”	[Unilever]	Skin cleanser
1964	A Perceptual Test of Ads for the New Westinghouse Top Loading Automatic Washer	Westinghouse Electric Corporation	Washing machine
1965	An Evaluation of Westinghouse’s Marketing Communication Program in Scientific American Magazine	Westinghouse Electric Corporation	
1965	Perceptual Evaluation of the New Focus Package Design	Miles Laboratories, Inc.	
1965	Perceptual Evaluation of Ten Cover Designs	Ladies’ Home Journal	Magazine covers
1968	An Experimental Study to Evaluate the Effectiveness of the Good Housekeeping Seal	Good Housekeeping Magazine	Guarantee seal on products (pure vegetable oil, toothpaste, facial tissues)
1968	Total Response Technique Evaluation of One Television Commercial Simoniz Wax - “Pixie”		Floor polisher
1964	Interest Tract (TM) Of NuSOFT “Scuba Diver” and Final Touch “Sweater Girl”	NuSoft Account Service Group	Fabric softener
1966	An Evaluation of the Georgia Pacific “Wood Panel” Commercial	The Georgia Pacific Company	Wood panels
1962	An Experimental Application of the Eye Camera to Record Album Covers	Radio Corporation of America	Album covers
1963	A study of Six Pictures Under Consideration for Use in Promoting V.I.P.		Movies
UN	A Perceptual Pre-Test of Advertising for THE COURTSHIP OF EDDIE’S FATHER	Metro-Goldwyn-Mayer, Inc.	Sitcom
1962	An Experimental Application of the Eye Camera to Advertising for MUTINY ON THE BOUNTY	Metro-Goldwyn-Mayer, Inc.	Movie
1963	A Study of Pupil Dilation Response to Advertising for THE WHEELER DEALERS/OF HUMAN BONDAGE	Metro-Goldwyn-Mayer, Inc.	Movies
UN	A Research Program to Evaluate Pre-Testing of 10-Second Promos	The National Broadcasting Company	Promos
1967	Perceptual Evaluation of Three Pilot Films “Occasional Wife” “The Hero” “The Monkees”	The National Broadcasting Company	TV series
1965	An Evaluation of Four Television Commercials Scripto “Shaggy Dog” “Hands and Knees” “Boardroom” “Bic”	The Scripto Company	Pens/lighters
1966	Perceptual Evaluation of Three Television Commercials Scripto “VU Lighter/ Scriptip” Paper-mate “150” Scripto “Wodmaster Pen/One Thousand and One Lights”	The Scripto Company	Pens/lighters
1966	Perceptual Evaluation of the Scripto Dual Product Commercial “One Thousand and One Lights/Wordmaster Pen”	The Scripto Company	Pens/lighters
1965	An Evaluation of the “Live Without” Commercial	American Telephone & Telegraph	Phones
1965	An Evaulation (*sic*) of Four Television Commercials “A.T.&T.”, “Conway”, “Diller”, and “Jimenez”	Pacific Northwest Bell	Phones
1966	Perceptual Evaluation of Three Alternate Campaign Approaches	The Council on Theological Education	
1968	Perceptual Evaluation of the Pilot Film “Two Good Guys”	The Columbia Broadcasting Company	TV series
1969	Perceptual Evaluation of the Pilot Film “The Queen and I”	The Columbia Broadcasting Company	TV series
1963	A Study of Pupil Dilation Response to Service Station Design Elements	Standard Oil Company	Gas station
1966	An Evaluation of Three Extension Phone Television Commercials “Lady on the Bell” “Young Married” “A.T.&.T.” Commercial	McCann-Erickson, Inc. - Humble Oil & Refining Company	Phones
1969	Total Response Technique Evaluation of One Television Commercial Humble - “Snow”	McCann-Erickson, Inc. - Humble Oil & Refining Company	Gasoline
1969	Total Response Technique Evaluation of One Television Commercial Humble - “Dune Buggy”	Humble Oil & Refining Company	Gasoline
1963	A Progress Report on the Pupil Dilation Method of Perceptual Research	N/A	Various (32 studies are mentioned in the summary)
1963	Statistical Analysis of Pupil Response Data	N/A	Various (newspapers, cookies, packages, models, silverware)
1965	Motivation Research with Special Emphasis on the Use of Projective Technique	N/A	N/a
1967	Pupillometrics and Advertising	N/A	Various (study on methodological aspects such as reliability, validity, data stability)
1967	A Study of Brightness Control in Testing TV Commercials	N/A	N/a

University work versus commercial work

The cooperation with Interpublic involved strategic decisions about how to separate commercial work from university research. In a letter to the president of Marplan, from June 1962, Hess advised keeping basic research separate from applied work: “*‘Basic basic’ research utilizing perceptual material similar to, but not actually, advertisements and T.V. commercials, should be carried out at the University of Chicago laboratory. This will keep our* *enterprise 'clean' and prevent possible complications*" (Box *M4138*, Folder *To New York*). In a draft entitled "*Proposal for Activities July 1960–July 1961, Perception Research Laboratory McCann-Erickson, Inc.*", Hess suggested: "*Because it seems wise to acquire more data faster when a really promising lead is found I propose that we enlarge the present facilities in either one of two ways, largely to study pupil size change. Either by adding one person full-time to help in the secretarial, tabulating, and measuring procedures, or by setting up a production laboratory on the fifteenth floor. The former would allow us to run a limited number of advertisements through the procedure of plotting eye movements and evaluating interest value. In addition to our basic research, probably twenty subjects per month could be so tested on ten to twenty advertisements.* (*See sample of Coca-Cola ad*) *This procedure would add approximately one-third of the present operating cost of the laboratory*". In the same draft, he also elaborated on the aforementioned proposed expansion of facilities in terms of space, equipment, and personnel, concluding that: "*Estimate of production capacity is 20 subjects on 10 to 20 ads or posters per week*" (Box *M4143*, Folder *General*).

Possible commercialization of the pupil apparatus

In correspondence between Hess and Marplan, the possibility of patenting the pupil apparatus appeared. Hess did not want a pupilometer patent to his name due to university regulations and recommended to do that on the name of Jim Polt: "*There is absolutely no way…that this apparatus can be patented in my name. The University statutes state specifically: 'Neither the University nor any members of its staff shall retain ownership, management, or licensing responsibilities for patents resulting from research or other activities carried out at the University or with the aid of its faciflities facilities* (*sic*)'" (Box *M4143*, Folder *Advertising correspondence*). We have not retrieved any patent, patent application, or a draft thereof, but it seems plausible that the pupil apparatus was commercialized. A letter from Marplan dated 1964 reads, for example: "*1. The order for an 'Eye Camera' should be addressed to Dr. Eckhard H. Hess, 1151 East 56th Street, Chicago 37, Illinois. 2. The order should specify* (*1*) '*Eye Camera*'" *apparatus of the type developed by Dr. Hess for Marplan USA….3. The total cost, exclusive of crating and freight, $1,500* (*U.S. dollars*)" (Box *M4165*, Folder *Correspondence 1964*). In another letter from Marplan dated 1964, an increase in the price of the pupil apparatus is mentioned: "*I have been informed by Dr. Hess that he will have to increase the price of the eye-camera. He will charge $1,750 for an apparatus equipped to operate on 110 volt current--$1,825 to operate on 220 volt current. As I understand it, the price increase was necessary to cover additional labor costs*" (Box *M4165*, Folder *Correspondence 1964*).

Active cooperation seeking

Hess was also active in reaching out for new projects. For example, on 12 August 1960, one week after his *Science* paper was accepted, Hess contacted *Playboy* magazine with a suggestion for cooperation: "*We have just published a paper in SCIENCE….In our study we found that a picture of the type represented by your 'Playmate of the Month' series results in large pupil size increases in most men. What I would now like to do is to use a series of pictures of this sort and test a number of men subjects on their pupil responses to these pictures*" (Box *M4166*, Folder *Early Pupil*).

In summary, Hess extensively cooperated with Marplan. This work had a large scope and involved active consultancy ("*a few days each month*", Hess, 1975b, p. 188).

References

Aboyoun, D.C., Dabbs, J.M., 1998. The Hess pupil dilation findings: sex or novelty? Soc. Behav. Personal. Int. J. 26, 415–419. https://doi.org/10.2224/sbp.1998.26.4.415.

Ahern, S., Beatty, J., 1979. Pupillary responses during information processing vary with scholastic aptitude test scores. Science 205, 1289–1292. https://doi.org/10.1126/science.472746.

Andreassi, J.L., 1980. Psychophysiology: Human Behavior and Physiological Response. Oxford University Press, New York.

Attard-Johnson, J., Bindemann, M., 2017. Sex-specific but not sexually explicit: pupillary responses to dressed and naked adults. R. Soc. Open Sci. 4, 160963. https://doi.org/10.1098/rsos.160963.

Attard-Johnson, J., Bindemann, M., Ó Ciardha, C., 2017. Heterosexual, homosexual, and bisexual men's pupillary responses to persons at different stages of sexual development. J. Sex Res. 54, 1085–1096. https://doi.org/10.1080/00224499.2016.1241857.

Bayer, M., Sommer, W., Schacht, A., 2011. Emotional words impact the mind but not the body: evidence from pupillary responses. Psychophysiology 48, 1553–1561. https://doi.org/10.1111/j.1469-8986.2011.01219.x.

Boersma, F., Wilton, K., Barham, R., Muir, W., 1970. Effects of arithmetic problem difficulty on pupillary dilation in normals and educable retardates. J. Exp. Child Psychol. 9, 142–155. https://doi.org/10.1016/0022-0965(70)90079-2.

Bradley, M.M., Lang, P.J., 2015. Memory, emotion, and pupil diameter: repetition of natural scenes. Psychophysiology 52, 1186–1193. https://doi.org/10.1111/psyp.12442.

Bradley, M.M., Sapigao, R.G., Lang, P.J., 2017. Sympathetic ANS modulation of pupil diameter in emotional scene perception: effects of hedonic content, brightness, and contrast. Psychophysiology 54, 1419–1435. https://doi.org/10.1111/psyp.12890.

Bradshaw, J.L., 1968. Pupil size and problem solving. Q. J. Exp. Psychol. 20, 116–122. https://doi.org/10.1080/14640746808400139.

Brandt, M.J., IJzerman, H., Dijksterhuis, A., Farach, F.J., Geller, J., Giner-Sorolla, R., Van't Veer, A., 2014. The replication recipe: what makes for a convincing replication? J. Exp. Soc. Psychol. 50, 217–224. https://doi.org/10.1016/j.jesp.2013.10.005.

Chapman, L.J., Chapman, J.P., Brelje, T., 1969. Influence of the experimenter on pupillary dilation to sexually provocative pictures. J. Abnorm. Psychol. 74, 396–400. https://doi.org/10.1037/h0027493.

Derksen, M., Van Alphen, J., Schaap, S., Mathôt, S., Naber, M., 2018. Pupil mimicry is the result of brightness perception of the iris and pupil. J. Cogn. 1, 1–16. https://doi.org/10.5334/joc.34.

Dovey, D., 2014. The eyes have it: pupil size can predict whether or not you'll make a good or bad decision. Retrieved from. http://www.medicaldaily.com/eyes-have-it-pupil-size-can-predict-whether-or-not-youll-make-good-or-bad-decision-303990.

Eisma, Y.B., De Winter, J.C.F., 2020. How do people perform an inspection time task? An examination of illusions, learning, and blinking. J. Cogn. 3, 34. https://doi.org/10.5334/joc.123.

Eisma, Y.B., Cabrall, C.D.D., De Winter, J.C.F., 2018. Visual sampling processes revisited: replicating and extending Senders (1983) using modern eye-tracking equipment. IEEE Trans. Hum. Mach. Syst. 48, 526–540. https://doi.org/10.1109/thms.2018.2806200.

Finke, J.B., Deuter, C.E., Hengesch, X., Schächinger, H., 2017. The time course of pupil dilation evoked by visual sexual stimuli: exploring the underlying ANS mechanisms. Psychophysiology 54, 1444–1458. https://doi.org/10.1111/psyp.12901.

Gagl, B., Hawelka, S., Hutzler, F., 2011. Systematic influence of gaze position on pupil size measurement: analysis and correction. Behav. Res. Methods 43, 1171–1181. https://doi.org/10.3758/s13428-011-0109-5.

Garrett, J.C., Harrison, D.W., Kelly, P.L., 1989. Pupillometric assessment of arousal to sexual stimuli: novelty effects or preference? Arch. Sex. Behav. 18, 191–201. https://doi.org/10.1007/BF01543194.

Goldwater, B.C., 1972. Psychological significance of pupillary movements. Psychol. Bull. 77, 340–355. https://doi.org/10.1037/h0032456.

Greenfield, P.M., 2017. Cultural change over time: why replicability should not be the gold standard in psychological science. Perspect. Psychol. Sci. 12, 762–771. https://doi.org/10.1177/1745691617707314.

Hall, W. J. (1959). *Slide projector* (Patent number 3,023,669). United States Patent Office.

Hayes, T.R., Petrov, A.A., 2016. Mapping and correcting the influence of gaze position on pupil size measurements. Behav. Res. Methods 48, 510–527.

Henderson, R.R., Bradley, M.M., Lang, P.J., 2018. Emotional imagery and pupil diameter. Psychophysiology 55, e13050. https://doi.org/10.1111/psyp.13050.

Hess, E.H., 1965. Attitude and pupil size. Sci. Am. 212, 46–55. https://doi.org/10.1038/scientificamerican0465-46.

Hess, E.H., 1968. Pupillometric assessment. In: Shlien, J.M. (Ed.), Research in psychotherapy, vol. 3. American Psychological Association, Washington D.C., pp. 573–583. https://doi.org/10.1037/10546-031

Hess, E.H., 1969. Ethological implications of the pupillary response. In: Summaries of the XIth International Ethological Conference (Rennes, France).

Hess, E.H., 1972. Pupillometrics. A method of studying mental, emotional, and sensory processes. In: Greenfield, N.S., Sternbach, R.A. (Eds.), Handbook of Psychophysiology. Holt, Rinehart and Winston, Inc., pp. 491–531

Hess, E.H., 1973a. Some new developments in pupillometrics. In: Dodt, E., Schrader, K.E. (Eds.), Die normale und die gestörte Pupillenbewegung/Normal and disturbed pupillary movements. Symposion der Deutschen Ophthalmologischen Gesellschaft (vom 10.–12. März 1972 in Bad Nauheim). Verlag J. F. Bergmann, München, pp. 246–262. https://doi.org/10.1007/978-3-642-80488-5_31.

Hess, E.H., 1973b. The Present State of Pupillometrics. Talk given at the University of Manitoba Symposium on Pupillometry. Winnipeg, Manitoba [audio recording].

Retrieved from the Drs. Nicholas and Dorothy Cummings Center for the History of Psychology. University of Akron, Ohio.
Hess, E.H., 1973c. What People Know about the Size of Eye Pupils (Paper presented at the Eighth Colloquium on the Pupil. Detroit, MI).
Hess, E.H., 1975a. The role of pupil size in communication. Sci. Am. 233, 110–119. https://doi.org/10.1038/scientificamerican1175-110.
Hess, E.H., 1975b. The Tell-tale Eye. How Your Eyes Reveal Hidden Thoughts and Emotions. Van Nostrand Reinhold Company.
Hess, E.H., Goodwin, E., 1974. The present state of pupillometrics. In: Janisse, M.P. (Ed.), Pupillary Dynamics and Behavior. Springer, Boston, MA, pp. 209–248. https://doi.org/10.1007/978-1-4757-1642-9_8.
Hess, E.H., Petrovich, S.B., 1987. Pupillary behavior in communication. In: Siegman, A. W., Feldstein, S. (Eds.), Nonverbal Behavior and Communication, 2nd ed. Lawrence Erlbaum Associates, Hillsdale, New Jersey and London, pp. 327–348.
Hess, E.H., Polt, J.M., 1960. Pupil size as related to interest value of visual stimuli. Science 132, 349–350. https://doi.org/10.1126/science.132.3423.349.
Hess, E.H., Polt, J.M., 1964. Pupil size in relation to mental activity during simple problem-solving. Science 143, 1190–1192. https://doi.org/10.1126/science.143.3611.1190.
Hess, E.H., Polt, J.M., 1966. Changes in pupil size as a measure of taste difference. Percept. Mot. Skills 23, 451–455. https://doi.org/10.2466/pms.1966.23.2.451.
Hess, E.H., Seltzer, A.L., Shlien, J.M., 1965. Pupil response of hetero-and homosexual males to pictures of men and women: a pilot study. J. Abnorm. Psychol. 70, 165–168. https://doi.org/10.1037/h0021978.
Hewig, J., Trippe, R.H., Hecht, H., Straube, T., Miltner, W.H., 2008. Gender differences for specific body regions when looking at men and women. J. Nonverbal Behav. 32, 67–78. https://doi.org/10.1007/s10919-007-0043-5.
Hope, J.A., Sherrill, J.M., 1987. Characteristics of unskilled and skilled mental calculators. J. Res. Math. Educ. 18, 98–111. https://doi.org/10.2307/749245.
Janisse, M.P., 1973. Pupil size and affect: a critical review of the literature since 1960. Can. Psychol. Psychologie Canadienne 14, 311–329. https://doi.org/10.1037/h0082230.
Janisse, M.P., 1974. Pupillometry: Some advances, problems and solutions. In: Janisse, M.P. (Ed.), Pupillary Dynamics and Behavior. Springer, Boston, MA, pp. 209–248. https://doi.org/10.1007/978-1-4757-1642-9_1.
Janisse, M.P., 1977. Pupillometry. In: The Psychology of the Pupillary Response. Hemisphere Publishing Corporation.
Klingner, J., Kumar, R., Hanrahan, P., 2008. Measuring the task-evoked pupillary response with a remote eye tracker. In: Proceedings of the 2008 Symposium on Eye Tracking Research & Applications. ACM, pp. 69–72. https://doi.org/10.1145/1344471.1344489.
Kooijman, L., Petermeijer, S.M., Eisma, Y.B., Dodou, D., De Winter, J.C.F., 2018. Replication of four studies of Eckhard Hess [preregistration]. Retrieved from. https://osf.io/8kbtc/download.
Kret, M.E., 2018. The role of pupil size in communication. Is there room for learning? Cognit. Emot. 32, 1139–1145. https://doi.org/10.1080/02699931.2017.1370417.
Krugman, H.E., 1964a. Some applications of pupil measurement. J. Mark. Res. 1, 15–19. https://doi.org/10.1177/002224376400100402.
Krugman, H.E., 1964b. Pupil Measurement at Marplan, November. The Brewers Digest, pp. 26–28.
Krugman, E.P., 2013. Consumer Behavior and Advertising Involvement: Selected Works of Herbert E. Krugman. Routledge, New York.
Laeng, B., Sirois, S., Gredebäck, G., 2012. Pupillometry: a window to the preconscious? Perspect. Psychol. Sci. 7, 18–27. https://doi.org/10.1177/1745691611427305.
Lewis, T., 2016. Exploring emotional contagion. Retrieved from. http://www.the-scientist.com/?articles.view/articleNo/46139/title/Exploring-Emotional-Contagion.
Loewenfeld, I.E., 1966. Comment on Hess' findings. Surv. Ophthalmol. 11, 291–294.
Loewenfeld, I.E., Lowenstein, O., 1993. The Pupil: Anatomy, Physiology, and Clinical Applications. Wiley-Blackwell.
Marquart, G., De Winter, J.C.F., 2015. Workload assessment for mental arithmetic tasks using the task-evoked pupillary response. PeerJ Comput. Sci. 1, e16 https://doi.org/10.7717/peerj-cs.16.
Martinez, V., 2015. Lying eyes, or something else? How blink rate, pupil dilation give insight to honesty and attraction. Retrieved from. https://www.medicaldaily.com/lying-eyes-or-something-else-how-blink-rate-pupil-dilation-give-insight-honesty-and-319164.
Mathôt, S., 2018. Pupillometry: psychology, physiology, and function. J. Cogn. 1, 16. https://doi.org/10.5334/joc.18.
Mathôt, S., Fabius, J., Van Heusden, E., Van der Stigchel, S., 2018. Safe and sensible preprocessing and baseline correction of pupil-size data. Behav. Res. Methods 50, 94–106. https://doi.org/10.3758/s13428-017-1007-2.
Mohammad, S.M., 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In: Proceedings of the 6th Annual Meeting of the Association for Computational Linguistics (Long Papers), 1, pp. 174–184. https://doi.org/10.18653/v1/P18-1017.
Nosek, B.A., Ebersole, C.R., DeHaven, A.C., Mellor, D.T., 2018. The preregistration revolution. Proc. Natl. Acad. Sci. 115, 2600–2606. https://doi.org/10.1073/pnas.1708274114.
Nuijten, M.B., Bakker, M., Maassen, E., Wicherts, J., 2018. Verify original results through reanalysis before replicating: a commentary on "Making replication mainstream" by Rolf A. Zwaan, Alexander Etz, Richard E. Lucas, & M. Brent Donnellan. Behav. Brain Sci. 41, e143. https://doi.org/10.1017/S0140525X18000791.
Nummenmaa, L., Hietanen, J.K., Santtila, P., Hyönä, J., 2012. Gender and visibility of sexual cues influence eye movements while viewing faces and bodies. Arch. Sex. Behav. 41, 1439–1451. https://doi.org/10.1007/s10508-012-9911-0.
Nunnally, J.C., Knott, P.D., Duchnowski, A., Parker, R., 1967. Pupillary response as a general measure of activation. Percept. Psychophys. 2, 149–155. https://doi.org/10.3758/BF03210310.
Paivio, A., Simpson, H.M., 1966. The effect of word abstractness and pleasantness on pupil size during an imagery task. Psychon. Sci. 5, 55–56. https://doi.org/10.3758/BF03328277.
Payne, D.T., Parry, M.E., Harasymiw, S.J., 1968. Percentage of pupillary dilation as a measure of item difficulty. Percept. Psychophys. 4, 139–143. https://doi.org/10.3758/BF03210453.
Peavler, W.S., McLaughlin, J.P., 1967. The question of stimulus content and pupil size. Psychon. Sci. 8, 505–506. https://doi.org/10.3758/BF03331723.
Polt, J.M., Hess, E.H., 1968. Changes in pupil size to visually presented words. Psychon. Sci. 12, 389–390. https://doi.org/10.3758/BF03331368.
Reilly, J., Kelly, A., Kim, S.H., Jett, S., Zuckerman, B., 2019. The human task-evoked pupillary response function is linear: implications for baseline response scaling in pupillometry. Behav. Res. Methods 51, 865–878. https://doi.org/10.3758/s13428-018-1134-4.
Reys, R.E., Reys, B.J., Nohda, N., Emori, H., 1995. Mental computation performance and strategy use of Japanese students in Grades 2, 4, 6, and 8. J. Res. Math. Educ. 26, 304–326. https://doi.org/10.2307/749477.
Rice, B., 1974. Rattlesnakes, French Fries and Pupillometric Oversell, February. Psychology Today, pp. 55–63.
Rieger, G., Cash, B.M., Merrill, S.M., Jones-Rounds, J., Dharmavaram, S.M., Savin-Williams, R.C., 2015. Sexual arousal: the correspondence of eyes and genitals. Biol. Psychol. 104, 56–64. https://doi.org/10.1016/j.biopsycho.2014.11.009.
Rushton, J.P., Brainerd, C.J., Pressley, M., 1983. Behavioral development and construct validity: the principle of aggregation. Psychol. Bull. 94, 18–38. https://doi.org/10.1037/0033-2909.94.1.18.
Sagal, B., 1965. Survival [television series episode]. Producer. In: Hamner, R. (Ed.), A Man Called Shenandoah. Bronze Enterprises, MGM Television. Director.
Schaefer, T., Ferguson, J.B., Klein, J.A., Rawson, E.B., 1968. Pupillary responses during mental activities. Psychon. Sci. 12, 137–138. https://doi.org/10.3758/BF03331236.
Scott, T.R., Wells, W.H., Wood, D.Z., Morgan, D.I., 1967. Pupillary response and sexual interest 85e-examined. J. Clin. Psychol. 23, 433–438. https://doi.org/10.1002/1097-4679(196710)23:4<433::AID-JCLP2270230408>3.0.CO;2-2.
Seitz, K., Schumann-Hengsteler, R., 2000. Mental multiplication and working memory. Eur. J. Cogn. Psychol. 12, 552–570. https://doi.org/10.1080/095414400750050231.
Siegle, G.J., Granholm, E., Ingram, R.E., Matt, G.E., 2001. Pupillary and reaction time measures of sustained processing of negative information in depression. Biol. Psychiatry 49, 624–636. https://doi.org/10.1016/S0006-3223(00)01024-6.
Simons, D.J., 2014. The value of direct replication. Perspect. Psychol. Sci. 9, 76–80. https://doi.org/10.1177/1745691613514755.
Skinner, N.F., 1980. The Hess, et al. study of pupillary activity in heterosexual and homosexual males: a re-evaluation. Perceptual and Motor Skills 51, 844. https://doi.org/10.2466/pms.1980.51.3.844.
Snowden, R.J., McKinnon, A., Fitoussi, J., Gray, N.S., 2019. Pupillary responses to static images of men and women: a possible measure of sexual interest? J. Sex Res. 56, 74–84. https://doi.org/10.1080/00224499.2017.1394959.
Sponsor, 1964, December. … in the eye of the beholder, 25–29. Sponsor. Retrieved from. https://worldradiohistory.com/Archive-Sponsor-Magazine/1964/Sponsor-1964-12-4.pdf.
SR Research, 2009. EyeLink® 1000 user manual. Version 1.5.0. Retrieved from. http://sr-research.jp/support/EyeLink%201000%20User%20Manual%201.5.0.pdf.
Stark, L., 1959. Stability, oscillations, and noise in the human pupil servomechanism. Proc. IRE 47, 1925–1939. https://doi.org/10.1109/JRPROC.1959.287206.
Stern, R.M., Ray, W.J., Quigley, K.S., 2001. Psychophysiological Recording. Oxford University Press, USA.
Windows of the soul/elixir of youth [television series episode]. P. Loizos & D. Gilling (Producers). In: Taylor, G.R. (Ed.), 1966. Horizon. BBC Two England.
Tronsky, L.N., 2005. Strategy use, the development of automaticity, and working memory involvement in complex multiplication. Mem. Cogn. 33, 927–940. https://doi.org/10.3758/BF03193086.
Van Bortel, F.J., 1968. Commercial applications of pupillometrics. In: Bass, F.M., King, C. W., Pessemier, E.A. (Eds.), Applications of the Sciences in Marketing Management. John Wiley and Sons, Inc., New York, London, Sydney, pp. 439–453.
Van Dale, 2019. Gratis Woordenboek [Free Dictionary]. Retrieved from. https://www.vandale.nl/opzoeken.
Van der Meer, E., Beyer, R., Horn, J., Foth, M., Bornemann, B., Ries, J., Wartenburger, I., 2010. Resource allocation and fluid intelligence: insights from pupillometry. Psychophysiology 47, 158–169. https://doi.org/10.1111/j.1469-8986.2009.00884.x.
Van der Wel, P., Van Steenbergen, H., 2018. Pupil dilation as an index of effort in cognitive control tasks: a review. Psychon. Bull. Rev. 25, 2005–2015. https://doi.org/10.3758/s13423-018-1432-y.

Watts, T.M., Holmes, L., Savin-Williams, R.C., Rieger, G., 2017. Pupil dilation to explicit and non-explicit sexual stimuli. Arch. Sex. Behav. 46, 155–165. https://doi.org/10.1007/s10508-016-0801-8.
West, D.V., 1962. In the Eye of the Beholder, April. Television Magazine, pp. 60–63. Retrieved from. https://worldradiohistory.com/Archive-Television-Magazine/Television-1962-Apr.pdf.
Woodmansee, J.J., 1966. Methodological problems in pupillographic experiments. In: Proceedings of the Annual Convention of the American Psychological Association. American Psychological Association, pp. 133–134.
Woodmansee Jr., J.J., 1965. An Evaluation of Pupil Response as a Measure of Attitude Toward Negroes (Doctoral dissertation). University of Colorado.
Zuckerman, M., 1971. Physiological measures of sexual arousal in the human. Psychol. Bull. 75, 297–329. https://doi.org/10.1037/h0030923.
Zwaan, R.A., Etz, A., Lucas, R.E., Donnellan, M.B., 2018. Making replication mainstream. Behav. Brain Sci. 41, e120 https://doi.org/10.1017/S0140525X17001972.