ORIGINAL PAPER

Measurement of Sexual Interests with Pupillary Responses: A Meta-Analysis

Janice Attard-Johnson[1] · Martin R. Vasilev[1] · Caoilte Ó Ciardha[2] · Markus Bindemann[2] · Kelly M. Babchishin[3]

Abstract

Objective measures of sexual interest are important for research on human sexuality. There has been a resurgence in research examining pupil dilation as a potential index of sexual orientation. We carried out a meta-analytic review of studies published between 1965 and 2020 (Mdn year $= 2016$) measuring pupil responses to visual stimuli of adult men and women to assess sexual interest. Separate meta-analyses were performed for six sexual orientation categories. In the final analysis, 15 studies were included for heterosexual men ($N = 550$), 5 studies for gay men ($N = 65$), 4 studies for bisexual men ($N = 124$), 13 studies for heterosexual women ($N = 403$), and 3 studies for lesbian women ($N = 132$). Only heterosexual and gay men demonstrated discrimination in pupillary responses that was clearly in line with their sexual orientation, with greater pupil dilation to female and male stimuli, respectively. Bisexual men showed greater pupil dilation to male stimuli. Although heterosexual women exhibited larger pupils to male stimuli compared to female stimuli, the magnitude of the effect was small and non-significant. Finally, lesbian women displayed greater pupil dilation to male stimuli. Three methodological moderators were identified—the sexual explicitness of stimulus materials, the measurement technique of pupillary response, and inclusion of self-report measures of sexual interest. These meta-analyses are based on a limited number of studies and are therefore preliminary. However, the results suggest that pupillary measurement of sexual interest is promising for men and that standardization is essential to gain a better understanding of the validity of this measurement technique for sexual interest.

Keywords Sexual interest · Sexual arousal · Sexual orientation · Meta-analysis · Pupil dilation

Introduction

The fact that the pupil of the human eye responds to changes in brightness levels within our visual environment is well known to most people. This is termed "the pupillary light reflex" and is responsible for an increase in pupil size (dilation) when the visual environment is darker and a decrease

---

in pupil size (constriction) in brighter settings (Ellis, 1981). In the twentieth century, a number of incidental observations of pupils responding to factors other than light, including interest and arousal, led to the pivotal research of Löwenstein and Loewenfeld. They developed the first technological method for measuring pupils, the electronic pupillograph, which captured motion pictures of the pupil and recorded their fluctuating movements (Löwenstein & Loewenfeld, 1958). With this approach, it became possible to conduct experimental investigations into the physiological and psychological factors underlying pupillary responses (Loewenfeld, 1958; Löwenstein & Loewenfeld, 1962). Over the next decade, researchers began to investigate an array of mental activities linked to pupil size changes, including memory load (Kahneman & Beatty, 1966), and general arousal activation (Hess & Polt, 1960; Nunnally et al., 1967).

Fluctuations in pupil size are controlled by the autonomic nervous system and are a result of dynamic shifts between sympathetic (results in dilation) and parasympathetic (results in constriction) activation (McDougal & Gamlin, 2008;

Zuckerman, 1971). A state of arousal leads to activation of the autonomic nervous system that is characterized by changes in skin conductance rate, blood pressure, breathing rate, and pupil dilation (Zuckerman, 1971). Pupil dilation can be evoked as a response to affective arousal, for example, during the viewing of highly pleasant (including sexually suggestive stimuli) and unpleasant stimuli (Bradley et al., 2008). Notably, the strength of the emotional state evoked by a stimulus (i.e., stimulus intensity) rather than the extent to which an emotion is positive or negative (i.e., stimulus valence) seems to elicit corresponding arousal responses in pupil size and change in skin conductance (Bradley et al., 2008). Sexually appetitive content, such as the viewing of erotic material or images depicting the observer's arousing sex, has also been shown to evoke pupil dilation compared to neutral content or images depicting the non-arousing sex (Rieger & Savin-Williams, 2012). For example, pupillary response and skin conductance correlate positively during the viewing of erotic movies and pictures (see, e.g., Bernick et al., 1971; Finke et al., 2017). The close correlation between skin conductance and pupil size during arousal suggests that pupillary responses are predominantly driven by direct activation of the sympathetic nervous system. This may be different to pupil dilation observed during other cognitive processes, for example, during information-processing and mental load, which appear to be mediated by parasympathetic inhibition (Bradley et al., 2008; Steinhauer et al., 2004).

Pupil Dilation as a Measure of Sexual Arousal/Interest

The ability to objectively measure sexual arousal and sexual orientation is important for research in understanding the theoretical basis of human sexuality. Existing measures of sexual orientation and sexual arousal include subjective self-report (e.g., Kinsey scale; Kinsey et al., 1948) and cognitive-based measures (e.g., IAT and Viewing Time; Babchishin et al., 2013; Lippa, 2012). However, these measures are recognized for their limitations; for example, subjective self-report would require the self-awareness of the individual as well as a willingness to respond truthfully, and there are debates around the precise mental processes that are being measured with cognitive-based tasks (Conaglen, 2004). Direct physiological measurement of sexual arousal is possible by recording genital responses (e.g., penile pleythsmography; Murphy et al., 2015). However, this is considered a highly invasive technique and may be of limited use for assessing individuals with sexual dysfunction. Pupillary responses could offer an additional or even alternative physiological measure of sexual arousal and sexual orientation to these measures, due to its non-invasive, instantaneous, and involuntary reaction to sexually appetitive content (Finke et al., 2017).

Hess et al. (1965) and Hess and Polt (1960) were the first to investigate whether changes in pupil size as a response to arousal are specific to preferred categories of pictorial sexual stimuli, which would therefore indicate subjects' sexual orientation. The pupils of heterosexual and gay men were recorded with a camera sampling at a rate of two frames per second, while they viewed paintings and photographs of nude men and women. The pupils of all five heterosexual men dilated to the female pictures, while four of the five gay men experienced greater pupil dilation to male stimuli.

In the 50 years that followed publication of these findings, there were attempts to replicate these results. However, these studies reported contradictory findings. While Hess et al. (1965) demonstrated clear sex-specific pupillary responses, others failed to find differences in the pupillary responses of men and women (Aboyoun & Dabbs, 1998; Scott et al., 1967). These approaches varied widely in their methodology, which included marked differences in participant sex and sexual orientation, method for recording pupil size, variations in stimuli (e.g., nude versus dressed; photographs versus artistic illustrations; highly controlled versus natural), and analytical approaches to pupillary responses (Aboyoun & Dabbs, 1998; Hamel, 1974; Hess et al., 1965; Scott et al., 1967). These methodological differences, in combination with often elementary approaches for measuring pupil size, likely gave rise to the inconsistent response patterns recorded. Furthermore, the studies were often highly time-consuming, for example, requiring manual frame-by-frame measurement and limited data (e.g., recording only three pupil measurements per slide; Garrett et al., 1989). Consequently, during that time period fewer than ten attempts to replicate Hess et al.'s findings using comparable methodologies followed, with variable success (Bernick et al., 1971; Garrett et al., 1989; Hamel, 1974; Nunnally et al., 1967; Peavler & McLaughlin, 1967; Schnelle et al., 1974; Scott et al., 1967). As a result, the study of pupillometry for measuring sexual interests was largely abandoned.

A resurgence of this research occurred some 40 years after Hess et al.'s (1965) work with the increased accessibility of contemporary eye-tracking technology (see, e.g., Attard-Johnson & Bindemann, 2017; Attard-Johnson et al., 2016, 2017; Finke et al., 2017; Rieger & Savin-Williams, 2012; Rieger et al., 2015). This technology uses video-based pupil and corneal reflection for tracking and recording eye gaze position and pupil size with high precision, recording pupil size every 1 ms to 20 ms depending on equipment used. (A comparison of methodological technique is provided in greater detail in the "Measurement Techniques" section.) A number of findings emerged from these studies. For example, when observers view erotic video footage (e.g., male–male or male–female interactions, or solo male/solo female actions), pupil dilation occurs to content that is in line with their self-reported arousal and self-identified sexual orientation

(Rieger & Savin-Williams, 2012). These response patterns have also been recorded during the viewing of static images depicting people in natural scenes (e.g., people wearing swim wear on beaches; Attard-Johnson et al., 2016, 2017), as well as for highly controlled sexually explicit and non-explicit images of adults (see Attard-Johnson & Bindemann, 2017). In these studies, for example, heterosexual men's pupils are larger when viewing female adults. In contrast, gay men tend to show increased pupil sizes for same-sex stimuli. Furthermore, pupillary responses to sexual stimuli have also been shown to correspond with direct measures of genital arousal (Rieger et al., 2015) and with reaction time measures of sexual interest (Ó Ciardha et al., 2018), providing further validation for pupillary responses as a measure of sexual interest.

These studies show that pupillary responses are not only able to indicate an arousal response to sexual stimuli generally, but can change in response to categories of specific interest to an observer (Attard-Johnson & Bindemann, 2017; Attard-Johnson et al., 2016, 2017; Rieger & Savin-Williams, 2012; Rieger et al., 2015) and, as such, could provide an index of sexual orientation. However, recent studies of pupillary response with contemporary eye-tracking equipment also provide some contradictory results, indicating that a better understanding of this approach is still required. This variation in findings is reflected, for example, in sex differences between male and female observers.

Sex Differences in Pupillary Response Patterns

Whereas men's physiological measures of arousal correlate strongly with their self-reported sexual orientation and subjective appraisal of the stimuli being perceived, this relationship is much weaker in women (Bailey, 2009; Chivers et al., 2010). This inconsistent pattern of responding also seems to extend to the measurement of arousal with pupillary responses to sexually appetitive images (Attard-Johnson et al., 2016, 2017; Rieger & Savin-Williams, 2012; Rieger et al., 2015). In these studies, heterosexual women sometimes record larger pupil sizes for the other-sex consistent with sexual orientation (Attard-Johnson & Bindemann, 2017), sometimes indistinguishable pupillary responses for same and other-sex stimuli (Rieger et al., 2015), and sometimes larger pupil sizes when viewing same-sex stimuli (Attard-Johnson et al., 2016). While this variation in response patterns is evident in heterosexual women, lesbian women have a greater tendency to respond in concordance with their self-reported sexual orientation and subjective arousal ratings (Rieger & Savin-Williams, 2012).

A factor that could contribute to inconsistencies in female sexual responding might reflect differences in the stimuli that are used in different studies. Stimuli that have been employed in the measurement of female sexual response more broadly vary in terms of modality but are typically sexually explicit.

All 132 studies published between 1969 and 2007 reporting genital and self-reported sexual arousal responses utilized stimuli depicting highly explicit sexual content comprising nudity, sexual acts, erotic video footage, erotic audio tapes, or sexual fantasy (studies reported in a meta-analysis by Chivers et al., 2010). Such stimuli have also been used extensively in more recent studies into measurement of genital responses in women (Dawson & Chivers, 2018). In contrast, research with pupillary response measurements differs not only in mode (e.g., static images versus video footage) but also in degree of sexual explicitness, from erotic videos depicting sexual acts (see Rieger & Savin-Williams, 2012; Rieger et al., 2015) to completely nude, semi-nude, and dressed images of men and women (see Attard-Johnson et al., 2016; Attard-Johnson & Bindemann, 2017). Although male sex-specific responding appears to be rather robust across these variations in stimuli, female responding is more varied (Attard-Johnson & Bindemann, 2017; Attard-Johnson et al., 2016; Finke et al., 2017; Rieger & Savin-Williams, 2012; Rieger et al., 2015).

These more varied responses in females could be a result of the preparation hypothesis (Dawson et al., 2013; Lalumière et al., 2020), which proposes that any sexual cues provoke automatic genital responses as a protective mechanism to prepare a woman for the possibility of a sexual encounter, and thus reduce the likelihood of pain or injury occurring (Chivers, 2017; Laan & Everaerd, 1995; Suschinsky & Lalumière, 2011). A time course of pupillary responses to explicit sexual stimuli reported in Finke et al. (2017) suggests that the pupils dilate indiscriminately to sexual stimuli within 400–600 ms of onset of image (when stimulus luminance is controlled for), followed by more specific responding. Specifically, early and late pupillary responses were associated with differential heart rate and skin conductance responses and suggest that these distinct responses are associated with different processes of the autonomic nervous system (Finke et al., 2017). Initial dilation (parasympathetically controlled), when not attributed to the light reflex, may be an immediate and spontaneous response to the detection of *any* biologically significant sexual cue. Later pupil dilation (sympathetically controlled) occurring after the initial 400 ms period may relate to stimulus content and sexual orientation (Finke et al., 2017). This can be seen in men where pupil size to all erotic stimuli (i.e., both other and same sex) started to differ significantly from baseline and from neutral non-erotic stimuli at 400 ms, whereas pupil size diverged after 600 ms for specific sexual stimuli according to sexual orientation (Finke et al., 2017). In women, however, the indiscriminate pupil responses between images of men and women continued throughout the duration of image presentation. The initial response to sexual cues provides some support for the preparation hypothesis and suggests that perhaps in women this response is sustained for longer and may be underpinned by a different biological process to that in men.

If the preparation hypotheses results in sustained responses in women, then images of high stimulus intensity depicting strong sexual cues, such as nude or erotic images of men and women, may trigger a heightened *non-specific* arousal response that attenuates any discrimination in arousal between image categories from women. If the sexual intensity of the images is reduced, for example, by viewing images of dressed men and women absent of all sexual cues, female responses may be less sensitive and become *more* discriminant across categories. Stimulus type and intensity are varied across studies measuring genital arousal, including audio-visual stimuli comparing arousal to human sex acts of varying intensity, primate sex acts, neutral scenes with no-person content (Chivers & Bailey, 2005), and audio narrative descriptions of sexual and non-sexual interactions (Chivers & Timmers, 2012). However, it is unclear from these studies whether the non-specific pattern of responding in women is a function of stimulus intensity. Furthermore, existing research measuring arousal with pupillary responses to non-sexually explicit stimuli of men and women is limited and also does not reveal a specific pattern. For example, while heterosexual women's pupils were larger for the other-sex when viewing both nude and dressed stimuli of the same identity (Attard-Johnson & Bindemann, 2017), in a different paradigm dilation was more strongly correlated with sexual orientation when women viewed sexually explicit videos compared with non-explicit videos of men and women discussing the weather (Watts et al., 2016). Furthermore, other studies report pupil dilation consistent with sexual orientation when women were exposed to only the faces of men and women (Laeng & Falkenberg, 2007). It is therefore possible that the specific experimental stimuli of a study can affect the response patterns found in women. However, the precise mechanism and the extent to which this occurs are not clear and may be attributed to an interaction of multiple factors. One of these factors may be the technique that is applied to measure changes in pupil size.

Measurement Techniques

Since the early studies, advancements in eye-tracking technology have made it easier to accurately measure pupil size changes. However, with limited research and no standardized approach, studies have varied widely in the methods of measurement and analysis (see Attard-Johnson et al., 2019), which could contribute to variations in pupillary response patterns across studies. The methods employed to measure pupillary responses differ particularly when comparing earlier measurement techniques, which were based on manual application of a ruler to calculate pupil size from still images of the eye, compared with modern techniques that apply automated infrared light to capture pupil responses.

One aspect in which these approaches vary relates to the rate of recording of pupil response, whereby early methods recorded far fewer pupil size measurements per stimuli compared to modern eye-tracking technology. Some earlier studies report only three recordings per trial using manual methods (Garrett et al., 1989) compared with a pupil measurement recorded every 1 to 20 ms for the duration of every trial with modern eye-tracking technology, such as the EyeLink (SR Research Ltd., 2005–2008), Tobii (2016), and SMI (Sensi-Motoric Instruments, Gaze Intelligence, 2020) eye trackers. Consequently, when scores for a trial are averaged across fewer pupil measurements, this could allow greater opportunity for erroneous pupil size changes unrelated to the task to have an undue influence on the overall task-evoked pupillary responses. Erroneous pupil dilation or constriction refers to noise in pupil dilation data arising from spontaneous internal cognitive processes unrelated to the task (e.g., working memory and cognitive load; Kahneman & Beatty, 1966), or even direction of eye gaze in relation to the stimulus. An additional advantage of modern eye trackers is that they can record the location of eye gaze alongside pupil responses. In this way, it is possible to eliminate erroneous pupil measurements that are linked to eye movements falling outside the boundaries of the stimulus or due to eye blinks.

Modern eye-tracking systems also differ from one another in their setup. For example, some eye trackers are desktop-mounted and involve the physical stabilization of the head using a chinrest to minimize head movements. Other remote eye trackers do not require physical head restraint but, instead, rely on computer algorithms to correct for head movements. The former generally provides greater accuracy while the latter allows more mobility (Brisson et al., 2013). There are several variations in the technological design, such as calibration quality and sampling rate, across a range of eye trackers that could provide a potential source of difference. The influence of these sources of error could be mitigated to some extent by reducing head movement with the use of a head or chin rest (Titz et al., 2018).

Environmental lighting is one factor that could provide a source of difference in accuracy between remote and desktop-mounted eye trackers. Although the effect of lighting has not been directly investigated in this context, one can imagine how lighting variation from the environment throughout an experiment can influence pupillary responses to the process being investigated. While researchers typically ensure consistent artificial lighting in the testing laboratory, remote eye trackers allowing for even minor head movements can pose a potential source of variation in light from the environment entering the pupil at different intensities. Subtle head movements, such as tilting the head backwards toward an overhead light or forwards and away from the light, can trigger a pupil light reflex (Ellis, 1981). Such changes in light intensity from the environment can cause pupil adaptation and recovery in

the time following (Mathôt, 2018), which can risk interfering with the task-related pupil responses being measured. However, a systematic and direct comparison of pupil response measurement across different eye-tracking systems has not been performed and therefore differences in accuracy and reliability are yet to be established. Therefore, we cannot exclude the possibility that differences in eye-tracking methods could account for at least some of the discrepancies in findings across studies.

Present Study

In recent years, there has been a marked increase in studies examining pupil responses as a method for assessing a person's sexual interest in other adults. Available evidence, however, suggests that pupil responses are not a reliable measure across and within observer groups that differ in sex and sexual orientation. Furthermore, the existing studies diverge in their methodological approaches to using pupil responses to assess sexual interests. Consequently, uncertainty exists at present as to the sensitivity of pupillary responses as a measure of sexual interests and the impact of measurement approach on such effects. In this context, the present study provides a meta-analysis of existing research to examine the extent to which pupil responses provide a measure of sexual interests in adult men and women. For this purpose, difference scores comparing pupillary responses to images of males and females were calculated separately for men and women. This within-subjects approach was taken to examine the applicability of the method in assessing sexual interest among these groups. The meta-analysis also includes factors that may moderate the validity of this measurement technique. Specifically, moderator analyses were conducted to examine the extent to which stimulus sexual explicitness (completely nude, partially nude with sexual regions obscured, or completed dressed), pupillary response measurement technique, and self-report measures used to confirm sexual orientation were related to the sensitivity of pupillary response for assessing sexual interests.

Based on previous research, we expected that pupil sizes are consistently larger for stimuli consistent with sexual orientation in men (i.e., larger pupils for other-sex stimuli in heterosexual men, and for same-sex stimuli in gay men). For heterosexual women, we expected a weak pattern of pupillary responses that may be larger for male stimuli. In contrast, we expected that lesbian women show stronger pupil dilation to same-sex images. Bisexual men and women would be expected to show undistinguishable pupillary responses between male and female stimuli.

Method

Studies available in peer-reviewed journals, books or book chapters, as well as unpublished dissertations, abstracts, conference papers, and preprints, were included in this review up until January 2020. Studies were identified by searching electronic research databases Web of Science and EBSCO (which includes PubMed, Medline, Science Direct, Open Dissertations, PsychINFO), by examining the reference lists of relevant studies, and via ResearchGate. Authors were contacted directly to request full texts, additional data, and information. The search terms employed were the following: pupil* (asterisk indicating variation such as pupillometry, pupil dilation, pupil responses, pupil size, pupillary) and sexual* (appeal, interest, arousal, preference, orientation), pupil* and attract* (attraction, attractive; see Figure S1 in Online Supplementary Materials).

To be included studies needed to meet the following criteria: They measured and compared pupil responses to visual stimuli of adult men and women to measure sexual interest; included at least one sample of male or female observers identifying as heterosexual, gay, or bisexual; stimuli (video or images) used depicted either a man or a woman but not both; a sample size of at least 5 observers per group; and contained sufficient information to calculate effect size d. In September 2019, the search yielded 40 documents following removal of duplicates. For studies reporting overlapping samples, the average effect size was taken if the sample size was identical, and the effect size for the study with the largest sample was taken if the sample size was not identical. In total, 16 eligible studies were identified (see Fig. 1). From the 17 studies, 16 included comparisons for heterosexual men, 7 for gay men, 4 for bisexual men, 14 for heterosexual women, 3 for lesbian women, and 2 for bisexual women. Before commencing analysis, another search was performed in January 2020 that revealed two additional studies for heterosexual men and women. The total number of studies eligible was increased to 19. See Table 1 for a summary of studies.

Samples

Of the 19 eligible studies, 17 were classified as published, and 2 (studies 39 and 40) were not published at the time the meta-analysis was conducted but had been submitted for publication. The year the studies were conducted ranged from 1965 to 2020 and samples were drawn from the UK ($k = 9$), the U.S. ($k = 5$), Germany ($k = 2$), the Netherlands ($k = 2$), and Norway ($k = 1$). Of those studies, 18 (95%) indicated that the participants were from university, college, or community and the participants of one study were drawn from a prison population.
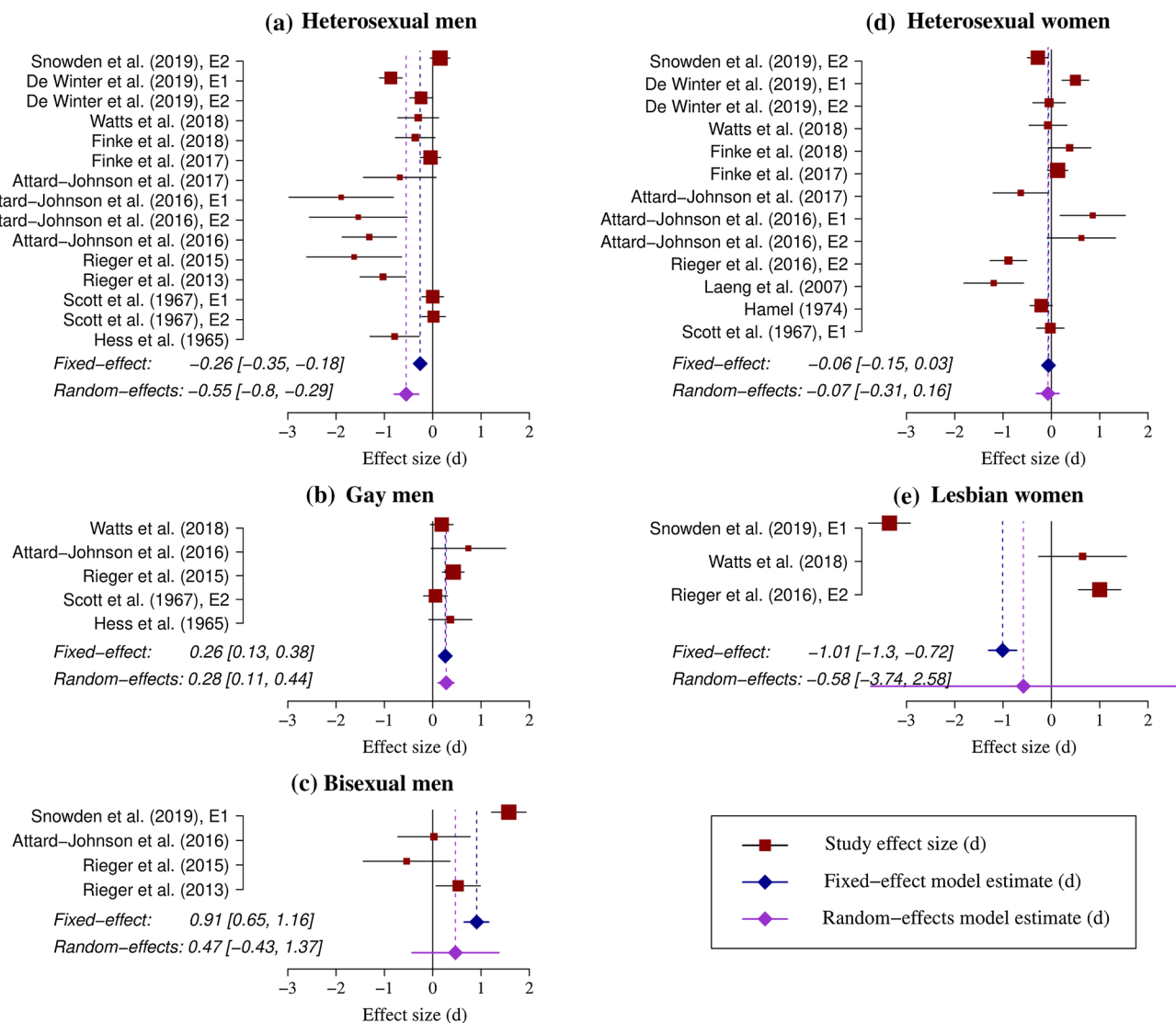
Fig. 1 Forest plots for the main meta-analysis results for heterosexual men (**a**), gay men (**b**), bisexual men (**c**), heterosexual women (**d**), and lesbian women (**e**). Plotted are the effect size estimates for each study and the meta-analysis estimates (with 95% CIs). The size of squares is proportional to the fixed-effect weight of each study. Positive effect sizes indicate greater pupil dilation to same-sex stimuli, and negative effect sizes indicate greater pupil dilation to other-sex stimuli. Study No. 4 (Snowden et al., 2019, E1) was excluded as an outlier from the analysis in panels **a**, **b**, and **d**; study No. 24 (Rieger et al., 2013) was also excluded as an outlier in panel **b** (see the Online Supplementary Materials for more details)

Coding Procedure

All studies were coded separately by the first and second author using a coding manual (see online supplement). On completion of the independent coding, consensus was reached following any differences between the raters. To assess interrater reliability for categorical variables ($n=8$), Cohen's kappa (k) was calculated and ranged from 0.64 to 1.00 (median = 0.82) for the following variables: sexual interest measure, location type, pupil response measure, stimulus mode, stimulus explicitness, stimulus body region, stimulus presentation duration, and pupil response analysis

methods. To calculate interrater reliability for the continuous variables, intraclass correlation coefficients for absolute agreements (ICC) were performed. These ranged from 0.66 to 1.00 (median = 1.00) for the following variables: sample size, mean sample age, mean SD for sample age, mean pupil response, mean SD for pupil response, correlations, and t-values.

Effect Size Calculation

The effect size was calculated in Cohen's d for within-subject designs using formulas 12.19–12.21 from Borenstein (2009).

Table 1 Summary of descriptive information of eligible studies included in meta-analysis

Sample: Heterosexual men

Study No.	Study	Country	N	Age	Setting	Sexual interest measure	Pupil response measure	Stimuli	Sexual explicitness	Duration per trial (ms)	Number of trials per category
1	Attard-Johnson et al. (2016)	UK	22	21.8	University/College	Self-report	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Images—Photographs	Partial nudity (obscured sexual regions)	5000	5
2	Attard-Johnson et al. (2016)	UK	21	19.5	University/College	Self-report	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Images—Photographs	Partial nudity (obscured sexual regions)	5000	5
4	Snowden et al. (2019)	UK	20	25.5	University/College/Community	Self-report (Kinsey Scale)	Tobii X2-60 Hz mobile eye-tracker (60 Hz rate recording every 16.67 ms)	Images—Photographs	Mixed nudity	2000	8
5	Snowden et al. (2019)	UK	18	21.6	University/College/Community	Self-report (Kinsey Scale)	Tobii X2-60 Hz mobile eye-tracker (60 Hz rate recording every 16.67 ms)	Images—Photographs	Mixed nudity	3000	10
7	Watts et al. (2018)	UK	6	25	University/College/Community	Self-report (Kinsey Scale)	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Video footage	Complete nudity	18,000	3
9	Hess et al. (1965)	USA	5	29	University/College/Community	Self-report	Manual frame-by-frame measurement of pupil diameter from video footage	Images—Photographs and paintings	Mixed nudity	10,000	5
10	Finke et al. (2018)	Germany	17	24.1	University/College	Self-report	SMI iView-X HiSpeed 500 eye-tracker (500 Hz sample rate)	Images—Photographs	Complete nudity	2500	15

Table 1 (continued)

Sample: Heterosexual men

Study No.	Study	Country	N	Age	Setting	Sexual interest measure	Pupil response measure	Stimuli	Sexual explicitness	Duration per trial (ms)	Number of trials per category
11	Finke et al. (2017)	Germany	14	25.3	University/College	Self-report	SMI iView-X HiSpeed 500 eye-tracker (500 Hz sample rate)	Images—Photographs	Mixed nudity	2500	48
14	Scott et al. (1967)	USA	10	–	University/College	Sexual interest not recorded	Manual frame-by-frame measurement of pupil diameter from video footage (2 f/s)	Images	Mixed nudity	10,000	8
15	Scott et al. (1967)	USA	5	26.5	Prison	Self-report	Manual frame-by-frame measurement of pupil diameter from video footage (2 f/s)	Images	Mixed nudity	10,000	8
17	Attard-Johnson et al. (2016)	UK	57	21.6	University/College	Self-report	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Images—Photographs	Partial nudity (obscured sexual regions)	10,000	18
24	Rieger et al. (2013)	UK	94	28	University/College/Community	Self-report (Kinsey Scale)	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Video footage	Complete nudity	30,000	1
29	Attard-Johnson et al. (2017)	UK	20	22.4	University/College	Self-report (Kinsey Scale)	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Images—Photographs	Mixed nudity	2669 (average duration)	18

Table 1 (continued)

Sample: Heterosexual men

Study No.	Study	Country	N	Age	Setting	Sexual interest measure	Pupil response measure	Stimuli	Sexual explicitness	Duration per trial (ms)	Number of trials per category
39	De Winter et al. (2021)	Netherlands	129	23.2	University/College	Sexual interest not recorded	EyeLink SR Research Remote infrared eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Images—Photographs	Partial nudity (obscured sexual regions)	10,000	2
40	De Winter et al. (2021)	Netherlands	102	23.3	University/College	Sexual interest not recorded	EyeLink SR Research Remote infrared eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Images—Line drawings	Partial nudity (obscured sexual regions)	10,000	2
41	Rieger et al. (2015)	USA	30	24.5	University/College/Community	Self-report (Kinsey Scale)	EyeLink SR Research Remote infrared eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Video footage	Complete nudity	18,000	3

Sample: Gay men

Study No.	Study	Country	N	Age	Setting	Sexual interest measure	Pupil response measure	Stimuli	Sexual explicitness	Duration per trial (ms)	Number of trials per category
4	Snowden et al. (2019)	UK	19	27.1	University/College/Community	Self-report (Kinsey Scale)	Tobii X2-60 Hz mobile eye-tracker (60 Hz rate recording every 16.67 ms)	Images—Photographs	Mixed nudity	2000	8
7	Watts et al. (2018)	UK	6	25	University/College/Community	Self-report (Kinsey Scale)	EyeLink SR Research Remote infrared eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Video footage	Complete nudity	18,000	3

Table 1 (continued)

Sample: Gay men

Study No.	Study	Country	N	Age	Setting	Sexual interest measure	Pupil response measure	Stimuli	Sexual explicitness	Duration per trial (ms)	Number of trials per category
9	Hess et al. (1965)	USA	5	29	University/College/Community	Self-report	Manual frame-by-frame measurement of pupil diameter from video footage	Images—Photographs and paintings	Mixed nudity	10,000	5
15	Scott et al. (1967)	USA	5	26.5	Prison	Self-report	Manual frame-by-frame measurement of pupil diameter from video footage (2 f/s)	Images	Mixed nudity	10,000	8
17	Attard-Johnson et al. (2016)	UK	20	24.5	University/College	Self-report	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Images—Photographs	Partial nudity (obscured sexual regions)	10,000	18
24	Rieger et al. (2013)	UK	107	23.28	University/College/Community	Self-report (Kinsey Scale)	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Video footage	Complete nudity	30,000	1
41	Rieger et al. (2015)	USA	29	23.66	University/College/Community	Self-report (Kinsey Scale)	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Video footage	Complete nudity	18,000	3

Table 1 (continued)

Sample: Bisexual men

Study No.	Study	Country	N	Age	Setting	Sexual interest measure	Pupil response measure	Stimuli	Sexual explicitness	Duration per trial (ms)	Number of trials per category
4	Snowden et al. (2019)	UK	13	26.9	University/College/Community	Self-report (Kinsey Scale)	Tobii X2-60 Hz mobile eye-tracker (60 Hz rate recording every 16.67 ms)	Images—Photographs	Mixed nudity	2000	8
17	Attard-Johnson et al. (2016)	UK	19	21.1	University/College	Self-report	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Images—Photographs	Partial nudity (obscured sexual regions)	10,000	18
24	Rieger et al. (2013)	UK	76	26.14	University/College/Community	Self-report (Kinsey Scale)	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Video footage	Complete nudity	30,000	1
41	Rieger et al. (2015)	USA	16	24.67	University/College/Community	Self-report (Kinsey Scale)	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Video footage	Complete nudity	18,000	3

Table 1 (continued)

Sample: Heterosexual women

Study No.	Study	Country	N	Age	Setting	Sexual interest measure	Pupil response measure	Stimuli	Sexual explicitness	Duration per trial (ms)	Number of trials per category
1	Attard-Johnson et al. (2016)	UK	22	21.8	University/College	Self-report	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Images—Photographs	Partial nudity (obscured sexual regions)	5000	5
2	Attard-Johnson et al. (2016)	UK	20	19.5	University/College	Self-report	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Images—Photographs	Partial nudity (obscured sexual regions)	5000	5
4	Snowden et al. (2019)	UK	28	23.1	University/College/Community	Self-report (Kinsey Scale)	Tobii X2-60 Hz mobile eye-tracker (60 Hz rate recording every 16.67 ms)	Images—Photographs	Mixed nudity	2000	8
5	Snowden et al. (2019)	UK	20	21.6	University/College/Community	Self-report (Kinsey Scale)	Tobii X2-60 Hz mobile eye-tracker (60 Hz rate recording every 16.67 ms)	Images—Photographs	Mixed nudity	3000	10
7	Watts et al. (2018)	UK	9	26.2	University/College/Community	Self-report (Kinsey Scale)	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Video footage	Complete nudity	18,000	3

Table 1 (continued)

Sample: Heterosexual women

Study No.	Study	Country	N	Age	Setting	Sexual interest measure	Pupil response measure	Stimuli	Sexual explicitness	Duration per trial (ms)	Number of trials per category
10	Finke et al. (2018)	Germany	18	24.1	University/College	Self-report	SMI iView-X HiSpeed 500 eye-tracker (500 Hz sample rate)	Images—Photographs	Complete nudity	2500	15
11	Finke et al. (2017)	Germany	14	25.3	University/College	Self-report	SMI iView-X HiSpeed 500 eye-tracker (500 Hz sample rate)	Images—Photographs	Mixed nudity	2500	48
14	Scott et al. (1967)	USA	10	–	University/College	Sexual interest not recorded	Manual frame-by-frame measurement of pupil diameter from video footage (2 f/s)	Images—Photographs	Mixed nudity	10,000	8
19	Hamel (1974)	USA	22	20	University/College	Sexual interest not recorded	Pupillometer system model V-1165-1R	Images	Mixed nudity	5,000	8
23	Laeng & Falkenberg (2007)	Norway	14	22.5	University/College	Other	SMI iView-X HiSpeed 500 eye-tracker (500 Hz sample rate)	Images—Photographs	No nudity (face only)	10,000	10 male; 5 female
27	Rieger et al. (2016)	UK	131	21.6	University/College/Community	Self-report (Kinsey Scale)	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Video footage	Complete nudity	30,000 & 180,000	12 & 3
29	Attard-Johnson et al. (2017)	UK	25	22.4	University/College	Self-report (Kinsey Scale)	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Images—Photographs	Mixed nudity	2669 (average duration)	18

Table 1 (continued)

Sample: Heterosexual women

Study No.	Study	Country	N	Age	Setting	Sexual interest measure	Pupil response measure	Stimuli	Sexual explicitness	Duration per trial (ms)	Number of trials per category
39	De Winter et al. (2021)	Netherlands	53	23.2	University/College	Sexual interest not recorded	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Images—Photographs	Partial nudity (obscured sexual regions)	10,000	2
40	De Winter et al. (2021)	Netherlands	45	23.3	University/College	Sexual interest not recorded	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Images—Line drawings	Partial nudity (obscured sexual regions)	10,000	2
19	Hamel (1974)	USA	22	20	University/College	Sexual interest not recorded	Pupillometer system model V-1165-1R	Images	Mixed nudity	5,000	8
23	Laeng and Falkenberg (2007)	Norway	14	22.5	University/College	Other	SMI iView-X HiSpeed 500 eye-tracker (500 Hz sample rate)	Images—Photographs	No nudity (face only)	10,000	10 male; 5 female
27	Rieger et al. (2016)	UK	131	21.6	University/College/Community	Self-report (Kinsey Scale)	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Video footage	Complete nudity	30,000 & 180,000	12 & 3
29	Attard-Johnson et al. (2017)	UK	25	22.4	University/College	Self-report (Kinsey Scale)	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Images—Photographs	Mixed nudity	2669 (average duration)	18

Table 1 (continued)

Sample: Heterosexual women

Study No.	Study	Country	N	Age	Setting	Sexual interest measure	Pupil response measure	Stimuli	Sexual explicitness	Duration per trial (ms)	Number of trials per category
39	De Winter et al. (2021)	Netherlands	53	23.2	University/College	Sexual interest not recorded	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Images—Photographs	Partial nudity (obscured sexual regions)	10,000	2
40	De Winter et al. (2021)	Netherlands	45	23.3	University/College	Sexual interest not recorded	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Images—Line drawings	Partial nudity (obscured sexual regions)	10,000	2

Sample: Lesbian women

Study No.	Study	Country	N	Age	Setting	Sexual interest measure	Pupil response measure	Stimuli	Sexual explicitness	Duration per trial (ms)	Number of trials per category
4	Snowden et al. (2019)	UK	17	26.1	University/College/Community	Self-report (Kinsey Scale)	Tobii X2-60 Hz mobile eye-tracker (60 Hz rate recording every 16.67 ms)	Images—Photographs	Mixed nudity	2000	8
7	Watts et al. (2018)	UK	9	26.2	University/College/Community	Self-report (Kinsey Scale)	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Video footage	Complete nudity	18,000	3
27	Rieger et al. (2016)	UK	106	24.97	University/College/Community	Self-report (Kinsey Scale)	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Video footage	Complete nudity	30,000 & 180,000	12 & 3

Table 1 (continued)

Sample: Bisexual women

Study No.	Study	Country	N	Age	Setting	Sexual interest measure	Pupil response measure	Stimuli	Sexual explicitness	Duration per trial (ms)	Number of TRI-ALS per category
4	Snowden et al. (2019)	UK	21	24.3	University/College/Community	Self-report (Kinsey Scale)	Tobii X2-60 Hz mobile eye-tracker (60 Hz rate recording every 16.67 ms)	Images—Photographs	Mixed nudity	2000	8
27	Rieger et al. (2016)	UK	108	22.9	University/College/Community	Self-report (Kinsey Scale)	EyeLink SR Research Remote infra-red eye-tracker (500/1000 Hz sample rate recording every 2 ms)	Video footage	Complete nudity	30,000 & 180,000	12 & 3

Cohen's d measured the average difference in pupil size during the viewing of images of men and women. Separate meta-analyses were conducted for the different gender and sexual orientation groups (i.e., heterosexual men, bisexual men, gay men, heterosexual women, and lesbian women). No analysis of bisexual females was possible due to the insufficient number of studies ($k = 2$). Positive effect sizes indicate larger pupil size for same-sex stimuli, whereas negative effect sizes indicate larger pupil size for other-sex stimuli. Cohen's d values of 0.20 reflect "small," 0.50 "moderate," and 0.90 "large" effects (Cohen, 1988). For comparison, examples of typical effect sizes for genital arousal measures while viewing male versus female sexual erotica are $d = 1.03$ for heterosexual men (Jabbour et al., 2020) and $d = 3.20$ for gay men (Semon et al., 2017), and $d = 0.41$ for Viewing Time measures in heterosexual and gay men (Ó Ciardha et al., 2018). An effect size for women of all sexual orientations may not provide a useful comparison since the direction of genital response patterns are varied across studies (see, for example, Chivers et al., 2004; Chivers, 2005; Rieger et al., 2015). However, an example of an effect size measuring genital arousal in heterosexual women is $d = 0.64$ with greater arousal to same-sex stimuli (Chivers et al., 2004) and $d = 0.22$ with no difference between male and female stimuli for Viewing Time measures (Xu et al., 2016).

All effect sizes where the 95% CI excludes 0 are equivalent to being statistically significant at $p < 0.05$. When confidence intervals for two variables do not overlap, these can be considered as being statistically different from each other, $p < 0.01$.

Meta-Analysis

Meta-analysis is a statistical method to aggregate weighted findings from multiple samples, presuming the aggregation of findings reflects a more accurate measure of the true effects of a particular variable of interest than individual samples. Findings across samples were aggregated using both fixed-effect and random-effects meta-analysis (Borenstein et al., 2009). Fixed-effect meta-analysis assumes that all studies are estimating the same underlying effect and any variability in this estimate is due to sampling error alone (Welton et al., 2012). Random-effects meta-analysis, on the other hand, relaxes this assumption by allowing for variability in the true effect size between studies. In this model, each study has its own unique underlying effect, which in turn is generated from a probability distribution of effect sizes. Thus, the unique underlying effects for studies can be informally thought of as random samples from a distribution of effect sizes (Welton et al., 2012).

When variability across samples is low ($Q <$ degrees of freedom), fixed-effect and random-effects meta-analyses produce identical results. While both fixed-effect and

random-effects results are presented, we focused on the fixed-effect model for interpretation as all the results are based on a small number of samples ($k < 30$) and the between-study variability component necessary for random-effects analyses (i.e., T^2, tau) becomes imprecise as the number of samples decreases (Schulze, 2007).

We used Cochran's Q and I^2 statistics to test the degree of heterogeneity across samples. Cochran's Q statistic tests whether variability across samples in an effect size is larger than would be expected by sampling error. The I^2 value is akin to a measure of effect size variability across samples and indicates the variability due to true heterogeneity, above and beyond what is expected by sampling error as opposed to chance. The I^2 statistic can be compared across analyses, whereas Q cannot when the number of samples (k) varies across analyses. As a general heuristic, I^2 values of 25% are considered low, 50% moderate, and 75% high variability (Higgins et al., 2003).

Meta-analyses can be strongly influenced by unusually small variance, typically due to unusually large sample sizes in between-subject d meta-analysis, but in within-subject meta-analyses such as the current study, by unusually large correlations estimates. Particularly in moderator analyses (when k decreases), the effects of other studies can largely disappear in the presence of a particularly heavy-weighted study. As in other meta-analyses (Hanson & Bussière, 1998), a study was considered an outlier and excluded if it had an extreme weight, the overall Q was significant, was the most extreme value, and it accounted for more than 50% of the overall Q. The exception to this rule was for analysis which only included three or fewer studies such that none of the studies were considered outliers. In addition to outliers, meta-analyses can also be strongly influenced by unusually large sample sizes. Therefore, to reduce the influence of an unusually heavy-weighted study, we artificially gave such a study only 10% more weight than the next largest study (see online supplement).

Moderator Analyses

When there was a reasonable amount of variability in an aggregated effect size ($I^2 > 20\%$) and sufficient samples reporting on the moderator of interest (2 for each level of the categorical moderator), between-level Q statistic analyses were conducted (Borenstein et al., 2009). A statistically significant between-level Q statistic analysis indicates that the moderator explains a significant portion of the variability across samples, above what would be expected by sampling error alone. Moderators examined included stimulus sexual explicitness (i.e., degree of nudity—complete nudity, partial, or dressed), pupillary measurement technique (i.e., eye tracker system used), and whether a self-report measure of sexual orientation was included.

Publication Bias

Publication bias occurs when studies in peer-reviewed journals have systematically different results from studies that are never published (Song et al., 2013). As such, publication bias can be a threat to meta-analysis validity if published studies are not representative of all studies that have been done on the topic (Rothstein et al., 2005). In the present study, two main publication bias tests were conducted: P-curves (Simonsohn et al., 2014a, 2014b, 2015) and funnel plots (Light & Pillemer, 1984; Sterne & Egger, 2001; Sterne et al., 2011). P-curves were generated using the "dmetar" package v. 0.0.9 (Harrer et al., 2019) in R v.3.62. (R Core Team, 2019), and funnel plot tests were conducted with the "meta" R package v. 4.10 (Balduzzi et al., 2019).

P-curves test whether the meta-analysis results can be explained by selective reporting practices such as p-hacking. A set of studies are said to have evidential value (i.e., represent a true effect) if the distribution of significant p values is right-skewed. This happens because smaller p values are more likely to occur in the presence of a true effect compared to p values close to 0.05. If the effect does not exist (and there is no p-hacking), the distribution of p values would be uniform. However, if there is p-hacking, the distribution would be left-skewed because there will be a disproportionate number of p values close to 0.05 (Simonsohn et al., 2014a).

The funnel plot is a technique where the study effect sizes are plotted against a measure of their precision, such as the standard error (Sterne & Egger, 2001). In the absence of bias, studies are expected to form a symmetrical funnel shape, where more precise studies appear narrowly at the top of the plot and less precise studies scatter more widely at the bottom. However, if the funnel shape appears asymmetric because studies are missing from one side of the plot, this could indicate the presence of publication bias (though publication bias is not the only possible cause of asymmetry; see Sterne et al., 2011).

Results

The characteristics of the approaches and protocols used for measuring pupillary responses for the studies included in this meta-analysis is summarized in Table 2. The majority of studies confirmed sexual interests of the participants with a measure of self-report sexual orientation ($k = 15$), while four studies (De Winter et al., 2021; Hamel, 1974; Scott et al., 1967) did not record sexual orientation and instead assumed all participants were heterosexual. All included studies used images ($k = 15$) or video footage ($k = 4$) comprising either an adult male or female presented completely nude ($k = 5$), partially nude (i.e., sexual regions only obscured with clothing or image manipulation; $k = 5$), dressed ($k = 1$), or mixed

Table 2 Meta-analysis results of pupil dilation as a measure of sexual interest

Group	Fixed-effect		Random-effects		Q	I^2 (%)	k	N
	d	95% CI	d	95% CI				
Heterosexual men	−.26	[−0.35, −0.18]	−.55	[−0.80, −0.29]	105.90*	86.8	15	550
Gay men	.26	[0.13, 0.38]	.28	[0.11, 0.44]	6.68	40.14	5	65
Bisexual men	.91	[0.65, 1.16]	.47	[−0.43, 1.37]	31.11*	90.36	4	124
Heterosexual women	−.06	[−0.15, 0.03]	−.07	[−0.31, 0.16]	73.33*	83.64	13	403
Lesbian women	−1.01	[−1.30, −0.72]	−.58	[−3.74, 2.58]	205.44*	99.03	3	132
Bisexual women	–	–	–	–	–	–	–	–

Heterosexual men analysis included studies 1, 2, 5, 7, 9, 10, 11, 14, 15, 17, 24, 29, 39, 40, 41; gay men included studies 7, 9, 15, 17, 41; bisexual men included studies 4, 17, 24, 41; heterosexual women included studies 1, 2, 5, 7, 10, 11, 14, 19, 23, 27, 29, 39, 40; lesbian women included studies 4, 7, 27; there were insufficient bisexual women studies for analysis

d: effect size in Cohen's d; CI: confidence interval; Q: χ^2 test statistic of heterogeneity; I^2: percentage of variability between studies that can be attributed to heterogeneity rather than sampling error. k: number of studies included in the analysis. N: total number of participants on which the analysis is based

*Statistically significant

stimulus sets depicting people in a range of explicitness (i.e., a mixture of nude, partially nude, and dressed images; $k = 8$). The number of trials per category ranged from 1 to 48 (median = 8), and the stimulus duration ranged from 2 s to 3 min (median = 10 s). The majority of studies conducted since the year 2000 measured pupillary responses using eye-tracking technology, such as the SR EyeLink 1000 or SMI. However, many of the earlier studies measured pupil size manually, typically by recording video footage of the eyes of observers, while these viewed stimulus images, and then manually measuring pupil size in image stills taken from this video footage (Hess et al., 1965).

The results from the meta-analysis are also presented in Table 2 and visualized in Fig. 1. The fixed-effect meta-analysis for heterosexual men yielded a small negative effect ($d = −0.26$, fixed-effect model), indicating that they exhibited greater pupil dilation to other-sex (i.e., female) compared to same-sex (i.e., male) stimuli.[1] There was significant variability in the studies ($I^2 = 87\%$), however, above and beyond what would be expected based on sampling error alone.

In contrast to heterosexual men, gay men showed greater pupil dilation to same-sex stimuli compared to other-sex stimuli. The effect size was similar in magnitude ($d = 0.26$ in the fixed-effect model and $d = 0.28$ in the random-effects model) and was statistically significant as the 95% CI again excluded 0. Additionally, the effects were relatively consistent across studies ($I^2 = 40\%$). Therefore, in line with their

reported sexual orientation, gay men showed stronger pupil dilation to male stimuli.

The results for bisexual men yielded a similar pattern to that of gay men. Bisexual men also showed greater pupil dilation to male stimuli, but the effect size was considerably larger than that of gay men ($d = 0.91$). There was large variability between studies ($I^2 = 90\%$).

The results for heterosexual women showed a very small, statistically non-significant, negative effect size ($d = −0.06$ in the fixed-effect and $d = −0.07$ in the random-effects meta-analysis), indicating slightly greater pupil dilation to other-sex (male) stimuli. Therefore, while the effect was in the expected direction, its magnitude was modest. The effect did not reach statistical significance as the 95% CIs included 0. There was large variability between studies ($I^2 = 83\%$).

Finally, lesbian women showed greater pupil dilation to other-sex (male) stimuli ($d = −1.01$). The meta-analysis estimate was not significant as the 95% CIs included 0 and variability between studies was high ($I^2 = 99\%$). This result is contrary to predictions as lesbian women were expected to exhibit greater pupil dilation to same-sex stimuli. However, this result needs to be interpreted with caution, as the effect size estimate was likely strongly influenced by one of the three included studies (No. 4; Snowden et al., 2019, Exp.1), which had a very large negative effect size ($d = −3.35$; see Fig. 1e). Even though study No. 4 had a smaller sample size ($N = 17$) compared to study No. 27 (Rieger et al., 2016, Exp.2; $N = 106$), the weight of the two studies in the meta-analysis was almost the same (45.3% vs 44.4% in the fixed-effect model, respectively). This was because study No. 4 had smaller standard deviations compared to study No. 27, which compensated for the smaller sample size in the variance calculation of the effect size. Thus, the combination of similar weights, the overall small number of available studies in the meta-analysis, and the large (negative) effect size of study

---

[1] The random-effects model yielded a larger effect size ($d = −.55$). This occurred mostly because the random-effects model tends to assign greater weight to smaller studies, which happened to have larger negative effect sizes in general (see Fig. 1a). Fixed-effect models are interpreted in the Results because random-effects require larger number of studies (closer to 30 non-overlapping samples) to estimate between study variability (tau) accurately.

No. 4 likely contributed to the pooled meta-analysis effect being negative. No analysis of bisexual females was possible due to the insufficient number of studies ($k = 2$).[2]

Moderator Analyses

The results of the moderator analysis are presented in Table 3 for heterosexual men and in Table 4 for heterosexual women. These analyses tested whether the effect sizes are moderated by: (1) how pupil size was measured; (2) whether a self-report measure of sexual interest was administered or not; (3) whether the correlation between group means was reported or estimated from other studies; and (4) whether participants were exposed to a high, low, or a mixed level of nudity.

The effect size was significantly moderated by pupil measurement for both heterosexual men and heterosexual women. For heterosexual men, the effect was much stronger when measured with an EyeLink eye-tracker, and much weaker when measured with other equipment. In contrast, for heterosexual women, the largest effect size was obtained with manual pupil size recordings demonstrating larger pupils for other-sex stimuli.

Effect sizes were not moderated by whether the study included a self-report measure of sexual interest. This suggests that studies that reported such a measure had comparable effect sizes to those that did not report it and instead presumed that participants were heterosexual.

The type of correlation estimation also significantly moderated the effect sizes for both heterosexual men and women. For men, the effect size was stronger when the correlation could be extracted from the paper compared to when it had to be estimated using data from other similar studies. The effect sizes for heterosexual women had similar magnitude for both actual and estimated correlations. Critically, however, the effect size for estimated correlations was negative (consistent with predictions), whereas the effect size for actual correlations was positive (contrary to predictions).

Finally, the level of sexual explicitness of the visual stimuli also significantly moderated effect sizes for both heterosexual men and women. In men, the strongest effect sizes were found for low explicitness (i.e., studies with stimuli of partially nude people with obscured sexual regions), followed by high explicitness (i.e., completely nude stimuli). Mixed explicitness (i.e., stimuli containing a combination of high and low) in men led to an effect size close to 0. In contrast, the effect size for heterosexual women was positive for low explicitness (contrary to predictions) but negative for high

and mixed explicitness (in line with predictions). In summary, while significant moderation was found for three out of the four considered variables, the pattern of results was generally inconsistent across heterosexual men and heterosexual women. This suggests that some of the differences in the moderator analyses may have to do, at least in part, with methodological differences in how pupil dilation is studied.

Publication Bias

Only the analyses for heterosexual men and heterosexual women had sufficient studies to be assessed for publication bias. The results are shown in Figure S3 in Online Supplementary Materials. Both p-curves were right-skewed, indicating that at least some of the studies show evidential value (i.e., represent a true effect). Therefore, the results are not likely to be explained by selective reporting based on p values. Note that estimating the "true" effect from p-curves is not recommended in cases with high heterogeneity, such as the present analyses (van Aert et al., 2016).

The funnel plots for the heterosexual male and heterosexual female analyses are presented in Figure S4 in Online Supplementary Materials. A visual investigation of the funnel plot for heterosexual women revealed no clear evidence of asymmetry. This was confirmed with Egger's regression test (Egger et al., 1997), which showed no significant evidence for funnel plot asymmetry, $t(11) = 0.15$, $p = 0.880$. Therefore, there was no evidence that studies with heterosexual women were influenced by publication bias. In contrast, the funnel plot for heterosexual men suggested possible asymmetry, as studies were missing from the lower right side of the plot. Studies No. 1, 2, and 41 had large, negative effect sizes but also large variance, indicating lesser precision in the estimate. While it is not immediately obvious what study characteristics contributed to this, there was a lack of similarly large *positive* effect sizes with lesser precision on the left side of the funnel plot, thus leading to the apparent asymmetry. Egger's test confirmed the significant presence of asymmetry for heterosexual men, $t(13) = 3.53$, $p = 0.004$.

The funnel plot asymmetry for heterosexual men suggests that publication bias may be present. To assess its influence on the results, we used Vevea and Woods' (2005) bias selection model. This method uses weight functions to model the mechanisms through which effect sizes may be suppressed by publication bias and returns an adjusted estimate that corrects for this bias (Vevea & Coburn, 2019). The results from four such weight functions are presented in Table S1 (see Online Supplementary Materials). The average adjusted effect size across the four models was -0.17 (range: -0.15, -0.21) for the fixed-effect meta-analysis and -0.61 for the

---

[2] This number reflects the number of studies with datasets that comprise a unique sample, which is a requirement for a meta-analysis. Therefore, when studies have the same sample, or overlapping samples, only one of the studies is included in the analysis.

Table 3 Moderator analyses for heterosexual men

	Fixed-effect		Random-effects		Q	I^2 (%)	k	N
	d	95% CI	d	95% CI				
Pupil measurement	−.26	[−0.35, −0.18]	−.55	[−0.80, −0.29]	105.90*	86.8	15	550
Manual	−.07	[−0.23, 0.09]	−.18	[−0.53, 0.18]	8.44*	76.19	3	20
EyeLink	−.68	[−0.81, −0.54]	−.85	[−1.20, −0.50]	36.69*	78.2	9	481
SMI	−.11	[−0.30, 0.08]	−.14	[−0.41, 0.14]	1.77	43.54	2	31
Tobii	–	–	–	–	–	–	–	–
$Q_{between}$					55.25*			
Sexual interest measure	−.26	[−0.35, −0.18]	−.55	[−0.80, −0.29]	105.90*	86.8	15	550
Not recorded [a]	−.36	[−0.49, −0.22]	−.37	[−0.87, 0.13]	28.42*	92.96	3	241
Self-report	−.21	[−0.31, −0.10]	−.62	[−0.94, −0.31]	74.53*	85.24	12	309
$Q_{between}$					2.95			
Correlation estimation	−.26	[−0.35, −0.18]	−.55	[−0.80, −0.29]	105.90*	86.8	15	550
Estimated	.07	[−0.06, 0.20]	.09	[0.00, 0.18]	1.07	0	3	33
Actual	−.50	[−0.61, −0.39]	−.75	[−1.04, −0.46]	61.91*	82.23	12	517
$Q_{between}$					42.92*			
Stimuli[b]	−.26	[−0.35, −0.18]	−.55	[−0.80, −0.29]	105.90*	86.8	15	550
High explicitness	−.27	[−0.45, −0.10]	−.63	[−1.22, −0.05]	21.29*	85.91	4	147
Low explicitness	−.66	[−0.82, −0.51]	−1.01	[−1.53, −0.50]	27.97*	85.7	5	331
Mixed	−.03	[−0.15, 0.08]	−.15	[−0.39, 0.09]	17.05*	70.68	6	72
$Q_{between}$					39.59*			

No eligible studies with non-nude stimuli. d: effect size in Cohen's d; CI: confidence interval; Q: χ^2 test statistic of heterogeneity; I^2: percentage of variability between studies that can be attributed to heterogeneity rather than sampling error. k: number of studies included in the analysis. N: total number of participants on which the analysis is based

*Statistically significant

[a]Includes studies where sexual interest was not reported and/or not recorded. Participants were assumed heterosexual

[b]High explicitness = completely nude, low explicitness = partially nude, and mixed = a mixture of high and low or not specified

random-effects meta-analysis (range: −0.48, −0.80).[3] Therefore, while the corrected random-effects estimate was similar to the original one, the corrected fixed-effect estimate was 65% smaller on average. In summary, the true effect size for heterosexual men may be smaller in magnitude (at least for the fixed-effect model) and this should be kept in mind when interpreting the results. However, even in the presence of such bias, the effect size was still negative, thus indicating that heterosexual men showed greater pupil dilation to female stimuli.

[3] The larger variability in the adjusted random-effects estimate likely occurs, at least in part, because the bias selection model also adjusts the between-study variance of the model.

Discussion

This meta-analysis examined whether changes in pupil size in response to viewing images of adult males and females can provide a consistent measure of sexual orientation in men and women. Overall, this was found to be the case for gay and heterosexual men, whereby heterosexual men showed greater pupil dilation to other-sex (i.e., female) stimuli, and gay men showed greater pupil dilation to same-sex (i.e., male) stimuli. The overall effect sizes for these observer groups ranged from small to moderate across studies, similarly to other measures of sexual interest using cognitive tasks (e.g., Viewing Time, Ó Ciardha et al., 2018), but smaller effect sizes relative to those obtained with genital arousal measurements (e.g., Jabbour et al., 2020; Semon et al., 2017). Results for bisexual men demonstrated the same direction of effect as gay men, of greater pupil dilation to male than to female stimuli. In fact, this effect was larger for bisexual than for gay men. Previous studies recording pupil dilation and genital arousal have found varied response patterns such that men who identify as bisexual inconsistently demonstrate arousal to both men and

Table 4 Moderator analyses for heterosexual women

	Fixed-effect		Random-effects		Q	I^2 (%)	k	N
	d	95% CI	d	95% CI				
Pupil measurement	−.06	[−0.15, 0.03]	−.07	[−0.31, 0.16]	73.33*	83.64	13	403
Manual	−.14	[−0.31, 0.04]	−.13	[−0.32, 0.05]	1.03	3.05	2	32
EyeLink	.02	[−0.14, 0.17]	.02	[−0.43, 0.48]	47.14*	87.27	7	305
SMI	.06	[−0.12, 0.24]	−.18	[−0.88, 0.53]	18.14*	88.98	3	46
Tobii	–	–	–	–	–	–	–	–
$Q_{between}$					7.02*			
Sexual interest measure	−**.13**	[−0.22, −0.03]	−.13	[−0.36, 0.10]	55.91*	80.33	12	350
Not recorded[a]	−.12	[−0.27, 0.04]	−.12	[−0.27, 0.04]	1.23	0	3	77
Self-report	−.13	[−0.25, 0.01]	−.14	[−0.48, 0.20]	54.66*	85.67	9	273
$Q_{between}$					0.02			
Correlation estimation	.02	[−0.08, 0.11]	.06	[−0.14, 0.26]	39.89*	74.93	11	258
Estimated	−**.19**	[−0.33, −0.05]	−**.22**	[−0.34, −0.10]	2.03	1.32	3	52
Actual	**.18**	[0.06, 0.31]	.19	[−0.06, 0.44]	22.38*	68.72	8	206
$Q_{between}$					15.48*			
Stimuli[b]	−.05	[−0.15, 0.04]	−.04	[−0.27, 0.20]	56.67*	82.35	11	375
High explicitness	−.10	[−0.27, 0.07]	−.26	[−0.86, 0.34]	20.94*	90.45	3	158
Low explicitness	**.35**	[0.16, 0.55]	**.42**	[0.03, 0.81]	9.09*	66.99	4	140
Mixed	−**.22**	[−0.35, −0.08]	−**.23**	[−0.38, −0.07]	4.17	27.99	4	77
$Q_{between}$					22.47*			

No eligible studies with non-nude stimuli. d: effect size in Cohen's d; CI: confidence interval; Q: χ^2 test statistic of heterogeneity; I^2: percentage of variability between studies that can be attributed to heterogeneity rather than sampling error. k: number of studies included in the analysis. N: total number of participants on which the analysis is based

*Statistically significant

[a]Includes studies where sexual interest was not reported and/or not recorded. Participants were assumed heterosexual

[b]High explicitness = completely nude, low explicitness = partially nude, and mixed = a mixture of high and low or not specified

women (Rieger & Savin-Williams, 2012; Rosenthal et al., 2011). However, in some studies this group has also exhibited a stronger arousal response to men (Rieger et al., 2005; Slettevold et al., 2019; Tollison et al., 1979). These diverse findings suggest that although sexual arousal is one facet in an individual's identification with a sexual orientation, other aspects are likely to be involved, too. One such factor is sexual curiosity. This has been found to interact with arousal responses in bisexual men, whereby those with lower levels of sexual curiosity are more aroused by same-sex men compared to those with higher level of sexual curiosity, who record bisexual arousal patterns (Rieger et al., 2013).

A different pattern was observed with female observers. Heterosexual women showed a trend of larger pupil sizes during the viewing of other-sex (i.e., male) stimuli, but the magnitude of this effect was small and not significant. The gender differences in the strength of responses observed here are consistent with those found in the wider literature on the measurement of sexual interest in men and women using other physiological measurement techniques, such as direct measures of genital arousal (Chivers et al., 2004).

Similarly to the inconsistent pattern of responding in bisexual men, it is possible that sexual arousal is only one element in women's sexual identification and may form a smaller part in the makeup of their sexual orientation than in heterosexual men (Peplau & Garnets, 2000). We return to this issue in the "Moderators of Pupil Response Measures" section.

The responses of bisexual women were not analyzed due to an insufficient number of studies (Rieger et al., 2016; Snowden et al., 2019). Lesbian women showed greater dilation to other-sex stimuli. While this last result was unexpected, the analysis included only 3 studies (total $N = 132$) and was affected by one outlier study with a large negative effect size. As the analyses of gay and bisexual groups were also based on a limited number of either 3 or 4 studies, the meta-analysis of these should be considered as preliminary until more evidence becomes available.

Finally, we note that in this study the difference scores in pupil responses for images of men and women were calculated separately for male and female observers. This provides separate response patterns and effect size estimates for men and women of different self-reported sexual orientation. This

approach is applied most frequently in this literature and was taken here to examine the potential applicability of pupillary response in assessing sexual interest *within* each group. However, some studies also include comparisons between male and female observers, or between sexual orientation groups, for each image category (Rieger et al., 2012; Savin-Williams et al., 2016). These between-subject comparisons would yield different effect sizes and may also produce a different pattern of results to this meta-analysis.

Moderators of Pupil Response Measures

We examined three moderators associated with methodological variations across studies: the sexual explicitness of stimuli, the methodological technique applied to measure pupillary response, and the inclusion of a subjective measure of sexual orientation or preference. Moderator analysis was only performed for heterosexual men and heterosexual women due to the insufficient number of studies for this analysis for the other participant groups. In heterosexual men, visual stimuli which comprised of partially nude images of people provided greater discrimination compared to studies using highly sexually explicit stimuli (i.e., complete nudity). Stimulus sets that comprised a mixture of explicitness level, such as dressed, partially nude, or completely nude, produced negligible effects. However, it is important to note that stimulus type might have been confounded with other possible moderators, such as the approach used to measure pupillary responses. Pupillary responses measured using an EyeLink eye trackers, for example, provided greater discrimination of pupil size change for male and female images in heterosexual men compared to other equipment.

These findings indicate that pupil response measurement and stimulus type can provide a source of variability. The specific effects of stimulus type and approach to measuring pupillary response are difficult to untangle because the three categories of explicitness (fully nude, partial or no nudity, and mixed nudity) do not span across the different types of pupillary response measures. For heterosexual men, for example, all studies with partially nude stimuli in the meta-analysis also applied EyeLink eye trackers to measure pupil response (Attard-Johnson et al., 2016; De Winter et al., 2021), while none of the studies using other techniques (i.e., SMI/Tobii eye trackers, manual measurement) included comparable stimuli of in low explicitness. In contrast, fully nude stimuli were employed by studies using EyeLink (Rieger et al., 2013; Watts et al., 2018) and SMI eye trackers (Finke et al., 2018), but not by any of the studies with Tobii eye trackers or manual pupil measurement. This could provide some explanation for a larger effect being observed for studies using partially nude compared to studies using more sexually explicit stimuli (i.e., complete nudity).

In heterosexual women, stimuli with low sexual explicitness demonstrated a moderate positive effect size (i.e., larger pupils when viewing persons of the same-sex), but a negative effect size (i.e., larger pupils when viewing other-sex stimuli) when mixed stimulus sets were used. The effect size for high sexual explicitness was negligible. These divergent findings could reflect the general pattern of inconsistency in female responding across studies and sexual arousal assessments more widely (Chivers, 2005; Chivers et al., 2010). Several theories have been proposed to explain inconsistencies in female sexual responding and comprehensive reviews have been published (see, for example, Chivers, 2005, 2017; Chivers et al., 2010). These theories include a range of methodological, biological, psychological and sociological factors. For example, potential differences could result from hormonal variations associated with fertility and menstrual cycle (Diamond, 2007; Shirazi et al., 2018), identification with the sexual pleasure that is perceived in other women leading to an arousal response to same-sex stimuli (Chivers, 2017), and greater malleability of sexuality by external contextual influences such as relationships (Baumeister, 2000). These possibilities have not been explored here (for a complete review of theories, see Chivers, 2017).

The comparatively small effect for stimuli of high sexual explicitness that was found in this meta-analysis could provide some support for the preparation hypothesis (Dawson et al., 2013; Lalumière et al., 2020), which suggests that any strong sexual cues provoke an indiscriminate arousal response to prepare a woman for the possibility of a sexual encounter (Chivers, 2017; Laan & Everaerd, 1995; Suschinsky & Lalumière, 2011). However, there is no clear explanation for the opposing effects of the mixed and low explicitness stimuli that were revealed in the current meta-analysis. Furthermore, the analysis considers difference scores of pupillary responses between male and female images, and from this it is therefore not possible to conclude whether the small effect is a result of no dilation *or* pupil dilation to a similar degree for both images. We also cannot exclude the possibility that other moderating factors could have contributed to this finding. For example, studies in both these subcategories included participants who were not asked to report their sexual orientation and were therefore assumed to be of heterosexual orientation (for example, De Winter et al., 2021; Hamel, 1974; Scott et al., 1967). Yet, the prevalence of same-sex behavior is reported to be higher in women compared to men (Diamond, 2016) with a difference of around 9% (7% of men versus 16% of women; Mercer et al., 2013). Thus, it is conceivable that the samples in these studies also contained a proportion of women who would have identified as bisexual or gay more so than samples of men. We note, however, that the moderator analysis also included measurement of sexual interest (i.e., whether researchers recorded

self-reported sexual interest or assumed participants were heterosexual) and this was not found to moderate effect sizes.

Taken together, the findings from the present meta-analysis suggest that differences in type of stimulus and measurement technique across studies might not only influence the degree of pupillary response effects of sexual interest, but also the direction of these effects. Additional primary studies are required to draw firm conclusions on the individual influence of any of the moderators examined in the current meta-analysis.

Limitations and Future Research Directions

Due to the low number of studies that were available, we were selective in the choices of variables that were used in the moderator analysis. Consequently, other potential variables associated with methodological techniques were not included in the meta-analysis, such as low-level stimulus characteristics. Image luminance, for example, has the potential to evoke bottom-up pupillary responses that might interfere with responses to the task-specific processes under investigation. One mechanism through which such interference could occur is the pupil light reflex, which refers to the constriction or dilation of the pupil in response to changes in the intensity of light levels that enter the eye (Ellis, 1981). The pupils begin to constrict within 200 ms after an increase in light, reaching a minimum diameter of approximately 2 mm at around 1500 ms (Mathôt, 2018). However, precise durations and pupil sizes vary depending on the intensity of the light change and individual differences (Ellis, 1981). In addition to luminance, color of the perceived light can also influence the duration of a constriction. For example, while blue light leads to sustained constriction of the pupil beyond 1500 ms, red light leads to a degree of re-dilation in the absence of light change, termed pupil escape (Mathôt, 2018). Variations in image luminance and color might, therefore, be a factor to consider when measuring pupillary responses to visual stimuli.

One method for controlling this factor is to equalize brightness across stimuli to match the mean luminance of the stimulus set (Attard-Johnson et al., 2016; Snowden et al., 2019). However, while this approach eliminates mean luminance differences across images, it could also create additional artifacts by introducing distortion of natural luminance levels *within* images, such as light–dark contrasts. Such image distortions might decrease the realism of images and could unwittingly interact with the attractiveness of a depicted person or draw the observer's attention to information in images that is irrelevant for the task at hand. In the study of pupillary responses as an index of sexual interest, only a few studies have so far adopted such luminance manipulations, with results that are difficult to assimilate, demonstrating a need for further research (cf. Attard-Johnson et al., 2016; Snowden

et al., 2019). However, most studies in this meta-analysis also demonstrate divergent pupillary responses to the same stimuli, based on differences in observers' sex and sexual orientation, indicating that low-level image characteristics are unlikely to explain the study findings.

Another factor that was not included in the meta-analysis here is the approach that is taken to analyze pupillary response data, which also varies across research studies. A recent paper compared four widely used methods for analyzing pupil responses (Attard-Johnson et al., 2019). Specifically, Attard-Johnson et al. analyzed (unadjusted) raw pupil scores obtained from eye-tracking, z-scores of this data, a conversion of the pupil data that captures differences between conditions in terms of the percentage change in eye pupil size, and pupillary responses that have been adjusted on the basis of a baseline measure. They reported that fundamental pupillary response patterns were consistent across all methods.

Finally, although it was possible to obtain information on all three methodological moderators from published sources or from direct contact with the authors, we could not acquire the actual correlations for pupil responses between male and female stimuli for some of the earlier studies included in the meta-analysis. Therefore, estimated correlations were used as a proxy when coding these studies. Actual correlation data produced a larger effect size compared to estimated correlations for heterosexual men. For heterosexual women, actual and estimated correlations produced effect sizes of similar magnitude but in different directions. In order to facilitate accumulation of knowledge, studies on pupil responses should provide correlations between sets of stimuli.

Implications for the Measurement of Sexual Interest

The current meta-analysis offers some tentative guidance for the measurement of pupillary responses of sexual interest. Across both heterosexual men and women, the low sexual explicitness stimuli (comprising partial or no nudity) appeared to provide clearer discrimination in pupillary response to sexual interest compared to high explicitness stimuli (complete nudity). We note, however, that for heterosexual women this effect was not in the direction consistent with their sexual orientation. Current evidence is too limited to determine whether some eye-tracking apparatus are better suited to this research. The desk-mounted EyeLink provided the strongest effect for heterosexual male observers, but this did not extend to heterosexual women for whom the largest effect was obtained for manual recording, and there were insufficient eligible studies for inclusion of remote Tobii eye trackers. Thus, further investigation comparing these methods is warranted, especially in women.

Furthermore, it is also possible that the affective state underlying these arousal patterns could be attributed to states evoked from other interpretations of the stimulus

being perceived by the observer as pleasant or unpleasant. For instance, an arousal response would also be evoked if the stimulus is perceived as threatening or unpleasant when viewing, for example, stimuli depicting illness or violence (Bradley et al., 2008). It may not be possible to establish for certain whether sexual arousal underpins pupil dilation responses to sexual stimuli. Equally, we cannot exclude the possibility that other 'pleasant' emotions, for instance affection or aesthetic appeal, could be underlying the pupil dilation responses to sexual stimuli. We would argue that it is implausible that, for example, heterosexual male arousal to female images is better explained by perceived threat or unpleasantness, or pleasant emotions given that correlations of pupillary responses to sexual stimuli with genital arousal patterns, subjective sexual arousal (Rieger et al., 2015) and sexual appeal ratings (Attard-Johnson et al., 2016). These correlations provide some evidence to support the notion that pupil dilation responses to sexual stimuli observed are related to sexual arousal—at least in men. However, we acknowledge that further research is needed to untangle the affective states underlying the arousal response when viewing these images.

This highlights the importance of obtaining confirmatory measure of sexual interest to validate against pupillary responses. In many studies this was obtained through subjective self-report of sexual orientation which, at least in men, demonstrates convergence with pupillary responses (Rieger et al., 2015), cognitive tasks (Ó Ciardha et al., 2018), and other physiological measures of sexual arousal (Chivers et al., 2010; Rieger et al., 2015). Few direct attempts have been made to examine the convergent validity of pupillary responses with self-report sexual orientation and cognitive-based measures of sexual interest (Attard-Johnson et al., 2016; Ó Ciardha et al., 2018; Rieger et al., 2015) and only one study with genital responses (Rieger et al., 2015). From these, men's pupillary responses demonstrated modest correlations with genital response (average $r = 0.59$; Rieger et al., 2015), viewing time (average $r = 0.51$; Ó Ciardha et al., 2018; Rieger & Savin-Williams, 2012), self-report sexual orientation (average $r = 0.56$; Ó Ciardha et al., 2018; Rieger et al., 2015; Rieger & Savin-Williams, 2012), and subjective sexual appeal and arousal ratings (average $r = 0.58$; Attard-Johnson et al., 2016; Rieger et al., 2015). In contrast, women's pupillary responses correlated weakly with genital response (average $r = 0.19$; Rieger et al., 2015), subjective sexual appeal, and arousal ratings (average $r = 0.15$; Attard-Johnson et al., 2016; Rieger et al., 2015), but demonstrated a moderate correlation with viewing time tasks (average $r = 0.47$; Ó Ciardha et al., 2018; Rieger & Savin-Williams, 2012) and self-report sexual orientation (average $r = 0.31$; Ó Ciardha et al., 2018; Rieger et al., 2015; Rieger & Savin-Williams, 2012). We recommend that future primary studies seeking to use pupillary responses as a measure of sexual interest include also one or more alternative measures of sexual orientation to strengthen our understanding of the validity of pupillary responses for the assessment of sexual interests in men and women.

Conclusion

While pupillary responses were first explored as a measure of observers' sexual interest in others in the 1960s, there has been an increase in research in this field with modern eye-tracking equipment in recent years. These studies vary in eye-tracking methodology and the approach to data analysis, stimulus materials, procedure, and participant characteristics. They also vary in their results, justifying the need for a meta-analysis of this field. The current study demonstrates that heterosexual and gay men show pupillary responses in line with their sexual orientation, whereas bisexual men exhibit greater pupil dilation during the viewing of male versus female stimuli. There were much fewer studies sampling women. In heterosexual women, differences in pupil response to male and female stimuli were small and non-significant, whereas lesbian women displayed greater dilation to male stimuli, and there were insufficient samples for bisexual women. These results therefore indicate that pupillary responses only provide a reliable index of the sexual interests of heterosexual and gay men. Although typical effect sizes for these pupillary responses are smaller compared to genital arousal measures, this approach offers a potentially valuable tool for measuring physiological arousal in a manner that is less invasive to participants and more accessible to researchers. However, the mixed findings, especially when moderators were included, also demonstrate a clear need to increase research on women, the convergent validity of this measurement technique, and for the development of standardized testing protocols and materials.

References

Aboyoun, D. C., & Dabbs, J. M. (1998). The Hess pupil dilation findings: Sex or novelty. *Social Behavior and Personality, 26*, 415–419. https://doi.org/10.2224/sbp.1998.26.4.415

Attard-Johnson, J., & Bindemann, M. (2017). Sex specific but not sexually explicit: Pupillary responses to dressed and naked adults. *Royal Society Open Science, 4*, 160963. https://doi.org/10.1098/rsos.160963

Attard-Johnson, J., Bindemann, M., & Ó Ciardha, C. (2016). Pupillary responses as an age-specific measure of sexual interest. *Archives of Sexual Behavior, 45*, 855–870. https://doi.org/10.0017/s10508-015-0681-3

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. *Journal of Sex Research, 54*, 1085–1096. https://doi.org/10.1080/00224499.2016.1241857

Attard-Johnson, J., Ó Ciardha, C., & Bindemann, M. (2019). Comparing methods for analysis of pupillary response. *Behavior Research Methods, 51*, 83–95. https://doi.org/10.3758/s13428-018-1108-6

Babchishin, K. M., Nunes, K. L., & Hermann, C. A. (2013). The validity of Implicit Association Test (IAT) measures of sexual attraction to children: A meta-analysis. *Archives of Sexual Behaviour, 42*, 487–499. https://doi.org/10.1007/s10508-012-0022-8

Bailey, J. M. (2009). What is sexual orientation and do women have one? In D. A. Hope (Ed.), *Contemporary perspectives on lesbian, gay and bisexual identities* (pp. 43–63). Springer.

Balduzzi, S., Rücker, G., & Schwarzer, G. (2019). How to perform a meta-analysis with R: A practical tutorial. *Evidence-Based Mental Health, 22*, 153–160. https://doi.org/10.1136/ebmental-2019-300117

Baumeister, R. F. (2000). Gender differences in erotic plasticity: The female sex drive as socially flexible and responsive. *Psychological Bulletin, 126*, 347–374. https://doi.org/10.1037/0033-2909.126.3.347

Bernick, N., Kling, A., & Borowitz, G. (1971). Physiological differentiation of sexual arousal and anxiety. *Psychosomatic Medicine, 33*, 341–352. https://doi.org/10.1097/00006842-197107000-00004

Borenstein, M. (2009). Effect sizes for continuous data. In H. M. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 221–235). Russel Sage Foundation.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). Fixed-effect versus random-effects models. *Introduction to meta-analysis*. John Wiley & Sons, Ltd.

Bradley, M. M., Miccoli, L., Escrig, M. A., & Lang, P. J. (2008). The pupil as a measure of emotional arousal and autonomic activity. *Psychophysiology, 45*, 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x

Brisson, J., Mainville, M., Mailloux, D., Beaulieu, C., Serres, J., & Sirois, S. (2013). Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. *Behavior Research Methods, 45*, 1322–1331. https://doi.org/10.3758/s13428-013-0327-0

Chivers, M. L. (2005). A brief review and discussion of sex differences in the specificity of sexual arousal. *Sexual and Relationship Therapy, 20*, 377–390. https://doi.org/10.1080/14681990500238802

Chivers, M. L. (2017). The specificity of women's sexual response and its relationship with sexual orientations: A review of ten hypotheses. *Archives of Sexual Behavior, 46*, 1161–1179. https://doi.org/10.1007/s10508-016-0897-x

Chivers, M. L., & Bailey, M. J. (2005). A sex difference in features that elicit genital response. *Biological Psychology, 70*, 115–120. https://doi.org/10.1016/j.biopsycho.2004.12.002

Chivers, M. L., Rieger, G., Latty, E., & Bailey, J. M. (2004). A sex difference in the specificity of sexual arousal. *Psychological Science, 15*, 736–744. https://doi.org/10.1111/j.0956-7976.2004.00750

Chivers, M. L., Seto, M. C., Lalumière, M. L., Laan, E., & Grimbos, T. (2010). Agreement of self-reported and genital measures of sexual arousal in men and women: A meta-analysis. *Archives of Sexual Behavior, 39*, 5–56. https://doi.org/10.1007/s10508-009-9556-9

Chivers, M. L., & Timmers, A. D. (2012). Effects of gender and relationship context in audio narratives on genital and subjective sexual response in heterosexual women and men. *Archives of Sexual Behavior, 41*, 185–197. https://doi.org/10.1007/s10508-012-9937-3

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.

Conaglen, H. (2004). Sexual content induce delay: A reexamination investigating relation to sexual desire. *Archives of Sexual Behavior, 33*, 359–367.

Dawson, S. J., & Chivers, M. L. (2018). The effect of static versus dynamic stimuli on visual processing of sexual cues in androphilic women and gynephilic men. *Royal Society Open Science, 5*, 172286. https://doi.org/10.1098/rsos.172286

Dawson, S. J., Suschinsky, K. D., & Lalumière, M. L. (2013). Habituations of sexual response in men and women: A test of the preparation hypothesis of women's genital response. *Journal of Sexual Medicine, 10*, 990–1000.

De Winter, J. C. F., Petermeljer, S. M., Kolljman, L., & Dodou, D. (2021). Replicating five pupillometry studies of Eckhard Hess. *International Journal of Psychophysiology*, *165*, 145–205. https://doi.org/10.1016/j.ijpsycho.2021.03.003

Diamond, L. M. (2007). A dynamical systems approach to the development and expression of female same-sex sexuality. *Perspectives on Psychological Science, 2*, 142–161. https://doi.org/10.1111/j.1745-6916.2007.00034.x

Diamond, L. M. (2016). Sexual fluidity in males and females. *Current Sexual Health Reports, 8*, 249–256. https://doi.org/10.1007/s11930-016-0092-z

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634. https://doi.org/10.1136/bmj.315.7109.629

Ellis, C. J. K. (1981). The pupillary reflex in normal subjects. *British Journal of Ophthalmology, 65*, 754–759. https://doi.org/10.1136/bjo.65.11.754

Finke, J. B., Behrie, A., & Schächinger, H. (2018). Acute stress enhances pupillary responses to erotic nudes: Evidence for differential effects of sympathetic activation and cortisol. *Biological Psychology, 137*, 73–82. https://doi.org/10.1016/j.biopsycho.2018.07.005

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

Garrett, J. C., Harrison, D. W., & Kelly, P. L. (1989). Pupillometric assessment of arousal to sexual stimuli: Novelty effects or preference? *Journal of Sexual Behavior, 18*, 191–201.

Hamel, R. F. (1974). Female subjective and pupillary reaction to nude male and female figures. *Journal of Psychology, 87*, 171–175. https://doi.org/10.1080/00223980.1974.9915687

Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology, 66*, 348–362.

Harrer, M., Cuijpers, P., Furukawa, T., & Ebert, D. D. (2019). *dmetar: Companion R package for the guide 'Doing meta-analysis in R'*. R package version 0.0.9000. Retrieved from http://dmetar.protectlab.org.

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., Seltzer, A. L., & Shlien, J. M. (1965). Pupil response of hetero- and homosexual males to pictures of men and women: A pilot study. *Journal of Abnormal Psychology, 70*, 165–168. https://doi.org/10.1037/h0021978

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557–560. https://doi.org/10.1136/bmj.327.7414.557

Gaze Intelligence. (2020). *SMI eyetracking support: Product manual*. Retrieved from https://gazeintelligence.com/smi-product-manual.

Jabbour, J. T., Hsu, K. J., & Bailey, J. M. (2020). Sexual arousal patterns of mostly heterosexual men. *Archives of Sexual Behavior, 49*, 2421–2429. https://doi.org/10.1007/s10508-020-01720-z

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*, 1583–1585. https://doi.org/10.1126/science.154.3756.1583

Kinsey, A. C., Pomeroy, W. B., & Martin, C. E. (1948). *Sexual behavior in the human male*. Saunders.

Laan, E., & Everaerd, W. (1995). Determinants of female sexual arousal: Psychophysiological theory and data. *Annual Review of Sex Research, 6*, 32–76. https://doi.org/10.1080/10532528.1995.10559901

Laeng, B., & Falkenberg, L. (2007). Women's pupillary responses to sexually significant others during the hormonal cycle. *Hormones and Behavior, 52*, 520–530. https://doi.org/10.1016/j.yhbeh.2007.07.013

Lalumière, M. L., Sawatsky, M. L., Dawson, S. J., & Suschinsky, K. D. (2020). The empirical status of the preparation hypothesis: Explicating women's genital responses to sexual stimuli in the laboratory. *Archives of Sexual Behavior*. https://doi.org/10.1007/s10508-019-01599-5

Light, R. J., & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Harvard University Press.

Lippa, R. A. (2012). Effects of sex and sexual orientation on self-reported attraction and viewing times to images of men and women: Testing for category specificity. *Archives of Sexual Behavior, 41*, 149–160. https://doi.org/10.1007/s10508-011-9898-y

Loewenfeld, I. E. (1958). Mechanisms of reflex dilatation of the pupil, historical review and experimental analysis. *Documenta Ophtalmologica, 12*, 185–448. https://doi.org/10.1007/BF00913471

Löwenstein, O., & Loewenfeld, I. E. (1958). Electronic pupillography: A new instrument and some clinical applications. *American Medical Association Archives of Ophthalmology, 59*, 352–363. https://doi.org/10.1001/archopht.1958.00940040058007

Löwenstein, O., & Loewenfeld, I. E. (1962). The pupil. Muscular mechanisms. In H. Dvidson (Ed.), *The eye* (Vol. 3, pp. 256–329). Academic Press.

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of Cognition, 1*, 1–23. https://doi.org/10.5334/joc.18

McDougal, D., & Gamlin, P. (2008). Pupillary control pathways. In R. H. Masland & T. Albright (Eds.), *The senses: A comprehensive reference* (Vol. 1, pp. 521–536). Academic Press. https://doi.org/10.1016/B978-012370880-9.00282-6

Mercer, C. H., Tanton, C., Prah, P., Erens, B., Sonnenberg, P., Clifton, S., Macdowall, W., Lewis, R., Field, N., Datta, J., Copas, A. J., Phelps, A., Wellings, K., & Johnson, A. M. (2013). Changes in sexual attitudes and lifestyles in Britain through the life course and over time: Findings from the National Surveys of Sexual Attitudes and Lifestyles (Natsal). *The Lancet, 382*, 1781–1794. https://doi.org/10.1016/S0140-6736(13)62035-8

Murphy, L., Ranger, R., Stewart, H., Dwyer, G., & Fedoroff, J. P. (2015). Assessment of problematic sexual interests with the penile plethysmograph: An overview of assessment laboratories. *Current Psychiatry Reports, 17*, 29. https://doi.org/10.1007/s11920-015-0567-z

Nunnally, J. C., Knott, P. D., Duchnowski, A., & Parker, R. (1967). Pupillary response as a general measure of activation. *Perception & Psychophysics, 2*, 149–155.

Ó Ciardha, C., Attard-Johnson, J. & Bindemann, M. (2018). Latency-based and psychophysiological measures of sexual interest show convergent andconcurrent validity. *Archives of Sexual Behavior, 47*, 637–649. https://doi.org/10.1007/s10508-017-1133-z

Peavler, S. W., & McLaughlin, J. P. (1967). The question of stimulus content and pupil size. *Psychonomic Science, 8*, 505–506. https://doi.org/10.3758/BF03331723

Peplau, L. A., & Garnets, L. D. (2000). A new paradigm for understanding women's sexuality and sexual orientation. *Journal of Social Issues, 56*, 329–350. https://doi.org/10.1111/0022-4537.00169

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.r-project.org/.

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. *Biological Psychology, 104*, 56–64. https://doi.org/10.1016/j.biopsycho.2014.11.009

Rieger, G., Chivers, M. L., & Bailey, J. M. (2005). Sexual arousal patterns of bisexual men. *Psychological Science, 16*, 579–584. https://doi.org/10.1111/j.1467-9280.2005.01578.x

Rieger, G., Rosenthal, A. M., Cash, B. M., Linsenmeier, J. A. W., Bailey, J. M., & Savin-Williams, R. C. (2013). Male bisexual arousal: A matter of curiosity? *Biological Psychology, 94*, 479–489. https://doi.org/10.1016/j.biopsycho.2013.09.007

Rieger, G., & Savin-Williams, R. C. (2012). The eyes have it: Sex and sexual orientation differences in pupil dilation patterns. *PLoS ONE, 7*, e40256. https://doi.org/10.1371/journal.pone.0040256

Rieger, G., Savin-Williams, R. C., Chivers, M. L., & Bailey, J. M. (2016). Sexual arousal and masculinity–femininity of women. *Journal of Personality and Social Psychology, 111*, 265–283. https://doi.org/10.1037/pspp0000077

Rosenthal, A. M., Sylva, D., Safron, A., & Bailey, J. M. (2011). Sexual arousal patterns of bisexual men revisited. *Biological Psychology, 88*, 112–115. https://doi.org/10.1016/j.biopsycho.2011.06.015

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley. https://doi.org/10.1002/0470870168

Schnelle, J. P., Kennedy, M., Rutledge, J. P., & Golden, S. B. (1974). Pupillary response as indication of sexual preference in a juvenile correctional institution. *Journal of Clinical Psychology, 30*, 146–150.

Schulze, R. (2007). Current methods for meta-analysis: Approaches, issues, and developments. *Zeitschrift für Psychologie, 215*, 90–103. https://doi.org/10.1027/0044-3409.215.2.90

Scott, T. R., Wells, W. H., Wood, D. Z., & Morgan, D. I. (1967). Pupillary response and sexual interest re-examined. *Journal of Clinical Psychology, 23*, 433–438. https://doi.org/10.1002/1097-4679(196710)

Semon, T. L., Hsu, K. J., Rosenthal, A. M., & Bailey, J. M. (2017). Bisexual phenomena among gay-identified men. *Archives of Sexual Behavior, 46*, 237–245. https://doi.org/10.1007/s10508-016-0849-5

Shirazi, T. N., Bossio, J. A., Puts, D. A., & Chivers, M. L. (2018). Menstrual cycle phase predicts women's hormonal responses to sexual stimuli. *Hormones and Behavior, 103*, 45–53. https://doi.org/10.1016/j.yhbeh.2018.05.023

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General, 143*, 534–547. https://doi.org/10.1037/a0033242

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). p-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9*, 666–681. https://doi.org/10.1177/1745691614553988

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better p-curves: Making p-curve analysis more robust to errors, fraud, and ambitious p-hacking, a reply to Ulrich and Miller (2015). *Journal of Experimental Psychology: General, 144*, 1146–1152. https://doi.org/10.1037/xge0000104

Slettevold, E., Holmes, L., Gruia, D., Nyssen, C. P., Watts-Overall, T. M., & Rieger, G. (2019). Bisexual men with bisexual and monosexual genital arousal patterns. *Biological Psychology, 148*, 107763. https://doi.org/10.1016/j.biopsycho.2019.107763

Snowden, R. J., McKinnon, A., Fitoussi, J., & Gray, N. S. (2019). Pupillary response to static images of men and women: A possible measure of sexual interest? *Journal of Sex Research, 56*, 74–84. https://doi.org/10.1080/00224499.2017.1394959

Song, F., & HooperLoke, L. Y. (2013). Publication bias: What is it? How do we measure it? How do we avoid it? *Open Access Journal of Clinical Trials*. https://doi.org/10.2147/OAJCT.S34419

SR Research Ltd. (2005–2008). EyeLink User Manual (version 1.4.0).

Steinhauer, S. R., Siegle, G. J., Condray, R., & Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *International Journal of Psychophysiology, 52*, 77–86. https://doi.org/10.1016/j.ipsycho.2003.12.005

Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology, 54*, 1046–1055. https://doi.org/10.1016/S0895-4356(01)00377-8

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *British Medical Journal, 343*, 1–8. https://doi.org/10.1136/bmj.d4002

Suschinsky, K. D., & Lalumière, M. L. (2011). Prepared for anything? An investigation of female genital arousal in response to rape cues. *Psychological Science, 22*, 159–165. https://doi.org/10.1177/0956797610394660

Titz, J., Scholz, A., & Sedlmeier, P. (2018). Comparing eye trackers by correlating their eye-metric data. *Behavior Research Methods, 50*, 1853–1863. https://doi.org/10.3758/s13428-017-0954-y

Tobii, A. B. (2016). *Tobii studio: User's manual*. Retrieved from https://www.tobiipro.com/siteassets/tobii-pro/user-manuals/tobii-pro-studio-user-manual.pdf.

Tollison, C. D., Adams, H. E., & Tollison, J. W. (1979). Cognitive and physiological indices of sexual arousal in homosexual, bisexual, and heterosexual males. *Journal of Behavioral Assessment, 1*, 305–314. https://doi.org/10.1007/BF01321372

van Aert, R. C. M., Wicherts, J. M., & van Assen, M. A. L. M. (2016). Conducting meta-analyses based on p values: Reservations and recommendations for applying p-uniform and p-curve. *Perspectives on Psychological Science, 11*, 713–729. https://doi.org/10.1177/1745691616650874

Vevea, J. L., & Coburn, K. (2019). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (3rd ed., pp. 383–432). Russell Sage Foundation.

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods, 10*, 428–443. https://doi.org/10.1037/1082-989X.10.4.428

Watts, T. M., Holmes, L., Raines, J., Orbell, S., & Rieger, G. (2018). Sexual arousal patterns of identical twins with discordant sexual orientations. *Scientific Reports, 8*, 14970. https://doi.org/10.1038/s41598-018-33188-2

Watts, T. M., Holmes, L., Savin-Williams, R. C., & Rieger, G. (2017). Pupil dilation to explicit and non-explicit sexual stimuli. *Archives of Sexual Behavior, 46*, 155–165. https://doi.org/10.1007/s10508-016-0801-8

Welton, N. J., Sutton, A. J., Cooper, N., & Ades, A. E. (2012). *Statistics in practice: Evidence synthesis for decision making in healthcare*. Wiley.

Xu, Y., Rahman, W., & Zheng, Y. (2016). Gender-specificity in viewing time among heterosexual women. *Archives of Sexual Behavior, 46*, 1361–1374. https://doi.org/10.1007/s10508-016-0795-2

Zuckerman, M. (1971). Physiological measures of sexual arousal in the human. *Psychological Bulletin, 75*, 297–329.