Examining Major Threats to Valid Measurement of Sexual Motivation: From Basic Research to Implications for Society and Science

---

Dissertation

zur Erlangung des akademischen Grades

eines Doktors der Naturwissenschaften

der Fakultät für Empirische Humanwissenschaften und

Wirtschaftswissenschaft

der Universität des Saarlandes

---

vorgelegt von

Marcel Weber

aus Landstuhl

Saarbrücken, 2025

Dekan:	Univ.-Prof. Dr. Peter Loos, Universität des Saarlandes
Berichterstatter:	Univ.-Prof. Dr. Malte Friese, Universität des Saarlandes
	Univ.-Prof. Dr. Cornelius J. König, Universität des Saarlandes
Tag der Disputation:	11.12.2024

Summary

Sexual motivation plays a significant role in almost everyone's life, influencing thoughts, feelings, behaviors, and decisions. Self-report measures are the primary means of unraveling the secrets of sexual motivation in the social sciences, allowing large-scale assessments of what goes on in people's minds and behind closed doors. However, the validity of conclusions drawn from sexual self-report measures is controversial, particularly with regard to gender differences in sexual motivation. While the literature highlights various biases that could potentially threaten the valid measurement of sexual motivation, little is known about their actual impact. In this dissertation, I fill some of this gap by identifying and addressing critical threats to valid measurement of sexual motivation.

Part I focused on how the measurement instrument affects validity. To date, most sexual motivation scales are vague in their conceptualization and lack a comprehensive validation process. In addition, most scales are regularly used for group comparisons (e.g., men vs. women) without demonstrating equivalent measurement properties across these groups. These factors lead to uncertainty about whether the scales measure sexual motivation and whether seemingly robust gender differences actually exist. We therefore developed and validated a novel scale of sexual motivation: the *Trait Sexual Motivation Scale* (*TSMS*). Across four preregistered studies, the theory-driven TSMS emerged as a reliable and valid measure of sexual motivation independent of gender and relationship status. Using this scale, we found higher male than female sexual motivation. This suggests that gender differences are not due to invalid group comparisons.

Part II focused on social desirability bias as a threat to valid measurement and conclusions. Because of gendered sexual norms, previous research has expected men to overreport and women to underreport sexual experiences, which may account for much of the measured gender differences. However, support for these predictions is scarce and largely limited

to laboratory research. In theory, the greater anonymity provided by online surveys promises more accurate responses, but little is known about whether this hope is justified for sexual self-reports. Here, we used the *Item Sum Technique*, an indirect questioning technique that maximizes people's anonymity, to create conditions under which honest self-reports are particularly likely. In this item sum group, sexual motivation and gender differences in sexual motivation were not significantly different from those found in a standard online survey group. These results suggest that there is little evidence of social desirability bias in online surveys and argue against the notion that such bias explains higher self-reported sexual motivation in men than in women.

Part III focused on the social norms that give rise to social desirability bias. Previous research predicted that sexual activity would be evaluated in opposite ways for women (socially punished) and men (socially rewarded), but this strong sexual double standard received little empirical support. We proposed an alternative model of sexual standards—*the Similarities and Differences (S&D) model*—that may explain the inconsistent findings by suggesting that male and female sexual norms are marked by both similarities and differences. Consistent with both models, participants perceived that high sexual activity is viewed more favorably for men than for women, while low sexual activity is seen more positively for women than for men. However, they also perceived that moderate levels of sexual activity, rather than very low or high levels, are viewed most favorably for both genders—a similarity predicted only by the S&D model.

In sum, using the TSMS in online studies promises valid future measurement of sexual motivation. The lack of evidence that higher sexual motivation in men is due to invalid scales or self-presentation contributes to debates about gender differences. In the general discussion, I seek to reconcile these findings with research showing no gender differences, present a new model of flexible self-presentation, and discuss the TSMS and S&D model's practical relevance. In doing so, I link basic research on sexual motivation with its broader scientific and societal implications.

Zusammenfassung

Sexuelle Motivation spielt im Leben fast aller Menschen eine wichtige Rolle. Sie beeinflusst Gedanken, Gefühle, Verhalten und Entscheidungen. Um umfassend zu ergründen, was in den Köpfen von Personen und hinter verschlossenen Türen vor sich geht, sind sexuelle Selbstberichte elementar. Gleichzeitig wird kontrovers diskutiert, inwiefern Befunden, die auf sexuellen Selbstberichten basieren, Vertrauen geschenkt werden darf. Während in der Literatur mehrfach Verzerrungen beschrieben wurden, welche die Validität sexueller Selbstberichte potenziell gefährden *könnten*, ist wenig darüber bekannt, ob diese tatsächlich zu falschen Schlussfolgerungen führen. Diese Dissertation zielt darauf ab, einen Teil dieser Lücke zu schließen, indem sie wesentliche Bedrohungen für die valide Erfassung sexueller Motivation identifiziert und untersucht.

Teil I beschäftigte sich damit, welche Rolle das Messinstrument bei der validen Erfassung sexueller Motivation spielt. Aktuell kommen meist Skalen zum Einsatz, deren zugrundeliegende Konzeptualisierung vage und deren Validierungsprozess unvollständig ist. Zudem werden diese Skalen häufig für Gruppenvergleiche (z. B. zwischen Männern und Frauen) verwendet, ohne zu prüfen, ob sie für diese Gruppen vergleichbare Messeigenschaften aufweisen. Diese Versäumnisse führen zu Unsicherheiten darüber, ob die Skalen wirklich sexuelle Motivation erfassen und ob die scheinbar robusten Geschlechtsunterschiede tatsächlich existieren. Wir haben daher eine neue Skala sexueller Motivation entwickelt und validiert: die *Trait Sexual Motivation Scale* (*TSMS*). In vier präregistrierten Studien zeigte sich die theoriegeleitete TSMS unabhängig vom Geschlecht und Beziehungsstatus einer Person als reliables und valides Messinstrument sexueller Motivation. Unter Einsatz der TSMS fand sich eine höhere mittlere sexuelle Motivation bei Männern als bei Frauen. Dies deutet darauf hin, dass Geschlechterunterschiede nicht auf unzulässige Gruppenvergleiche zurückzuführen sind.

Teil II beschäftigte sich damit, wie sozial erwünschtes Antwortverhalten die Validität sexueller Selbstberichte und darauf basierender Schlussfolgerungen beeinflusst. Aufgrund geschlechtsspezifischer sexueller Normen wurde in der bisherigen Forschung angenommen, dass Männer in ihren sexuellen Erlebnisberichten eher übertreiben und Frauen eher untertreiben. Diese Tendenzen könnten einen erheblichen Teil gefundener Geschlechterunterschiede erklären. Die Belege für diese Annahmen sind jedoch spärlich und stammen fast ausschließlich aus Laborstudien. Theoretisch sollte das hohe Ausmaß an Anonymität, das Online-Studien kennzeichnet, unverzerrte Selbstberichte begünstigen. Ob sich diese Hoffnungen erfüllen, ist für sexuelle Selbstberichte allerdings weitgehend ungeklärt. Hier wurde die *Item Sum Technique* verwendet, eine indirekte Befragungstechnik, welche die Anonymität der Teilnehmenden maximiert. Dadurch sollen Bedingungen geschaffen werden, unter denen wahrheitsgemäße Selbstberichte besonders wahrscheinlich sind. Die auf diese Weise erfasste sexuelle Motivation und entsprechende Geschlechterunterschiede wurden mit denjenigen aus einer Standard-Online-Befragung verglichen. Es zeigten sich keine signifikanten Unterschiede. Diese Ergebnisse liefern wenig Hinweise auf sozial erwünschtes Antwortverhalten in Online-Befragungen und stellen die Annahme infrage, dass diese Verzerrungen für die höhere gemessene sexuelle Motivation bei Männern im Vergleich zu Frauen verantwortlich sind.

Teil III befasste sich mit sozialen Normen als Grundlage für sozial erwünschtes Antwortverhalten. Frühere Studien sagten voraus, dass sexuelle Aktivität bei Männern sozial belohnt und bei Frauen sozial bestraft wird. Für diese gegensätzlichen männlichen und weiblichen Sexualnormen, die als starker sexueller Doppelstandard bezeichnet werden, fanden sich jedoch nur wenige Belege. Wir haben ein alternatives Modell sexueller (Doppel-)Standards aufgestellt: das *Similarities and Differences Model* (*S&D-Modell*). Dieses sagt vorher, dass männliche und weibliche Sexualnormen sowohl Ähnlichkeiten als auch Unterschiede aufweisen

und könnte so Inkonsistenzen in früherer Forschung erklären. Gemäß der Wahrnehmung der Studienteilnehmenden wird in der Gesellschaft ein hohes Ausmaß sexueller Aktivität bei Männern positiver beurteilt als bei Frauen und umgekehrt ein geringes Ausmaß sexueller Aktivität bei Frauen positiver beurteilt als bei Männern. Diese Befunde sind sowohl mit dem starken sexuellen Doppelstandard als auch mit dem S&D-Modell vereinbar. Allerdings nahmen die Teilnehmenden ebenfalls wahr, dass nicht besonders niedriges oder hohes, sondern ein moderates Ausmaß sexueller Aktivität bei beiden Geschlechtern am besten beurteilt wird – eine Ähnlichkeit männlicher und weiblicher Sexualnormen, die nur das hier präsentierte S&D-Modell erklären kann.

Zusammenfassend verspricht der Einsatz der TSMS in Online-Studien eine valide Erfassung sexueller Motivation in zukünftiger Forschung. Für die Annahme, dass die höhere sexuelle Motivation von Männern im Vergleich zu Frauen auf die Verwendung nicht valider Skalen oder auf sozial erwünschtes Antwortverhalten zurückzuführen ist, konnten in unseren Studien keine Belege gefunden werden. Diese Ergebnisse liefern einen wertvollen Beitrag zur anhaltenden Debatte über Geschlechterunterschiede. In der allgemeinen Diskussion stelle ich die aktuellen Ergebnisse solchen Studien gegenüber, die keine Geschlechterunterschiede gefunden haben. Darüber hinaus stelle ich ein neues Modell der flexiblen Selbstpräsentation vor und diskutiere die praktische Relevanz der TSMS und des S&D-Modells. Auf diese Weise schlage ich eine Brücke zwischen der Grundlagenforschung zur sexuellen Motivation und ihren breiteren Implikationen für Wissenschaft und Gesellschaft.

Acknowledgments

This dissertation may bear my name, but it is just as much the product of the wonderful people who have accompanied me along the way. I will forever be grateful to my supervisor, Malte Friese, for his unwavering support, his kind and appreciative nature, and all that I have learned from him, both personally and professionally. I also wish to extend my thanks to Cornelius J. König for agreeing to review this thesis.

Thank you, Alex, Dorota, Hannah, Helena, Janet, Kathi, Lasse, Lina, Lisa, Max, Moritz, and Stefanie for being the best colleagues anyone could ever hope for. You were always there for me—supporting me, cheering me on, and making me laugh a million times. I will treasure these moments forever! Another big thank you goes out to my wonderful student assistants, Paula and Tobias, for all their invaluable support. Thanks also to Benedikt, David, and Julius. Without your guidance and encouragement, I might never have embarked on this wonderful journey.

Lastly, I would like to express my deepest gratitude to my family and friends for their unconditional support. Thank you to my five(!) wonderful grandparents, whose footprints will be felt forever, even though they can no longer be here today. To my mom and dad, I thank you from the bottom of my heart for always giving me the freedom to find and follow my own path and for standing by me unconditionally. Dominik, you have been so much along the way—listener, advisor, and most importantly, brother—and you have filled each of these roles wonderfully. Ingrid, Michael and Susanne, Yannick and Lena: Thank you for always making me feel at home when I'm with you. Chrisi, Matthias, Paddi, Steffi, and Stephan: Thank you for distracting me when needed, for cheering me up, and for your endless understanding. I couldn't ask for better friends, you guys are amazing! Finally, and most importantly, this last thank you is for you, Kristin. My biggest support, my love, my wife—the most important person in my life. Thank you for everything!

Index of Publications

The main parts of this publication-oriented dissertation ("publikationsorientierte Dissertation") are based on three manuscripts, all of which are in first authorship of the author of this dissertation. At the time of submission of this dissertation, one of the manuscripts has been published (Part I), one has been submitted for publication (Part II), and one has been accepted for publication (Part III).

Part I: Weber, M., Reis, D., & Friese, M. (2024). Development and validation of the Trait Sexual Motivation Scale (TSMS). *Journal of Personality Assessment, 106*, 267-282. https://doi.org/10.1080/00223891.2023.2206896

Part II: Weber, M., Kilger, H., & Friese, M. (2024). *How valid are self-reports of sexual motivation? Using the item sum technique to examine self-presentation tendencies in online research* [manuscript submitted for publication in *Motivation Science*].

Part III: Weber, M., & Friese, M. (in press). Sexual (double) standards revisited: Similarities and differences in the societal evaluation of male and female sexuality. *Social Psychological and Personality Science*. https://doi.org/10.1177/19485506241237288

Content

Introduction

Calling sexual motivation a construct of universal relevance is hardly an exaggeration. Defined as the intrinsic interest in sexual activity and sexual pleasure, it is the inner engine moving people's pursuit for sex (Frankenbach et al., 2022; Stark et al., 2015). Sexual motivation is one of the pillars of human reproduction (Zuk & Simmons, 2018, p. 104) and a key predictor of sexual and romantic satisfaction (Kim et al., 2021; Muise et al., 2013). For those who start dating, higher sexual motivation is associated with stronger romantic interest (Eastwick et al., 2023). For partners who are in a romantic relationship, discrepancies in sexual motivation are a frequent reason for couple therapy (Henry & Miller, 2004). In short, sexual motivation helps the species to survive and the individual to thrive.

Most of what we (think to) know about sexual motivation comes from studies in which participants answered questions about their sexual experiences and behaviors. There are good reasons for this elevated position of sexual self-reports. Sexual motivation cannot be reliably observed from the outside, as most events take place in the mind or behind closed doors. Physiological indicators such as penile circumference or vaginal blood volume are an important addition to the assessment repertoire, but they are invasive, expensive, and mostly limited to an assessment in artificial laboratory environments. Thus, for pragmatic reasons alone, self-report measures play a significant role in the large-scale assessment of sexual motivation.

However, concerns about the validity of conclusions drawn from sexual self-reports are pervasive. Conceptual and empirical deficits result in uncertainty about whether items and instruments measure sexual motivation, constituting a *defining threat* to validity. If the meaning of what is measured is different between groups, researchers face the *delicate threat* of comparing apples with oranges. Tendencies to over- or underreport sexual events to conform to (gendered) sexual norms are a *social threat* to validity.

These aspects are in no means technical details. They all pose a real danger of drawing false conclusions about the nature of human sexuality. Knowledge about how much these threats *actually* undermine validity is critical but in large parts a matter of speculation to this point. One very visible consequence of this uncertainty are heated debates about gender differences in sexual motivation. Numerous studies have found higher average sexual motivation in men than in women (Baumeister et al., 2001; Frankenbach et al., 2022). However, these studies have not ended the debate about gender differences, nor will future studies until the extent to which these differences are a product of bias is understood.

This thesis aims to reduce the uncertainty surrounding valid measurement of and valid conclusions about sexual motivation. I first introduce three major threats to validity. I then present three research projects (Parts I to III) that help estimate how these threats shape sexual self-reports and therefore lead to biased conclusions. Part I focuses on threats inherent to insufficiently validated scales and presents the Trait Sexual Motivation Scale (TSMS) as a solution. Part II addresses people's reluctance to respond accurately to sexual questions and examines whether online surveys can provide an anonymous environment that facilitates valid self-reports of sexual motivation. Part III aims to promote a better understanding of the perceived societal norms that are likely to give rise to social desirability bias. Throughout these parts, a special emphasis is on the research question that is emblematic of the far-reaching consequences that validity-related uncertainties can have for research and practice: whether women and men differ in their sexual motivation.

Gender Differences in Sexual Motivation

In the often rational world of science, I have come across few debates as passionate, sometimes heated, as those surrounding gender differences in sexual motivation. Despite the multidimensional nature of gender (Strang et al., 2023), the question is discussed almost

exclusively in terms of two genders: Is the average sexual motivation of men higher than that of women?

An accurate answer to this question has important scientific and societal implications. Entire theories are based on the assumption that men have a higher sexual motivation than women. If this is a misconception, they would collapse or require major revision (Frankenbach et al., 2022). Pronounced gender differences in sexual motivation imply that discrepancies in sexual motivation are common in mixed-sex relationships, indicating a potential risk to relationship satisfaction and well-being (Davies et al., 1999; Mark, 2014).

Among researchers, there seem to be two irreconcilable camps: those who believe that the question has long been answered clearly and affirmatively (e.g., Baumeister et al., 2001), and others who doubt the correctness of this conclusion (e.g., Conley et al., 2011). Online reactions to a publication in which colleagues and I meta-analytically approached this much-discussed question (Frankenbach et al., 2022) show a comparable polarization in the general population (Reddit, 2022). How can such a simple group comparison lead to such different interpretations?

Possible heterogeneity in the empirical findings cannot provide a basis here. Higher average male than female sexual motivation was consistently found in an extensive narrative review with numerous indicators (Baumeister et al., 2001), in a large-scale international project covering 53 countries (Lippa, 2009), and in a comprehensive meta-analysis summarizing data from over 600,000 people (Frankenbach et al., 2022). At the beginning of the millennium, Baumeister and colleagues (2001) concluded that the “combined quantity, quality, diversity, and convergence of the evidence render the conclusion [that men have higher average sexual motivation than women] indisputable”.

This conclusion has not gone unchallenged. Despite the robustness of the findings, strong arguments have been made that these measured gender differences are (in part) not valid, but

rather the result of methodological flaws and biases (Conley et al., 2011; Dawson & Chivers, 2014). In what follows, I introduce and address three major threats to valid measurement of sexual motivation in general, and to valid conclusions about gender differences in sexual motivation in particular.

Threats to the Validity of Sexual Self-Reports

The attempt to understand the world and the search for truth are guiding principles of science. From a test theory perspective, this is reflected in the elevated position of validity among the criteria defining the quality of empirical research. The prominent understanding of (construct) validity as the overlap between what is measured (i.e., the *measurement level*) and what should be measured (i.e., the *construct level*) can serve as a feasible working definition in this thesis (Pospeschill, 2022, p. 24). Conversely, threats to validity are factors that may weaken the direct correspondence between the construct level and the measurement level, potentially leading to false conclusions about the world.

The Defining Threat: Wait … Are We Actually Measuring Sexual Motivation?

A convincing demonstration of validity is a critical feature of any measure. Cronbach and Meehl (1955), who have been pioneers in research on validity, posited that construct validation is necessary whenever "an investigator believes that his instrument reflects a particular construct, to which are attached certain meanings." While there is no definite and universal procedure for demonstrating validity, there are steps that have been highlighted by several theorists (e.g., Flake et al., 2017; Loevinger, 1957; Simms, 2008). In the structural phase of validation, researchers typically examine the reliability of the measure (e.g., internal consistency, test-retest reliability) and its factor structure, using approaches to reduce the complexity in the data (e.g., exploratory or confirmatory factor analysis). In the external phase of validation, it is common practice to show that the measure correlates in expected ways with other measures—that is, it correlates weakly

with measures of theoretically unrelated constructs (i.e., discriminant validity) and correlates highly with measures of the same or very similar constructs (i.e., convergent validity). In addition, criterion validity can be assumed when the measure predicts important external criteria that are assessed at the same time (i.e., concurrent validity) or with some delay (i.e., prospective validity). Despite the undisputed importance of these steps, comprehensive validation processes are generally the exception rather than the norm (Flake et al., 2017), and this is no different for measures of sexual motivation in particular. In fact, the impressive number of measures used to study sexual motivation is dominated by items and scales that have undergone no or incomplete validation processes (see Stark et al., 2015).

Although much less considered than the empirical steps, conceptual requirements are no less important. As the concept of validity is about the overlap between what is and what should be measured, a clear theoretical conceptualization of the construct under study (e.g., how is sexual motivation supposed to manifest?) should be an integral element of any validation process. Without this conceptualization, it is difficult to draw the line between phenomena that are direct indicators of the construct and those that reflect related constructs or external criteria (Frankenbach et al., 2022). For instance, suppose that Emily and Samantha do not differ in how frequently they think about sex, desire sex, and have sex. However, Emily enjoys a greater variety of sexual activities and holds more positive attitudes towards sex than Samantha. Would you say that these two persons differ in their overall strength of sexual motivation? Some might argue that the phenomena that distinguish the two reflect external constructs such as personal preferences, past experiences, and internalized cultural knowledge rather than sexual motivation (Frankenbach et al., 2022). Others have used these phenomena as indicators of sexual motivation (Baumeister et al., 2001). To resolve this ambiguity, an underlying conceptualization is essential, and conceptual vagueness is harmful.

For sexual motivation and its measures, there *is* a great deal of conceptual vagueness. While there is broad consensus on how to define the construct[1] (e.g., Baumeister et al., 2001; Frankenbach et al., 2022; Spector et al., 1996; Stark et al., 2015), few theory-driven suggestions have been made as to what constitutes the elements of sexual motivation. Without the integration of theory and empiricism that is a core element of validity, researchers face the *defining threat* that what is measured may not adequately reflect the target construct (e.g., sexual motivation). This can have serious implications for conclusions relevant to practice, such as biased associations between sexual motivation and measures of other constructs.

Some colleagues and I have recently presented a conceptualization that may provide a viable basis for valid measurement of sexual motivation (Frankenbach et al., 2022). We situated sexual motivation within the broader context of personality research by applying two general principles about the nature of traits. First, we propose that sexual motivation is a latent construct that manifests itself in cognitive, affective, and behavioral events such as sexual fantasies, desire, and self-stimulation (Roberts, 2009). Second, we propose that substantial within-person variation in state sexual motivation does *not* imply that it is inappropriate to consider between-person variation in trait levels of sexual motivation. Instead, we adopt an integrated state/trait view according to which the mean across a person's state-level manifestations can be used as a reliable estimate of that person's trait sexual motivation (Fleeson, 2001, 2004). In short, a person high in trait sexual motivation is someone who often thinks about sex, desires sex, and has sex.

The Delicate Threat: Comparing Apples with Oranges

As a construct of universal relevance, sexual motivation has been studied across multiple populations including different age groups (from childhood to the elderly; Li, 2022; Marshall,

[1] Adding another layer of complexity, the construct that we refer to as sexual motivation appears under various other names in the literature, including sex drive, sexual desire, sexual interest, and libido (see Spector et al., 1996). Throughout all Parts of this manuscript, I will consistently use the term "sexual motivation."

2012) and combinations of sexual orientation and gender identity (from individuals who self-identify as cisgender and heterosexual to those who self-identify along the queer spectrum; e.g., Goldey et al., 2024; Skorska et al., 2023).

To ensure that a measure qualifies for use in and comparison across groups, researchers should provide evidence that the measure has very similar psychometric properties across these groups. This equivalence of measures for different groups of people is typically referred to as measurement invariance (Greiff & Scherer, 2018; Putnick & Bornstein, 2016). Critically, measurement invariance is not about whether there is a difference between groups, which is an open empirical question. Rather, measurement invariance is about whether a measure can be appropriately used in and compared across these groups in the first place. It is a critical prerequisite for many mathematical operations, especially for calculating mean differences between groups, but it is rarely tested empirically.

Two factors for which sexual motivation is particularly likely to manifest itself in somewhat different ways are gender and relationship status. Imagine that Arthur and Lisa participate in a survey about sexual motivation. Three of the items presented to them are “How strong is your desire to engage in sexual activity with a partner?,“ “I am constantly looking for a new sex partner,” and “It is easy for me to initiate sexual activities”. Arthur, who has been single for a while, wonders what “a partner” means: a person he could have a relationship with? Any sexual partner, including a stranger he has just met? Similarly, does the third item imply that he could easily approach a person he does not know to have sex with them? Although he disagrees with the second item, he is glad that “a new partner” leaves little room for interpretation. As a person who is in a romantic relationship and married, Lisa has no difficulty answering the first item, interpreting “a partner” as a reference to her spouse. She finds the second item a bit irritating: wouldn’t agreeing with this statement mean that she wants to cheat on her partner? She

chooses “strongly disagree”. She also hesitates to agree with the third item because she has learned that men should make the first move. She believes that she wants sex a little more often than her partner, but she tries to get him in the mood rather than initiating sexual activities herself. Both Arthur and Lisa felt that some of the items were unclear and did not capture their true strength of sexual motivation.

The characters and scenarios are fictional, but the threat is real. The interpretation of these items and how well their content represent a person’s sexual motivation is likely to differ between people who are single versus in a romantic relationship, and between men and women. The first item is part of the Sexual Desire Inventory (SDI, Spector et al., 1996), which is one of the most commonly used measures of sexual motivation. It has been used in dozens of studies and exists in numerous language versions, many of which have undergone validation (e.g., Callea & Rossi, 2021; Kuhn et al., 2014; Vallejo-Medina et al., 2020). The second and third items can be found in the Trait Sexual Motivation Questionnaire (TSMQ, Stark et al., 2015), a thoroughly validated scale that distinguishes between different facets of sexual motivation. These and other self-report scales have contributed greatly to a better understanding of sexual motivation. However, any uncertainty about whether these scales have similar psychometric properties across gender and relationship status limits their widespread use and their convincing demonstration of validity. For example, some researchers who have recognized these issues have chosen to change the wording of items or exclude partner-related items altogether when studying single people (Park & MacDonald, 2022; Vallejo-Medina et al., 2020). While plausible, these workarounds are unsatisfactory because they prevent the comparability of results across subgroups.

Particularly high levels of equivalence, referred to as strong or scalar measurement invariance, are required when a measure is not simply to be administered to different groups, but is to be used to compare the means of these groups. This is directly relevant to the validity of

measured gender differences in sexual self-reports. If the psychometric properties depend on gender, researchers may face the *delicate threat* of comparing "apples with oranges" (Greiff & Scherer, 2018), and seemingly robust gender differences might not reflect true differences, but rather be an artificial product of a lack of equivalence (Nye & Drasgow, 2011).

The Social Threat: Self-Presentation Entering the Stage

In addition to these conceptual and methodological concerns, there are threats to validity that are attributable to the interplay between the person answering the sexual questions and the situational context in which they do it. Because of the socially sensitive nature of sexuality, people may be reluctant to answer sexual questions completely accurately. In many situations, it may be tempting to present oneself in a way that is likely to be perceived positively by others. This type of response bias, often referred to as social desirability bias (Paulhus, 1988), is a permanent threat to the validity of sexual self-reports. Readers are invited to take a quick self-test: If asked in public, how easy would it be for you to be honest about how many hours of sleep you got last night? And what if the question was instead how many people you have slept with in your life?

Previous findings suggest that sexual self-presentation is not limited to real-life contexts but can also be found in scientific studies. For instance, Meston and colleagues (1998) found substantial correlations between sexual self-reports reported in the laboratory and social desirability scales assessing tendencies toward impression management and self-deception. These associations occurred despite the researchers' best intentions to minimize socially desirable responding (e.g., large testing room, visual barriers between the tables). Self-presentation tendencies pose a critical and *social threat* to validity. Biased frequency estimates of sexual events and biased associations between sexual motivation and other constructs are two of the ways in which socially desirable responding can lead to flawed conclusions.

However, self-presentation has probably received the most attention in the literature as a possible basis for the gender differences found in sexual self-reports. Previous research predicted that men and women who answer sexual questions are likely to adjust their responses in opposite ways, "such that men may be motivated to […] exaggerate the frequency and variability of their sexual encounters, whereas women may be motivated to understate theirs" (Alexander & Fisher, 2003, p. 28). These predictions were based on the prominent assumption that there is a strong sexual double standard in society: that sexual activity is socially rewarded for men but socially punished for women (Marks & Fraley, 2005).

Despite their plausibility and prominence, both the underlying assumption of markedly different male and female sexual standards and the potential consequence of gendered social desirability bias have found limited support. A recent meta-analysis revealed that high levels of sexual activity were evaluated more favorably for men than for women, but this double standard was inconsistent and rather small (Endendijk et al., 2020). It has been speculated that today, gendered evaluations may be the exception rather than the norm (Jonason, 2008; Jonason & Marks, 2009).

In a creative attempt to examine male and female self-presentation strategies, Alexander and Fisher (2003) compared (gender differences in) sexual self-reports between different groups of participants. In the exposure threat condition, participants were led to believe that the experimenter might see their responses, thereby stimulating socially desirable responding. In the bogus pipeline condition, participants were hooked up to an alleged polygraph, thereby encouraging honest responses. They found an interaction between gender and condition that was primary driven by women, who reported significantly more sexual events in the bogus pipeline condition. These results provide some support for gendered social desirability bias. However,

they were not consistently found across outcomes and studies (Fisher, 2013; Fisher & Brunell, 2014).

In our meta-analysis, we used bias indicators to estimate the extent to which measured gender differences are attributable to social desirability bias (Frankenbach et al., 2022). A common feature of these indicators is that for individuals who self-identify as heterosexual, true gender differences in the population should be (close to) zero because the type of activity being measured requires the simultaneous presence of male and female individuals (e.g., number of lifetime sexual partners). Thus, higher male than female values for these bias indicators could be interpreted cautiously as evidence that gender differences are partially biased. On average, we did indeed find higher male than female self-reports for these bias indicators. However, the gender differences were not consistent across the bias indicators and were substantially smaller than those for the indicators of sexual motivation. This suggests that social desirability bias can explain at most a small fraction of the gender differences in sexual motivation.

More dedicated research on sexual double standards and (gendered) social desirability bias is needed to better understand the validity of self-reported sexual motivation and gender differences therein. Investigating whether sexual self-reports are accurate in online surveys is particularly important because they have become the standard in recent years and offer anonymous conditions that may facilitate honest responses. However, most evidence for social desirability bias comes from laboratory studies that involve some degree of interpersonal interaction (Alexander & Fisher, 2003; Meston et al., 1998), which increases the likelihood of socially desirable responding.

The Present Research

A healthy and happy sex life, of which sexual motivation is a crucial feature, contributes to overall well-being. Social psychologists rely heavily on self-reporting to illuminate the

darkness of what goes on in people's minds and behind closed doors. However, there are good reasons to doubt what this flashlight seems to show. The present research addresses critical threats to the validity of sexual self-reports. In doing so, it contributes to a better understanding of controversial findings that are relevant to everyday life, such as gender differences in sexual motivation.

Conceptually, I assume a causal path from the construct level to the measurement level (Borsboom et al., 2004): differences in latent sexual motivation cause differences in self-reported sexual motivation. However, this relationship is not perfect. In addition to random noise, there are systematic influences that may weaken the validity of sexual self-reports and thus lead to spurious conclusions. First, many self-report scales lack a proper process of development and validation. This brings in the *defining threat* and the *delicate threat* to validity. Because of conceptual and empirical shortcomings, it is unclear (1) whether these scales truly measure sexual motivation and (2) whether they have the same meaning across different groups. Part I addresses these threats. It presents the theory-driven development and extensive empirical validation of a novel measure of sexual motivation: the Trait Sexual Motivation Scale (TSMS).

Second, self-presentation tendencies are another critical and *social threat* to validity. Most support for the actual relevance of social desirability bias comes from laboratory studies in which participants were placed within a larger group of individuals (e.g., Meston et al., 1998) or had to interact with an experimenter (e.g., Alexander & Fisher, 2003; Fisher, 2013). Moreover, while these studies found partial support for gender differences in socially desirability responding, not all of these effects were consistent with predictions of a strong sexual double standard (e.g., little evidence of male overreporting). Part II examines the accuracy of (gender differences in) sexual self-reports in online surveys. To get a complete picture, we combined several approaches to prevent, detect, and control for social desirability bias. In Part III, we addressed uncertainties

regarding the perceived societal norms likely to influence self-presentation tendencies. We proposed and tested a novel model of sexual (double) standards that equally emphasizes similarities and differences between male and female sexual norms. We reasoned that this S&D model could explain inconsistencies in previous research on gendered sexual norms and gendered social desirability bias. Table 1 provides an overview of the three manuscripts included in this thesis, which cover a total of eight studies.

Table 1

The Present Research: Overview of the Parts Included in This Thesis

Part	Title	Studies	Authors	Status (as to August 2024)
I	Development and Validation of the Trait Sexual Motivation Scale (TSMS)	4	Weber, M., Reis, D., & Friese, M.	Paper published in the *Journal of Personality Assessment* (2024)
II	How Valid Are Self-Reports of Sexual Motivation? Using the Item Sum Technique to Examine Self-Presentation Tendencies in Online Research	2(1)[a]	Weber, M., Kilger, H., & Friese, M.	Manuscript submitted for publication in *Motivation Science* (currently under review)
III	Sexual (Double) Standards Revisited: Similarities and Differences in the Societal Evaluation of Male and Female Sexuality	2(1)[a]	Weber, M. & Friese, M.	Manuscript accepted for publication in *Social Psychological and Personality Science* (in press)

Note. [a]Parts II and III each contained two studies. These studies were analyzed and presented together in the corresponding manuscripts.

The author of this dissertation is the first author on all three manuscripts. One manuscript has been published (Part I), one has been submitted for publication (Part II), and one has been

accepted for publication (Part III). This research was conducted in the spirit of openness and transparency, with open data and open materials (i.e., analysis scripts, questionnaires, and codebooks) provided for all studies. I used the AI-powered writing software DeepL Write (Kutylowski, 2024) for language checks (i.e., grammatical correctness, word choice) on the texts that I have written. I did *not* use any AI-powered software to create these texts or the ideas upon which they are based.

Part I: Development and Validation of the Trait Sexual Motivation Scale (TSMS)

The first part focuses on the role that the instrument used to measure sexual motivation plays in facilitating valid conclusions. An in-depth analysis of past research led to the conclusion that existing self-report measures of sexual motivation lack a theoretical conceptualization, a comprehensive validation process, or both, compromising their validity and widespread use. We therefore developed and comprehensively validated the Trait Sexual Motivation Scale (TSMS), featuring the substantive, structural, and external phases expressed in classic and recent guidelines for scale validation processes (e.g., Flake et al., 2017; Loevinger, 1957; Simms, 2008).

In the substantive phase, we generated a first version of the scale that is consistent with our theoretical conceptualization of sexual motivation (Frankenbach et al., 2022). In the structural phase, we empirically derived the final version of the TSMS (Study 1) and tested its internal consistency, stability, factorial validity, and measurement invariance for gender and relationship status (Study 2). Measurement invariance was critical to the present research because the degree to which a measure is invariant or equivalent across groups (e.g., similar psychometric properties for men and women) determines the appropriateness of mathematical operations such as calculating differences between groups (e.g., gender differences). In the external phase, we tested whether the TSMS correlates in expected ways with other sexual and non-sexual measures (i.e., nomological validity) and whether it predicts sexual criteria, both cross-sectionally (i.e.,

concurrent validity, Study 3) and prospectively in daily life (i.e., predictive validity, Study 4). In sum, Part I contributes to the valid measurement of sexual motivation by addressing the defining threat of measuring something other than the intended construct and the delicate threat of biased mean differences due to non-equivalent psychometric properties across groups.

Part II: How Valid Are Self-Reports of Sexual Motivation? Using the Item Sum Technique to Examine Self-Presentation Tendencies in Online Research

The second part focuses on social desirability bias in online research. Although this social threat to validity has a prominent place in sexuality research, little is known about its actual impact on gender differences in sexual motivation and sexual self-reports in online surveys. Unlike laboratory studies, online surveys eliminate the need for interaction with experimenters or other participants, potentially counteracting norm-consistent self-presentation tendencies. Our primary approach to examining the validity of online self-reports was to compare self-reported sexual motivation and gender differences therein between two groups. A first *direct questioning* (*DQ*) *group* simply answered sexual questions, resembling a standard online survey. A second *item sum* (*IS*) *group* answered the same sexual questions. However, instead of entering their true values, these participants were instructed to report the sum of their answers to one sexual and two non-sexual questions as a joint response. Because the non-sexual questions ask for information that only the participants have access to (e.g., the last two digits of a friend's phone number), there is no way to determine how much the sensitive and non-sensitive questions contributed to each participant's joint response. Under these conditions of maximum anonymity, honest self-reporting is particularly likely. Despite the random nature of each participant's response, mean level estimates for the sexual questions across the IS group can be calculated using knowledge of the population characteristics of the non-sensitive questions (e.g., phone number digits are approximately uniformly distributed, expected mean of the sum of two digits = 9).

To provide a more complete picture, this *indirect approach* to examining validity using the Item Sum Technique (Trappmann et al., 2014) was complemented by three additional approaches: (1) a *logic approach* that examines gender differences on questions for which substantial true gender differences are unlikely, (2) a *subjective approach* that considers self-rated levels of honesty, and (3) a *control approach* that calculates and controls for associations between self-reported sexual motivation and social desirability scales. Part II contributes to resolving uncertainties surrounding the valid measurement of sexual motivation in two important ways. First, it examines how social desirability bias affects sexual self-reports in online surveys. Second, it provides insight into the extent to which this social threat to validity may explain gender differences in sexual motivation.

Part III: Sexual (Double) Standards Revisited – Similarities and Differences in the Societal Evaluation of Male and Female Sexuality

The third part aims to promote the valid assessment of sexual self-reports by providing novel insights into the perceived societal norms that may give rise to socially desirable responding. Past predictions on biased sexual self-reports were built on the idea that sexual activity is socially punished for women but socially rewarded for men—a strong sexual double standard implying that women are likely to downplay and that men are likely to exaggerate sexual events (Alexander & Fisher, 2003; Marks & Fraley, 2005). However, empirical evidence in favor of the strong sexual double standard is mixed, at best, and some of its critical assumptions (e.g., the more favorable evaluation of low sexual activity levels for women than for equally inactive men) have rarely been tested.

Here, we present the S&D model, which posits that perceived societal norms (i.e., people's perceptions of how society values different levels of sexual activity) for men and women are equally characterized by marked similarities *and* differences, both of which are

necessary for an adequate characterization of sexual norms. We predicted that participants would perceive that high and low levels of sexual activity would be evaluated differently for men and women, but that the most favorable evaluations would occur for moderate levels of sexual activity. This new perspective on sexual (double) standards challenges the prominent notion that sexual norms and self-presentation tendencies are polar opposites for men and women. In this way, Part III holds important scientific implications for the valid measurement of sexual motivation.

PART I

Development and Validation of the Trait Sexual Motivation Scale (TSMS)

Marcel Weber

Saarland University

Dorota Reis

Saarland University

Malte Friese

Saarland University

Manuscript published in the *Journal of Personality Assessment* (2024).

This is an Accepted Manuscript of an article published by Taylor & Francis in the Journal of Personality Assessment on May 22, 2023, available at https://doi.org/10.1080/00223891.2023.2206896

Development and Validation of the Trait Sexual Motivation Scale (TSMS)

Marcel Weber[1], Dorota Reis[1], and Malte Friese[1]

[1]Department of Psychology, Saarland University

Author Note

Marcel Weber: https://orcid.org/0000-0002-0507-0070

Dorota Reis: https://orcid.org/0000-0002-0027-9713

Malte Friese: https://orcid.org/0000-0003-0055-513X

Corresponding Author

Correspondence concerning this article should be addressed to Marcel Weber, Department of Psychology, Saarland University, Campus A2 4, 66123 Saarbrücken, Germany. E-mail: marcel.weber@uni-saarland.de.

Abstract

Sexual motivation, the interest in sexual activity, affects people's thinking, feeling, and behavior. Common scales used to assess sexual motivation suffer from drawbacks that limit their validity and applicability. We therefore developed and validated the *Trait Sexual Motivation Scale* (TSMS), a brief, theory-driven self-report scale, over the course of four preregistered studies (N_{total} = 2,083). Results indicated good model fit, high internal consistency and stability of the second-order (i.e., trait sexual motivation) and first-order (i.e., cognition, affect, behavior) factor scores, and scalar measurement invariance for gender and relationship status. The TSMS correlated as expected with sexual and non-sexual constructs and predicted sexual outcomes cross-sectionally and prospectively in everyday life. Overall, the TSMS emerged as an economical, reliable, and valid measure of sexual motivation.

Keywords: sexual motivation, scale validation, nomological network

Development and Validation of the Trait Sexual Motivation Scale (TSMS)

Few spheres of life are as universally relevant as sexuality, and sexual motivation is a key aspect of people's sexuality. Typically understood as the intrinsic interest in sexual activity (Baumeister et al., 2001; Frankenbach et al., 2022; Stark et al., 2015), sexual motivation affects people's thinking, feeling, decision-making, and behavior. Sexual motivation can form, strengthen, and jeopardize both casual sexual encounters and long-term romantic relationships (Birnbaum, 2014; Birnbaum et al., 2019; Kim et al., 2021). In short, sexual motivation is highly relevant in almost every adult's life.

To elucidate the implications of sexual motivation for people's lives, how it differs or is similar across groups of people, how it relates to other sexual and non-sexual constructs, experiences, and behaviors, and how it may be affected by life circumstances, researchers need a valid measure of the construct. In the present research, we developed and validated the *Trait Sexual Motivation Scale* (TSMS), a brief, theory-driven self-report scale, over the course of four preregistered studies. Our goals were threefold and can be structured using Loevinger's (1957) seminal approach to construct validation. In the *substantive phase*, we relied on general literature on measurement and construct validation (Flake et al., 2017; Simms, 2008) to consider the specific needs of researchers interested in sexual motivation, concluded that a new instrument to assess sexual motivation is needed, and developed a first version of the TSMS. In the *structural phase*, we refined the scale and tested its properties. In the *external phase*, we specified the nomological network of (the) trait sexual motivation (scale) and tested the scale's ability to predict sexual outcomes and its incremental predictive value over key demographic variables and alternative measures of sexual motivation.

Measuring Sexual Motivation

As with any other trait, people differ in their typical level of sexual motivation (Baumeister et al., 2001; Frankenbach et al., 2022). The characteristics, antecedents, and

consequences of trait sexual motivation have attracted the interest of many disciplines, including social, personality and clinical psychology, medicine, sociology, and biology. Research from these fields has revealed the manifold implications of sexual motivation: Higher sexual motivation is associated with higher sexual and relationship satisfaction in couples (Kim et al., 2021). On average, men's sexual motivation is higher than women's (Baumeister et al., 2001; Frankenbach et al., 2022). Discrepancies in sexual motivation can pose a challenge for romantic relationships (Davies et al., 1999; Mark, 2012). Hormonal levels (e.g., estradiol, progesterone) predicted day-to-day fluctuations in sexual motivation in naturally cycling women (Roney & Simmons, 2013). Abnormally low or high sexual motivation can be detrimental to individuals and may therefore be clinically relevant (e.g., hyposexual desire disorder, Clayton et al., 2018).

In order to draw valid conclusions about sexual motivation and its relevance in daily life, good theorizing and measurement of sexual motivation are essential. Integrating work on measurement and construct validation in general (Flake et al., 2017; Loevinger, 1957) and the needs of sexual motivation research specifically, we see the following theoretical, empirical, and practical desiderata for instruments measuring sexual motivation.

First, before we measure, we should know what we want to capture. Simply put, what we can learn is limited if the definition and theoretical conceptualization of the construct remain vague. For sexual motivation, a precise definition is particularly important as the same construct appears under different names (e.g., sexual motivation, sex drive, libido; Spector et al., 1996) and different constructs appear under the same name (e.g., sex drive: intrinsic sexual motivation in general versus its biological component only; Baumeister et al., 2001; Levine, 2003). In addition, a clear theoretical conceptualization provides the basis for developing items that adequately capture the construct and helps to draw the line between valid indicators of sexual motivation and sexual criteria that may be related but are not part of the core construct (Frankenbach et al., 2022).

Second, validity cannot be taken for granted: Researchers need to make sure that instruments actually measure what they are supposed to measure: Are measurement models derived from theory supported by the data (i.e., factorial validity)? Are associations with various constructs consistent with theory and prior research (i.e., nomological validity)?

Third, from a practical perspective, measures of sexual motivation should be widely applicable and allow for the statistical operations that researchers using the scale are likely to consider important. Sexual motivation scales are regularly administered to and compared across different populations, most notably across gender and relationship status. However, some scales include items that refer to a "partner" that may be difficult for single people to answer (e.g., How strong is your desire to engage in sexual activity with a partner?," Spector et al., 1996). Other scales assess people's motivation to seek new sexual encounters that may be difficult to answer for people who are in a sexually exclusive relationship (e.g., "I am constantly looking for a new sex partner," Stark et al., 2015). As a consequence, researchers have changed the wording of items (Park & MacDonald, 2022) or decided to omit partner-related items altogether when studying single people (Vallejo-Medina et al., 2020). These changes may affect the measurement properties of the scale and make comparisons across studies difficult. In addition, a prerequisite for valid intergroup comparisons is that the measure has the same meaning across groups (e.g., men and women). The extent to which such *measurement invariance* (or measurement equivalence) holds in the data is pivotal information because it determines which statistical operations can be validly performed with the scale (Luong & Flake, 2022; Putnick & Bornstein, 2016). In turn, measurement invariance is tied to the validity of statistical conclusions: Without knowledge of measurement invariance across the groups of interest, researchers risk comparing "apples with oranges" (Greiff & Scherer, 2018) and seemingly robust group differences may be biased (Nye & Drasgow, 2011).

Fourth, a valid sexual motivation scale should predict relevant sexual outcomes such as sexual cognitions, feelings, and behaviors. A particularly strong demonstration of criterion-oriented validity would be the prospective prediction of sexual outcomes in people's daily lives.

Fifth, although subordinate to the other desiderata, we consider brevity to be another beneficial characteristic. Given that sexual motivation is often assessed as one of several constructs in large-scale online or experience-sampling studies, there is a growing need for an economical yet valid way of assessing sexual motivation.

Several measures of sexual motivation have been developed. The Sexual Desire Inventory (SDI, Spector et al., 1996), for instance, has contributed greatly to understanding the importance of (couple discrepancies in) sexual motivation for relationship and sexual satisfaction (Davies et al., 1999; Kim et al., 2021). (Variants of) the sex drive subscale of the Sexual Attitudes and Feelings Scale (SAF, Lippa, 2006) have been used in large-scale studies examining gender differences in sexual motivation across many nations and cultures (Lippa, 2009). The Sex Drive Questionnaire (SDQ, Ostovich & Sabini, 2004) has been used to examine the relationship between sexual motivation and sociosexuality (Simpson & Gangestad, 1991; for an overview of sexual motivation measures, see Stark et al., 2015). Although these and other measures have been instrumental in providing important insights into the nature of sexual motivation, none of them fully satisfies the key desiderata discussed in this section. Instead, all of these scales have either theoretical, empirical, and/or practical limitations. These limitations may compromise their construct validity, their widespread application, and their suitability for comparisons across groups and studies. We therefore concluded that researchers and practitioners would benefit from a new instrument that is based on a coherent theoretical conceptualization of sexual motivation and that has undergone an extensive validation process.

The Present Research

In the present research, we developed and validated the TSMS. This process followed the phases of construct validation introduced by Loevinger (1957), which will serve to structure the remainder of this article. In the *substantive* part of the validation process, we present a theoretical conceptualization of sexual motivation and generate a first version of the TSMS. In the *structural* part (Studies 1 & 2), we generate the final version of the scale and scrutinize its psychometric properties. In the *external* part (Studies 3 & 4), we test associations between the TSMS and other constructs and criteria (i.e., nomological and criterion validity).

Open Science Statement

We preregistered our research goals, hypotheses, and analytic strategies for all four studies prior to data collection. Exploratory non-preregistered analyses are transparently stated as such. Preregistration documents, data, scripts, and materials are openly available on the Open Science Framework (https://osf.io/ux9nk/). All studies were approved by the Institutional Review Board of Saarland University.

Part I: Substantive Phase

Theoretical Conceptualization of Trait Sexual Motivation

Profound theorizing about the conceptualization of the construct under investigation and its causal impact on test scores is an often overlooked but critical part of any validation process (Borsboom et al., 2004; Flake et al., 2017). The TSMS is based on a recent theoretical conceptualization that combines insights from trait theory with research on sexual motivation (Frankenbach et al., 2022). According to this conceptualization, trait sexual motivation manifests as relatively consistent patterns of thoughts, feelings, and behaviors, similar to other traits (McCrae & Costa, 2003; Roberts, 2009). Thus, people high in trait sexual motivation think about sex, desire sex, and have sex more often than people low in this trait (Frankenbach et al., 2022).

Sexual motivation also varies as a state within individuals. Even a person with a strong sexual motivation does not seek sexual pleasure all the time. The seeming conundrum between stable patterns of sexual thoughts, feelings, and behaviors on the one hand and strong intraindividual variability on the other hand is elegantly resolved by the idea of traits as density distributions of states (Fleeson, 2001; Fleeson & Jayawickreme, 2015; Roberts, 2009). Simply put, state sexual motivation varies over time as a function of various situational influences. However, over longer time spans (e.g., one week), the central tendency of the distribution of states is a reliable indicator of a person's trait sexual motivation, giving way to stable individual differences. Thus, measures can validly assess sexual motivation if they assess typical patterns in sexual motivation indicators over extended periods of time.

This integrated trait/state perspective fits well with the seminal sexual incentive motivation model, according to which sexual motivation requires the simultaneous presence of a sexually relevant stimulus (e.g., seeing or fantasizing about a potential partner) and an activated neural system (i.e., the central motive state; Ågmo & Laan, 2022b; Toates, 2009). The interplay of these two components, mediated by sexual arousal and sexual approach motivation, determines the occurrence of (partnered) sexual activity. Individuals high in sexual motivation may then be those who, on average, respond more readily to (a wider range of) sexually relevant stimuli (Ågmo & Laan, 2022a, 2022b). Previous work has used a large and heterogeneous variety of variables as indicators of sexual motivation (Baumeister et al., 2001). Without a clear rationale, it is difficult to determine which variables are valid indicators of sexual motivation and which may be related but distinct from the construct. The present conceptualization has clear implications for the measurement of sexual motivation. It specifies that the higher-order latent construct of sexual motivation manifests in the frequency of sexual cognitions (including thoughts, fantasies, or daydreams), sexual feelings (including desire or lust), and sexual behaviors (including solo masturbation or partnered sexual activity). These are the primary indicators of sexual motivation.

Scale Development

Our aim was to create a brief sexual motivation scale based on the theoretical conceptualization by Frankenbach and colleagues (2022). To this end, we sought to create items that are (a) easy to comprehend, (b) gender-neutral, and (c) representative of the three facets of sexual cognition, affect, and behavior. We developed a first eight-item version of the TSMS with three cognitive (e.g., "How often do you think about sex?"), three affective (e.g., "How often do you feel sexual desire?"), and two behavioral items (e.g., "How often are you sexually active [self-stimulation plus sex with another person?]"), all referring to frequencies "in a typical week" (see Table S1 in the supplementary online material [SOM] for the complete set of items).

These items allow individuals to report all cognitive, affective, and behavioral events, regardless of their origin and regardless of the person to whom they are directed or with whom they are performed. This is particularly salient for the two behavioral items that assess the sum of individual and dyadic sexual events (e.g., masturbation and activities performed with another person). These item wordings allowed for capturing the various ways in which latent sexual motivation can manifest in sexual behavior, independent of a person's preferences and situational circumstances (e.g., availability of a sex partner). Imagine person S who is single and person R who is in a romantic relationship. Assume that the frequency with which S and R think about sex, have sexual desires, and become sexually active is identical, but that solitary sexual activities (e.g., self-stimulation) are more common for S, whereas dyadic activities are more common for R. The sexual motivation of both persons would arguably be very similar, even though it expresses itself somewhat differently in terms of the behavioral facet due to their different life circumstances. Being agnostic towards the specifics of sexual events helps the TSMS operate similarly regardless of respondents' relationship status and gender[1].

[1]We value all gender identities. Because academic (and social) discussions about possible gendered expressions of sexual motivation usually contrast male and female sexuality, we follow previous research and focus on individuals who self-identify as male and female.

Part II: Structural Phase

Study 1: Scale Refinement

We designed a first preregistered study to transition from the initial item pool to a final version of the TSMS that is (1) easy to answer (for participants), (2) easy to process (for researchers), and (3) economical. Our specific aims were to empirically derive meaningful response categories and to shorten the scale to six items. A sample of N = 766 participants (49.9% female; 50.0% romantically involved; 78.1% heterosexual; age in years: M = 26.57, SD = 5.87, range: 18-41) recruited through Prolific.co completed the initial 8-item version of the TSMS using an open response format. As preregistered, the sample was randomly split into two subsamples. In the exploratory subsample, we used graphical and descriptive analyses to explore different ways of combining the open-ended responses into seven response categories. Separately for each item, we agreed on a winning solution of categorized data that approximated a normal distribution and facilitated a meaningful interpretation (e.g., 3-4x a week = every other day; 5-7x = up to once a day). These winning solutions were then tested in the independent confirmatory subsample. In both subsamples, the skewness and kurtosis of the categorized data were small (i.e., absolute values $\leq$ 1.26). Histograms and Q-Q plots showed no or marginal deviation from normally distributed data for all but two items (c3, b1). The proportion of participants answering zero was considerably higher for item b1 than item b2. We suspected that this was due to participants interpreting the word “plus” (item b1, see Table S1) as “having both events at the same time”. Therefore, we adjusted the item wording to remove this ambiguity (i.e., “How often do you pleasure either yourself or another person sexually? [Please provide the total of all events.]”). In sum, Study 1 provided the final six-item version of the TSMS with empirically derived response categories (see Table 1).

Table 1

Final 6-Item Version of the TSMS

Items	Response Categories
Sexual thoughts In a typical week: c1: How often do you think about sex? c2: How often do you have sexual fantasies?	1 = *0x a week*; 2 = *1-2x a week*; 3 = *3-4x a week (= every other day);* 4 = *5-7 times a week (= up to once a day)*; 5 = *8-14 times a week (= up to twice a day)*; 6 = *15-21 times a week (= up to three times a day)*; 7 = *more than 21 times a week (= more than three times a day).*
Sexual behaviors In a typical week: b1: How often do you pleasure either yourself or another person sexually? (Please provide the total of all events.) b2: How often do you either masturbate or have sex with someone else? (Please provide the total of all events.)	1 = *0x a week*; 2 = *1x a week*; 3 = *2x a week*; 4 = *3-4x a week (= every other day)*; 5 = *5-7 times a week (= up to once a day)*; 6 = *8-14 times a week (= up to twice a day)*; 7 = *more than 14 times a week (= more than twice a day)*.
Sexual feelings In a typical week: a1: How often do you feel sexual desire? a2: How often do you feel "turned on"?	1 = *0x a week*; 2 = *1-2x a week*; 3 = *3-4x a week (= every other day);* 4 = *5-7 times a week (= up to once a day)*; 5 = *8-14 times a week (= up to twice a day)*; 6 = *15-21 times a week (= up to three times a day)*; 7 = *more than 21 times a week (= more than three times a day).*

Note. Cognitive (c), affective (a), and behavioral (b) items. Item labels (e.g., c1) are intended to provide orientation for readers but were not presented to participants. Items c2, b1, and a1 form the Brief Trait Sexual Motivation Scale (BTSMS).

Study 2: Reliability, Factorial Validity, and Measurement Invariance

Study 2 was designed to provide first evidence of the reliability, factorial validity, and measurement invariance of the TSMS. Regarding reliability, we examined (1) the extent to which different items capture the same first- and second-order factor(s) (i.e., *internal consistency*), and (2) the extent to which differences in the (latent) trait remain stable over time (i.e., *stability* over four weeks and three months). Regarding factorial validity, the theoretical conceptualization of sexual motivation presented in Part I implies a measurement model with the second-order factor trait sexual motivation, the three first-order factors cognition, affect, and behavior, and their

respective indicators. We used confirmatory factor analysis (CFA) to examine the adequacy of this measurement model. Finally, we assessed the extent to which the psychometric properties of the TSMS are the same for men and women, and for singles and those in romantic relationships—that is, the measurement invariance of the scale across gender and relationship status.

Method

Participants and procedure. In total, $N = 665$ adult participants were recruited through Prolific.co. They agreed to complete the initial survey (T1) and two short follow-up surveys four weeks (T2) and three months (T3) later. After applying our preregistered exclusion criteria, a final sample of $N = 658$ participants remained for T1 (50.0% female; 50.3% romantically involved; 74.4% heterosexual; age in years: $M = 27.24$, $SD = 6.26$, range: 18-41), of whom 85.9% and 69.8% also completed the surveys at T2 and T3, respectively ($N_{T2} = 565$, $N_{T3} = 459$). The three surveys were compensated separately (T1: £0.50; T2 and T3: £0.25 each); overall, the hourly wage was £7.50 (i.e., U.S. $10.33 at the time the study was launched). Participants who completed all three surveys received a 10% bonus (i.e., £0.10). In each survey, participants provided consent and then filled out the final version of the TSMS. Next, they answered background questions (e.g., age, sexual orientation) and data-quality questions (e.g., self-rated data quality, anonymity; T1 only).

Preregistered analytic strategy.

Internal consistency and stability. First, we expected the first-order subscales (i.e., cognition, affect, behavior) to be internally consistent. Second, the global scale score should also be internally consistent—particularly when accounting for facet-specific differences. Third, the TSMS should measure sexual motivation as a relatively stable trait. Therefore, we expected that differences in TSMS scores would be stable over four weeks and three months, respectively. Reliability coefficients for internal consistency were the Spearman-Brown coefficient (ρ_{SB}) for the two-item subscales (Eisinga et al., 2013) and the partial coefficient omega ($\omega_{partial}$) as a measure of

the reliability of the second-order factor when controlling for facet-specific variance components. For stability, we preregistered the stability estimator (Röseler et al., 2020), which takes into account the internal consistency of a scale and is therefore recommended over the test-retest reliability coefficient (r_{tt}).

Factorial validity. To assess model fit, we primarily relied on the comparative fit index [CFI] and the standardized root mean square residual [SRMR], as recommended by Hu and Bentler (1999, see also Niemand & Mai, 2018). We further report the χ^2 test statistic and additional fit indices (i.e., Tucker-Lewis index [TLI], root mean square error of approximation [RMSEA], non-preregistered). For the preregistered fit indices, we applied both traditional fixed cutoffs (CFI ≥ .95 and SRMR ≤ .09; Hu & Bentler, 1999) and dynamic cutoffs that are tailored to the specific parameters of the scale and model under investigation (McNeish & Wolf, 2021; Niemand & Mai, 2018). We calculated the dynamic cutoffs using the web tool flexiblecutoffs.org (CFI ≥ .987 and SRMR ≤ 0.022; N = 658; df = 6, Niemand & Mai, 2018).[2] We used the R package *lavaan* (version 0.6-12; Rosseel, 2012) and applied effects coding (i.e., $M_{\text{loadings}} = 1$, $M_{\text{intercepts}} = 0$ at each level, see Little et al., 2006) to specify and test the proposed model.

Measurement invariance. To test whether the TSMS is measurement invariant across gender and relationship status, we used multigroup CFA (French & Finch, 2008; Xu & Tracey, 2017). This stepwise approach generates parallel measurement models for different groups (e.g., men and women) by specifying a series of nested models with increasingly strict restrictions. Following recommendations for higher-order models (Chen et al., 2005; Rudnev et al., 2018), we specified five nested models each for gender and relationship status (additional restrictions on top of those mentioned for previous models in parentheses): (M1) a configural model (no restrictions); (M2) a first-order metric model (equal first-order factor loadings across groups); (M3) a first- and

[2] Dynamic cutoffs are not yet available for higher-order models. We thus determined dynamic cutoffs for an isomorphic one-level model with correlated latent factors “Cognition”, “Affect”, and “Behavior”.

second-order metric model (equal second-order factor loadings across groups); (M4) a first-order scalar model (equal intercepts of measured variables across groups); (M5) a first- and second-order scalar model (equal intercepts of first-order latent factors across groups). The models were then compared sequentially in terms of meaningful changes in CFI and SRMR (metric stages: $\Delta CFI \geq .01$ and $\Delta SRMR \geq .03$, scalar stages: $\Delta CFI \geq .01$ and $\Delta SRMR \geq .03$; Chen, 2007) as well as McDonald's NCI ($\Delta MNCI \geq -.007$; Kang et al., 2016). A detailed overview of which levels of measurement variance allow which operations is beyond the scope of this paper and can be found elsewhere (e.g., Chen et al., 2005; Putnick & Bornstein, 2016). One critical level, however, is the scalar measurement invariance of the second-order factor trait sexual motivation (M5): This level allows for the comparison of mean scale scores across groups (e.g., gender differences in sexual motivation).

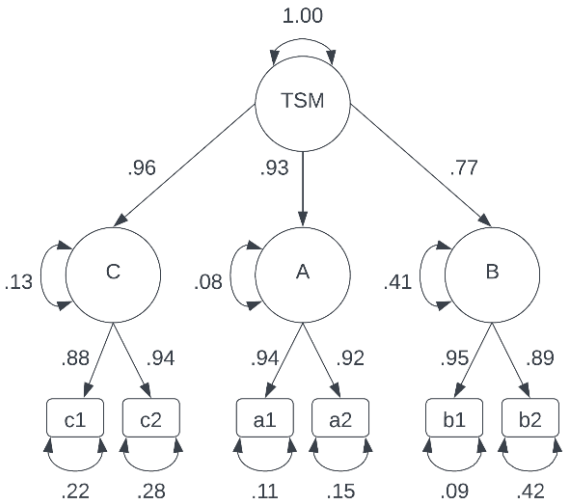
Results

Preliminary analyses, internal consistency, and stability. There were no missing values. For all items, skewness and kurtosis were small (i.e., absolute values < 1) and comparable to the results of Study 1. Internal consistency was high for all subscales ($\rho_{SB} \geq .86$) and for the total TSMS score when accounting for the first-order facets ($\omega_{partial} = .96$). In addition, TSMS scores were highly stable across four weeks (stability estimator = .92) and three months (stability estimator = .92).

Factorial validity. Figure 1 illustrates the variances and factor loadings of the proposed second-order model. Preregistered and non-preregistered fit indices in the total sample and in all subsamples indicated good model fit (Table 2). Non-preregistered exploratory analyses suggested that the proposed model described the data better than a simple one-factor model (see SOM, Table S2, for details).

Figure 1

Second-Order Confirmatory Factor Analysis of the Final Version of the TSMS



Note. Factor loadings and (residual) variances are standardized. C = Cognition; A = Affect; B = Behavior; TSM = Trait Sexual Motivation

Table 2

Second-Order Model: Model Fit in Total Sample and Subsamples

Sample	*N*	χ^2	*df*	*p*	CFI	SRMR	TLI	RMSEA
Total	658	32.24	6	< .001	.992	.010	.981	.082
Women	329	7.32	6	.292	.999	.009	.998	.026
Men	329	26.12	6	< .001	.987	.016	.967	.101
Single	327	17.61	6	< .001	.993	.008	.984	.077
In a relationship	331	31.67	6	< .001	.985	.016	.962	.114

Measurement invariance and group differences. For gender, successive comparisons of the first four models revealed no differences (Table 3). Setting equal intercepts of the first-order latent factors across groups (model 5a) caused a noticeable increase in MNCI. However, the changes in CFI and SRMR were marginal, and model 5a fit the data adequately (CFI = .982, SRMR = .057). For relationship status, none of the model comparisons revealed any marked difference, and the most restrictive model 5b fit the data adequately (CFI = .984, SRMR = .042). These results suggest that the TSMS is measurement invariant at the scalar level for both gender and relationship status.

Table 3

Measurement Invariance of the TSMS Across Gender and Relationship Status

Model	χ^2	*df*	CFI	SRMR	MNCI	$\Delta\chi^2$	Δdf	*p*	ΔCFI	ΔSRMR	ΔMNCI
Gender											
Model 1a	33.44	12	.993	.012	.984	–	–	–	–	–	–
Model 2a	35.40	15	.993	.017	.985	1.96	3	.581	.000	.005	.001
Model 3a	39.48	17	.993	.036	.983	4.08	2	.130	-.001	.019	-.002
Model 4a	49.37	20	.990	.039	.978	9.89	3	.020	-.002	.003	-.005
Model 5a	77.60	22	.982	.057	.959	28.23	2	< .001	-.009	.018	-.019
Relationship Status											
Model 1b	49.28	12	.989	.012	.972	–	–	–	–	–	–
Model 2b	54.65	15	.988	0.20	.970	5.37	3	.147	-.001	.007	-.002
Model 3b	63.12	17	.987	0.38	.966	8.47	2	.015	-.002	.019	-.004
Model 4b	75.26	20	.984	0.41	.959	12.15	3	.007	-.003	.002	-.007
Model 5b	78.32	22	.984	0.42	.958	3.06	2	.217	-.000	.001	-.001

Note. Model 1: configural models (no constraints); Model 2: first-order metric models (i.e., models 1 plus equal first-order factor loadings across groups); Model 3: first- and second-order metric models (i.e., models 2 plus equal second-order factor loadings across groups); Model 4: first-order scalar models (i.e., models 3 plus equal intercepts of measured variables across groups); Model 5: first- and second-order scalar models (i.e., models 4 plus equal intercepts of first-order latent factors across groups; Chen et al., 2005; Rudnev et al., 2018).

Measurement invariance at the scalar level allowed us to compare latent TSMS scores between men and women and between single and romantically involved persons. Replicating previous research (Baumeister et al., 2001; Frankenbach et al., 2022; Lippa, 2009), average TSMS scores were higher for men than for women ($z = 10.57$, $p < .001$, *standardized mean difference* = 0.86). We found no evidence that TSMS scores differed between those in romantic relationships and those who were single ($z = 0.83$, $p = .407$, *standardized mean difference* = 0.07)[3].

Discussion

Study 2 revealed (1) internally consistent factor scores, (2) high relative stability of trait sexual motivation (scores) after periods of four weeks and three months, (3) an adequate fit of the proposed second-order model, and (4) scalar measurement invariance for gender and relationship status. Thus, the TSMS enables *comparing* trait sexual motivation *between* people of different genders and relationship statuses. Consistent with previous research (Baumeister et al., 2001; Frankenbach et al., 2022), TSMS scores indicated higher sexual motivation in men compared to women, providing initial evidence of convergent validity at the group level.

Part III: External Phase

Study 3: Nomological Validity, Criterion Validity, and Incremental Validity

The aims of Study 3 were twofold. First, we aimed to map the nomological network of sexual motivation as measured by the TSMS by examining associations with other sexual and non-sexual constructs. Second, we tested the extent to which the TSMS predicts sexual criteria (e.g., pornography use, time spent with sexuality) in isolation (i.e., criterion validity) and over and above gender, age, and an alternative measure of sexual motivation (i.e.,

[3] For reference, group differences based on manifest scores were $t(656) = 11.57$, $p < .001$, $d = 0.90$, for gender, and $t(656) = 0.94$, $p = .346$, $d = 0.07$ for relationship status.

incremental validity). We preregistered ranges of expected values for the nomological associations and minimum expected associations for the associations with sexual criteria.

Method

Participants, power, and procedure. The recruitment strategy and exclusion criteria were the same as those described in Study 1. Responses from N = 461 participants were collected through Prolific.co. The final sample consisted of N = 450 participants (51.1% female; 50.2% romantically involved; 69.8% heterosexual; age in years: M = 27.10, SD = 6.12, range: 18-40). An effect size sensitivity analysis revealed that this sample size provides 80% power to detect small associations of r = .13 and 90% power to detect small to medium associations of r = .15 in the long run. Participants were paid £1.60, equivalent to an hourly wage of £8.00 (i.e., U.S. $11.28 at the time the study was launched). After giving informed consent, participants answered the TSMS, questions about sexual outcomes, other sexual and non-sexual constructs, and background and data quality questions.

Measures and preregistered analytic strategy. We preregistered all measures, the expected dimensionality of all multi-item measures, and the expected associations with all nomological and criterion measures. We report manifest associations between the TSMS and the nomological and criterion measures, respectively[4]. Qualitative descriptions of associations (e.g., "very small") follow the benchmarks suggested by Funder and Ozer (2019).

Nomological measures. The included measures were expected to cover the continuum from very low (i.e., discriminant) to very high (i.e., convergent) associations with the TSMS. We expected very high correlations with alternative measures of sexual

[4] We had pre-registered latent analyses using structural equal modeling, but one model showed insufficient fit. All other models fitted the data well and conclusions were identical to those drawn based on the manifest models.

motivation ($r > .70$)[5], moderate to very large negative associations with sexual restraint ($-.40 \leq r \leq -.20$), small to large positive associations with sociosexuality ($.10 \leq r \leq .30$), and small to moderate negative associations with self-control ($-.20 \leq r \leq .00$). In addition, we expected small to moderate associations with the Big Five personality dimensions ($-.20 \leq r \leq .20$), which were likely to be positive for openness and extraversion, and negative for conscientiousness and agreeableness.

Sexual Attitudes and Feelings Scale: Sex Drive (SAF). Participants completed the 5-item SAF (Lippa, 2006) as an alternative measure of sexual motivation (e.g., "I have a high sex drive," $\omega = .84$).

Sexual Desire Inventory (SDI). The SDI (Spector et al., 1996) assesses sexual desire with 14 items (e.g., "When you first see an attractive person, how strong is your sexual desire?"). Recent evidence suggests good fit of a model with three correlated factors ("solitary," "attractive-person based," and "partnered," see Mark et al., 2018). However, some researchers have raised concerns about the appropriateness of this model for single persons (Vallejo-Medina et al., 2020). Furthermore, it is common to calculate an overall SDI score (e.g., Jones et al., 2018), a practice that is consistent with our idea of a higher-order sexual motivation, but is not reflected in the model. We therefore preregistered adding a second-order factor (i.e., total sexual desire) and limiting confirmatory analyses to people in a relationship. Internal consistency was high (total sexual desire: $\omega_{partial} = .88$; subscales: $\rho_{SB\ attractive\ person} = .89$, $\omega_{solitary} = .89$, $\omega_{partnered} = .89$).

[5] To demonstrate convergent and incremental validity, we used alternative measures of sexual motivation based on a theoretical conceptualization closely related to the one proposed here. If one were to distinguish between excitatory and inhibitory processes as proposed by the dual control model of sexual behavior, variants of the SES/SIS scales would have been natural candidates (Carpenter, et al., 2011; Janssen et al., 2002).

Sexual Restraint Scale (SRS). People differ in their sexual restraint—that is, in how much they resist (versus give in to) sexual urges. Six items from the SRS (Gailliot & Baumeister, 2007) were used to assess this construct (e.g., "I am very good at controlling my sexual urges," ω = .87).

Sociosexual Orientation Inventory (SOI-R). People with unrestricted sociosexuality positively evaluate, desire, and/or engage in uncommitted sexual relationships (Penke & Asendorpf, 2008; Simpson & Gangestad, 1991). The SOI-R consists of three subscales capturing sociosexual attitudes, desire, and behavior. Participants answered the three items that form the attitudinal subscale (e.g., "Sex without love is OK," ω = .79; Penke & Asendorpf, 2008)[6].

Brief Multidimensional Self-Control Scale (BMSCS). Self-control refers to the ability to control dominant responses, including thoughts, emotions, and behavioral impulses, and to avoid conflicts between dominant responses and long-term goals. We used the 8-item BMSCS to assess self-control (e.g., "I focus daily on my long-term goals," ω = 81; Nilsen et al., 2020).

Big Five Inventory 2 – Short Version (BFI-2-S). Previous research suggests small to medium associations between sexual motivation and Big Five personality traits, which were most pronounced for openness (positive relation), extraversion (positive), and conscientiousness (negative; Allen & Walter, 2018). We used the 30-item BFI-2-S (Soto &

[6] We assessed only sociosexual attitudes because the validity and interpretation of items measuring sociosexual desire (e.g., "How often do you have fantasies about having sex with someone with whom you do *not* have a committed romantic relationship?") and behavior (e.g., With how many different partners have you had sex within the past 12 months?") are likely to depend strongly on a person's relationship status (Lippa, 2009). Although sociosexual desire is thought to be a specific form of general sexual desire (Penke & Asendorpf, 2008), it is impossible to empirically disentangle these concepts for single people, whose desire by definition cannot refer to a relationship partner. For those in sexually exclusive relationships, behavioral items are likely to be of limited information because, unlike singles, having multiple sex partners requires sexual infidelity.

John, 2017), which assesses three facets of each Big Five dimension with two items each (.78 $\leq \omega_{partial} \leq$ .87 for the five dimensions).

Criterion validity measures. The criteria were four types of sexual experiences and behaviors that we expected to be associated with trait sexual motivation—(1) pornography use, (2) orgasm frequency, (3) time spent with sexuality, and (4) age at first masturbation. We used or adapted face-valid one-item measures from previous research (see Table 4). We preregistered pornography use, orgasm frequency, and time spent with sexuality as primary outcomes. Age at first masturbation served as a secondary outcome for which we were less certain about obtaining an association with trait sexual motivation. *If* we found a negative correlation (i.e., higher sexual motivation linked with first masturbation earlier in life), this would strongly corroborate the proposed trait understanding of sexual motivation, because current levels of sexual motivation would then be linked to a sexual milestone that (in many cases) took place many years ago.

For each criterion, we tested the predictive value of the TSMS in isolation (i.e., *bivariate model*), over and above gender and age, which have been identified as important predictors of sexual events (i.e., *incremental model*), and compared the incremental values of the TSMS and the SAF as an alternative measure of sexual motivation[7] (i.e., *comparative model*). For bivariate associations, we expected very large positive associations with the primary outcomes (*r*s > .40) and a smaller negative association with age at first masturbation ($r <$ -.20).

[7] We did not preregister comparing the TSMS with the SDI due to concerns about the SDI's validity for singles (Vallejo-Medina et al., 2020).

Results

Preliminary analyses: Reliability, factorial validity, and gender differences (non-preregistered replication). There were no missing values, and skewness and kurtosis were again small (i.e., absolute values ≤ 1.02). In a series of non-preregistered analyses, we replicated the high internal consistency ($\omega_{partial}$ = .96; $\rho_{SB\ cognition}$ = .86; $\rho_{SB\ affect}$ = .92; $\rho_{SB\ behavior}$ = .92) and factorial validity (CFI = .992, SRMR = .011) of the TSMS. In addition, we again found higher sexual motivation in men than in women (*standardized mean difference* = 0.77).

Nomological validity. Table 5 shows the associations between the TSMS and the nomological measures. Almost all of the associations were as expected. TSMS scores and the alternative measures of sexual motivation and sexual desire, respectively, were highly correlated (SAF: r = .71; SDI: r = .66), suggesting strong convergent validity. Associations between the TSMS and other sexual constructs were consistent with our predictions in terms of direction and magnitude (SRS [sexual restraint]: r = -.25; SOI [sociosexuality]: r = .22). Also as expected, the associations with the non-sexual measures (i.e., Big Five, self-control) were small to moderate (-.13 ≤ rs ≤ .05), suggesting strong discriminant validity. Figure 2 illustrates the nomological network of sexual motivation, furthering our understanding of the construct.

Criterion validity and incremental validity. Bivariate models revealed significant positive associations between the TSMS and the primary criteria of pornography use, orgasm frequency, and time spent with sexuality (0.45 ≤ β ≤ 0.60, ps < .001, .198 ≤ R^2 ≤ .445), and with the secondary criterion of age at first masturbation (β = -0.24, p < .001, R^2 = .056).

Table 4

Criterion Validity: Constructs, Instruments, Items, Reliability, and Preregistered Predictions

Criterion	Item wording	Response options	Prediction
Pornography use	"During the past year, how often did you view pornographic material (such as internet sites, magazines, or movies)?" ("New Family Structures Study", documented in Regnerus et al., 2016)	1 = never, 2 = once a month or less; 3 = 2-3 days a month; 4 = 1-2 days a week; 5 = 3-5 days a week; 6 = (almost) every day	$r > .40$
Orgasm frequency	"During the last year, how many orgasms did you have in a typical week? It does not matter how the orgasm was achieved (e.g., masturbation, sexual encounters, wet dreams)." (Klein et al., 2015)	Open response format (__ orgasms a week)	$r > .40$
Time spent with sexuality	"Please think of a typical day in the last year: Please estimate the amount of time you spent with sexual fantasies, sexual urges, and sexual behavior." (adapted from Klein et al., 2015)	1 = *not at all;* 11 = *more than 3 hours*	$r > .40$
Age at first masturbation	"At what age did you masturbate for the first time?" (adapted from Pinkerton et al., 2003)	Open response format (At the age of __ years)	$r < -.20$

Table 5

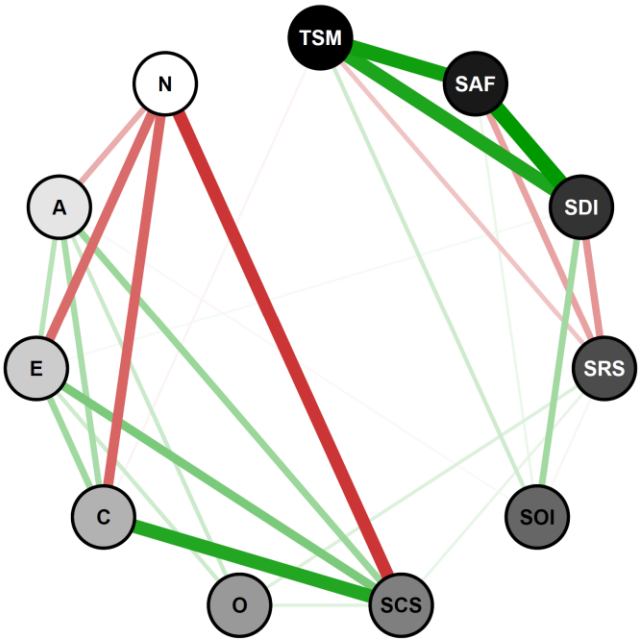
Nomological Validity of the TSMS: Constructs, Instruments, Preregistered Predictions, and Results

Constructs	Instruments	$r_{expected}$	$r_{observed}$	*t*	*df*	*p*
Sexual motivation	Sexual Attitudes and Feelings Scale, Subscale "Sex Drive" (SAF; Lippa, 2006)	$r \geq .70$	.71	21.09	448	< .001
Sexual desire	Sexual Desire Inventory 2, "Total Sexual Desire" (SDI; Spector et al., 1996)	$r \geq .70$	.66	13.26	224	< .001
Sexual restraint	Sexual Restraint Scale (SRS; Gailliot & Baumeister, 2007)	$-.40 \leq r \leq -.20$	-.25	-5.46	448	< .001
Sociosexuality	Revised Sociosexual Orientation Inventory (SOI-R, Subscale "Attitudes"; Penke & Asendorpf, 2008)	$.10 \leq r \leq .30$	.22	4.89	448	< .001
Self-control	Brief Multidimensional Self-Control Scale (BMSCS; Nilsen et al., 2020)	$-.20 \leq r \leq .00$	-.05	-1.02	448	.308
Big Five	Big Five Inventory, short version (BFI-2-S; Soto & John, 2017)					
O: Openness		$.00 \leq r \leq .20$	.03	0.73	448	.468
C: Conscientiousness		$-.20 \leq r \leq .00$	-.13	-2.78	448	.006
E: Extraversion		$.00 \leq r \leq .20$	-.03	-0.70	448	.485
A: Agreeableness		$-.20 \leq r \leq .00$	-.07	-1.50	448	.135
N: Neuroticism		$-.20 \leq r \leq .20$	.02	0.43	448	.668

Note. t, df, and *p* refer to two-tailed tests of correlations with the TSMS against zero (i.e., no correlation).

Figure 2.

Nomological Network of Trait Sexual Motivation as Measured by the TSMS



Note. Associations between the TSM(S) and measures of sexual motivation (SAF), sexual desire (SDI), sexual restraint (SRS), sociosexuality (SOI), self-control ([BM]SCS), and the BFI-2-S subscales openness (O), conscientiousness (C), extraversion (E), agreeableness (A), and neuroticism (N). Thicker lines represent stronger positive (*green*) or negative (*red*) associations; small associations ($|r| < .1$) are hidden for clarity.

Table 6 summarizes the results of the incremental validity analyses. Incremental models indicated that both measures of sexual motivation explained additional variance beyond gender and age in all outcomes. The comparative model revealed that incremental effects of the SAF beyond gender, age, and the TSMS were found for time spent with sexuality ($\Delta R^2 = .025$), but not for any other criterion ($\Delta R^2 \leq .002$). In contrast, the TSMS had incremental effects beyond gender, age, and the SAF on all four outcomes ($.043 \leq \Delta R^2 \leq .129$). Also, all criteria were more strongly associated with the TSMS than with the SAF.

Table 6

Criterion Validity of the TSMS

	Pornography use			Orgasm frequency			Time spent with sexuality			First masturbation (age)		
Model	β	R^2	ΔR^2	β	R^2	ΔR^2	β	R^2	ΔR^2	β	R^2	ΔR^2
M1. Covariate Model		.285***			.075***			.028**			.013	
Gender	1.04***			0.54***			0.33***			-0.11		
Age	-0.17***			0.01			-0.06			0.11*		
M2a. Incremental Model: TSMS		.482***	.197a***		.328***	.253a***		.203***	.175a***		.071***	.058a***
Gender	0.68***			0.13			0.02			0.10		
Age	-0.19***			-0.01			-0.07			0.11*		
TSMS	0.48***			0.54***			0.45***			-0.26***		
M2b. Incremental Model: SAF		.361***	.076a***		.199***	.124a***		.185***	.157a***		.029**	.016a**
Gender	0.85***			0.31***			0.06			-0.02		
Age	-0.17***			0.02			-0.05			0.10*		
SAF	0.29***			0.37***			0.42***			-0.13**		
M3. Comparative Model		.483***	.001b/.122c***		.328***	.000b/.129c***		.228***	.025b***/.043c***		.073***	.002b/.044 c***
Gender	0.68***			0.13			0.04			0.09		
Age	-0.19***			-0.01			-0.06			0.11*		
SAF	-0.04			0.03			0.22***			0.06		
TSMS	0.51***			0.52***			0.30***			-0.30***		

* $p < .05$; ** $p < .01$; *** $p < .001$. [a]compared to M1; [b]compared to M2a; [c]compared to M2b. To present easily interpretable values, we standardized all continuous variables and left gender in its original metric (e.g., average change in SD units of pornography use if a person is male rather than female).

Discussion

Replicating the results of Study 2, the TSMS was a highly reliable measure of sexual motivation with excellent model fit. Associations between the TSMS and alternative measures of sexual motivation or sexual desire were very high, indicating convergent validity. Associations with sexual restraint, sociosexuality, personality dimensions, and self-control were small to moderate and consistent with previous findings in terms of direction and magnitude, indicating discriminant validity. These associations provide new insights into the nomological network of sexual motivation.

We further found that the TSMS predicted pornography use, orgasm frequency, how much time participants spent with sexuality, and age of first masturbation. The latter finding is particularly noteworthy because it refers to a milestone in sexual development that occurred, on average, more than a decade earlier. These associations remained similar when controlling for gender and age, indicating incremental validity over these demographic variables. Finally, the TSMS showed superior predictive value compared to the SAF.

Study 4: Predictive Validity for Sexual Events in Everyday Life

Study 3 established the criterion validity of the TSMS by demonstrating its ability to predict the frequency of typical sexual behaviors. One potential criticism of Study 3 is that the predictors and criteria were assessed in the same session and required similar cognitive strategies (e.g., recalling and aggregating sexual behaviors over longer periods of time). Thus, it is possible that the associations were partially shaped by shared method variance. In addition, in Study 3, the TSMS "predicted" typical frequencies of sexual behaviors in the past. Study 4 therefore used experience sampling to test whether the TSMS predicts the frequency of future sexual events that are directly indicative of sexual motivation (i.e., sexual cognition, affect, and behavior), criterion outcomes (i.e., pornography use, time spent with

sexuality), and further sexual experiences (i.e., sexual excitability, self-rated sexual motivation) in everyday life.

Method

Participants, power, and procedure. A total of *N* = 241 participants volunteered to take part in a 14-day experience-sampling study. Participants were again recruited through Prolific.co. The study consisted of three phases, the first two of which were relevant to this project (see preregistration). First, participants completed an intake survey in which they answered the TSMS as well as background and data quality questions (see Study 2). Second, *N* = 213 participants who met the preregistered inclusion criteria (see Study 2) entered a 14-day experience-sampling phase. Three mobile survey invitations per day (i.e., 42 in total, approximately 1 minute each) were sent through the Prolific system at 10 a.m., 3 p.m., and 8 p.m., and could be accessed within 60 minutes after receipt.[8] In each mobile survey, participants reported on sexual experiences and behaviors since receiving the previous signal. Participants who completed at least one mobile survey formed the final sample *(N* = 209; 50.2% female; 49.8% romantically involved; 70.8% heterosexual; age in years: *M* = 27.02, *SD* = 6.23, range: 18-40). They completed a total of *k* = 4,973 mobile surveys (23.8 on average per person; 56.7% of all mobile surveys). Sensitivity analyses calculated based on Arend and Schäfer (2019) revealed that in the long run, this would give us an 80% chance of detecting moderate associations (β = .21) between the TSMS and sexual experiences and behaviors. Including bonuses that depended on the number of surveys completed, participants could earn up to £11.00 (i.e., U.S. $14.60 at the time the study was launched; intake: £1.00; mobile surveys: £0.15 each; follow-up: £0.70; bonus: up to £3.00).

[8] For technical reasons, some response periods exceeded 60 minutes. Rerunning the analyses based on responses collected within 60 minutes only did not alter any of the conclusions.

Measures and preregistered analytic strategy. During the intake session, participants completed the TSMS. In each of the mobile surveys, they reported on sexual events and downstream criteria in daily life. Face-valid one-item measures were used to maximize clarity and minimize attrition. First, participants answered four items assessing the frequencies of cognitive, affective, and solitary and dyadic behavioral sexual events since the last signal ("How many sexual thoughts and fantasies did you have?," "How often did you feel sexual desire or 'turned on'?," "How often did you masturbate or pleasure yourself sexually?," "How often did you have sex with another person?"). Participants then completed four items assessing the criterion outcomes and further sexual experiences. Pornography use was assessed in a dichotomous format ("Have you used sexually exciting or pornographic material [such as Internet sites, magazines, or movies]?," 1 = *no*, 2 = *yes*). Time spent with sexuality, self-rated sexual motivation, and sexual excitability ("Since receiving the last signal, [I spent a considerable amount of time with sexuality (e.g., fantasies, desire, activities, pornography)/I had a strong sex drive/it did not take much to get me sexually excited]") were assessed using 7-point scales (1 = *strongly disagree* to 7 = *strongly agree*).

To test the predictive validity of the TSMS, we used the R package *lme4* (version 1.1-30; Bates et al., 2014) to run (generalized) linear mixed models ([G]LMM) with observations in everyday life (level 1) nested within participants (level 2). TSMS scores were the manifest means of the six TSMS items. We used standard LMMs for normally distributed outcomes, GLMMs with a binomial distribution for binary outcomes, and GLMMs with a Poisson probability distribution for count outcomes.

Results and Discussion

The results are summarized in Table 7. We found strong positive associations between the TSMS and all seven outcomes (all *p*s < .001). Non-preregistered exploratory analyses further suggested that the TSMS was still significantly associated with all outcomes

after gender and age were added to the models (all *p*s < .001). Taken together, the TSMS prospectively predicted sexual experiences and behavior in everyday life and showed incremental effects above and beyond important demographic predictors.

Table 7

Predictive Validity of the TSMS: Associations with Sexual Outcomes in Everyday Life

Criterion	*B*	*SE*	$CI_{B\ 95\%}$	*z*	*p*	Effect
Event Frequencies						
Cognition	0.50	0.04	[0.42, 0.57]	13.41	< .001	1.64[a]
Affect	0.51	0.04	[0.44, 0.59]	13.27	< .001	1.67[a]
Behavior	0.55	0.06	[0.44, 0.67]	9.58	< .001	1.74[b]
Criterion Outcomes						
Pornography use	0.54	0.08	[0.39, 0.69]	6.98	< .001	1.71[b]
Time spent with sexuality	0.42	0.04	[0.34, 0.51]	9.63	< .001	0.15[c]
Further Sexual Experiences						
Self-rated sex drive	0.63	0.05	[0.53, 0.73]	12.80	< .001	0.24[c]
Sexual excitability	0.58	0.05	[0.49, 0.68]	12.21	< .001	0.21[c]

Note. Intercepts are not displayed for the sake of clarity. Behavior (0 = *no*, 1 = *yes*) contains information from solitary and dyadic events (preregistered) to prevent zero inflation. Effect sizes are [a]incidence rate ratios for event frequencies (IRR values greater than 1 indicate a positive association between TSMS scores and event frequencies), [b]odds ratios for binary outcomes (positive association: OR > 1), and [c]$R^2_{marginal}$ (i.e., proportion of the total variance explained by the fixed effect; Nakagawa et al., 2017) for continuous outcomes (positive association: $R^2_{marginal} > 0$).

General Discussion

Sexual motivation is a central personality characteristic that shapes people's sexual experiences and behavior in both solitary and social contexts. Existing measures of sexual motivation leave open questions regarding their underlying theoretical conceptualization of

sexual motivation, their (lack of) validation process, and their measurement invariance across gender and relationship status. We therefore developed and validated the theory-driven TSMS across four preregistered studies, following Loevinger's (1957) seminal structure for construct validation.

Substantive Phase: Theory-Driven Scale Development

We derived the items and factor structure of the TSMS from a theoretical conceptualization positing that the sexual motivation trait is a relatively stable and latent construct that manifests in cognitive, affective, and behavioral events whose frequencies can be used to measure the trait (Frankenbach et al., 2022).

Structural Phase: Categories, Factorial Validity, and Measurement Invariance

We empirically derived response categories for the TSMS (Study 1). All items were approximately normally distributed. CFA revealed excellent fit of the proposed measurement model (Studies 2 and 3). Multigroup CFA further revealed scalar measurement invariance allowing for intergroup comparisons regarding gender and relationship status. Replicating previous findings, these revealed a stronger sexual motivation in men compared to women.

External Phase: Nomological Associations and Criterion-Related Validity

The TSMS correlated highly with alternative measures of sexual motivation (SAF) and sexual desire (SDI), indicating strong convergent validity. Medium associations with sexual restraint and sociosexual attitudes and small associations with non-sexual constructs, including self-control and the Big Five, provided evidence for discriminant validity. Overall, the observed associations were consistent with our predictions, therefore indicating strong nomological validity. In addition, the TSMS predicted concurrently measured sexual outcomes (e.g., orgasm frequency, age at first masturbation), had incremental value over gender, age, and the SAF, and prospectively predicted the frequency of sexual events (i.e., sexual fantasies/desires/activities), sexual criterion measures (e.g., pornography use), and

other sexual experiences (e.g., sexual excitability) in everyday life. These results support the high practical relevance of the TSMS.

Developing an Ultra-Short Scale: The Brief Trait Sexual Motivation Scale (BTSMS)

With its six items, the TSMS is an efficient measure of sexual motivation. For use in studies in which each item is costly (e.g., panel studies, experience-sampling studies), we reasoned that an even briefer scale with decent psychometric properties would be welcome. We therefore developed the *Brief Trait Sexual Motivation Scale* (BTSMS) by combining the three items with the highest first-order factor loadings in Study 2 (i.e., c2: "sexual fantasies"; a1: "sexual desire"; b1: "pleasuring oneself/another person"). Repeating key analyses revealed that the BTSMS is internally consistent ($\omega = .85$), highly stable over four weeks and three months (stability estimator $\geq .98$), and replicates known gender differences (*standardized mean difference* = 0.88, Study 2). In addition, just like the six-item TSMS, the BTSMS showed meaningful nomological associations and significantly predicted all sexual outcomes cross-sectionally (Study 3) and prospectively (Study 4). For details, please consult the SOM, Tables S3-S5.

Theoretical and Practical Implications

Sexual motivation often plays an important role in research on romantic relationships and sexuality. To our knowledge, the TSMS is the first scale that has been constructed to apply equally to participants who are female, male, single or in a romantic relationship (Stark et al., 2015; Vallejo-Medina et al., 2020) and has also been empirically shown to be measurement invariant at the scalar level for gender and relationship status, allowing for mean-level comparisons across these groups. These excellent psychometric properties may be useful for researchers interested in gender differences (e.g., Baumeister et al., 2001; Frankenbach et al., 2022) or sexual desire discrepancies in couples. A vibrant literature seeks to answer the question of whether and under what circumstances differences in sexual

motivation between partners in a romantic relationship may impact sexual and relationship satisfaction (e.g., Kim et al., 2020; Mark, 2012). In research involving heterosexual couples, access to a scale that ensures the valid interpretation of empirical mean differences between genders is of particular value. Without evidence of measurement invariance, it is unknown whether and to what extent empirical within-couple differences in sexual motivation reflect actual differences on the construct level (Sakaluk et al., 2021).

On the theoretical level, previous research has debated whether the terms sexual motivation and sexual desire refer to the same or different constructs (e.g., Spector et al., 1996; Stark et al., 2015). Very high correlations between dedicated measures of sexual motivation (TSMS, SAF) on the one hand and a dedicated measure of sexual desire (SDI) on the other hand suggest that on the empirical level, these scales appear to measure the same construct. This suggests that the field may suffer from a jangle fallacy, in which different terms falsely suggest that they refer to different constructs, when in fact they refer to the same (Gonzalez et al., 2021). Note that in the theoretical conceptualization that guided the present research (Frankenbach et al., 2022), sexual affect (including desire) represents one of three facets of the overarching construct of sexual motivation. Other work that takes sexual desire as the overarching construct focusses on this affective facet and does not, or less prominently, include cognitive and behavioral facets (Birnbaum, 2018). Future work would benefit from more clearly delineating these constructs (or declaring them synonymous).

Strength, Limitations, and Future Research

One strength of the present research is its systematic orientation toward common standards of scale development and construct validation (Flake et al., 2017; Loevinger, 1957). Throughout this process, we relied on a theoretically grounded conceptualization of sexual motivation and used state-of-the-art methodology. For example, we used separate samples to empirically derive and test the response categories to avoid overfitting the

solutions to one specific dataset (Study 1), implemented advances in study-specific decision criteria (i.e., dynamic cutoffs; Niemand & Mai, 2018) when assessing model fit (Studies 2-3), and tested predictive validity within an intensive-longitudinal experience sampling design (Study 4). Another strength is that we adhered to open science practices by preregistering all studies, including exclusion criteria, fit indices, cutoffs, and decision criteria for model fit analyses, as well as expectations regarding nomological and criterion associations. All preregistrations, materials, and data are openly available on the OSF.

Notwithstanding these strengths, some limitations also warrant mention. First, all samples were recruited through online crowdsourcing platforms. Past research suggests that online samples are demographically heterogeneous (Goodman et al., 2013) and that Prolific.co samples provide high-quality data (Peer et al., 2021). In our studies, equal numbers of male and female, and single and romantically involved participants speak to at least some heterogeneity. Few failed attention checks and theoretically meaningful convergent and divergent associations further indicate that responses were valid. Thus, we are confident that the quality of our data is high, but future work will provide more conclusive evidence about the validity of the TSMS beyond online samples. Second, all participants were US residents. As sexuality is influenced by societal norms, this may also affect sexual motivation, its manifestations, and associations with related concepts. Preliminary findings from our lab suggest that the present results may generalize to at least some other Western societies, but more dedicated work examining cross-cultural similarities and differences is needed. In particular, future research should test whether the TSMS is measurement invariant across different cultures and languages. Third, because all measures were self-reported, associations between the TSMS and sexual criteria may be inflated by common-method bias (Podsakoff et al., 2012). In general, third-party reports can be a remedy to this problem. However, third-party reports are problematic when the issue is based on

subjective perceptions or behavior that is difficult to observe (Brannick et al., 2010), both of which apply to sexual motivation. Therefore, we refrained from third-party reports, but followed recommendations to minimize common method bias by collecting predictors and criteria with different response formats (Podsakoff et al., 2012). Finally, our agenda was guided by Loevinger's (1957) perspective on validity, as expressed in recent recommendations for scale validation processes (Flake et al., 2017). An alternative way to establish validity would be to demonstrate that experimentally manipulating sexual motivation changes test scores in predicted ways (Borsboom et al., 2004, 2009). Previous research showing that individuals report greater sexual desire and more frequent sexual behavior following exposure to sexual compared to neutral stimuli (e.g., movies, stories) suggests that an appropriate experimental manipulation would also cause situational shifts on a state-adapted version of the TSMS (Both et al., 2004; Goldey & van Anders, 2012). The extent to which experimental manipulations may or may not alter sexual motivation at the trait level is an interesting question and avenue for future research.

Conclusion

Sexual motivation plays a fundamental role in people's day lives. We developed the TSMS, a brief six-item sexual motivation scale that is grounded in theory and allows for an economical, reliable, and valid assessment of trait sexual motivation irrespective of gender and relationship status.

Acknowledgments

We thank Paula Laute and Tobias Ludwig for their valuable assistance in creating the online surveys and preparing the manuscript.

Declaration of Interest Statement

The authors report there are no competing interests to declare.

Data Availability Statement

The data that support the findings of this study are openly available on the Open Science Framework at https://osf.io/ux9nk/.

References

Ågmo, A., & Laan, E. (2022a). Sexual incentive motivation, sexual behavior, and general arousal: Do rats and humans tell the same story? *Neuroscience & Biobehavioral Reviews, 135*, Article 104595. https://doi.org/10.1016/j.neubiorev.2022.104595

Ågmo, A., & Laan, E. (2022b). The sexual incentive motivation model and its clinical applications. *The Journal of Sex Research*, 1–20. https://doi.org/10.1080/00224499.2022.2134978

Allen, M. S., & Walter, E. E. (2018). Linking big five personality traits to sexuality and sexual health: A meta-analytic review. *Psychological Bulletin, 144*(10), 1081–1110. https://doi.org/10.1037/bul0000157

Arend, M. G., & Schäfer, T. (2019). Statistical power in two-level models: A tutorial based on Monte Carlo simulation. *Psychological Methods, 24*(1), 1–19. https://doi.org/10.1037/met0000195

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *ArXiv:1406.5823 [Stat]*. http://arxiv.org/abs/1406.5823

Baumeister, R. F., Catanese, K. R., & Vohs, K. D. (2001). Is there a gender difference in strength of sex drive? Theoretical views, conceptual distinctions, and a review of relevant evidence. *Personality and Social Psychology Review, 5*(3), 242–273. https://doi.org/10.1207/S15327957PSPR0503_5

Birnbaum, G. E. (2014). Sexy building blocks: The contribution of the sexual system to attachment formation and maintenance. In *Mikulincer M., Shaver P. R. (Eds.), Mechanisms of social connection: From brain to group* (pp. 315–332). American Psychological Association. https://doi.org/10.1037/14250-018

Birnbaum, G. E. (2018). The fragile spell of desire: A functional perspective on changes in sexual desire across relationship development. *Personality and Social Psychology Review*, *22*(2), 101–127. https://doi.org/10.1177/1088868317715350

Birnbaum, G. E., Kanat-Maymon, Y., Mizrahi, M., Recanati, M., & Orr, R. (2019). What fantasies can do to your relationship: The effects of sexual fantasies on couple interactions. *Personality and Social Psychology Bulletin*, *45*(3), 461–476. https://doi.org/10.1177/0146167218789611

Borsboom, D., Cramer, A. O. J., Kievit, R. A., Scholten, A. Z., & Franic, S. (2009). The end of construct validity. In *R. W. Lissitz (Ed.), The concept of validity: Revisions, new directions, and applications* (pp. 135–170). IAP Information Age Publishing.

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Both, S., Spiering, M., Everaerd, W., & Laan, E. (2004). Sexual behavior and responsiveness to sexual stimuli following laboratory-induced sexual arousal. *The Journal of Sex Research*, *41*(3), 242–258. https://doi.org/10.1080/00224490409552232

Brannick, M. T., Chan, D., Conway, J. M., Lance, C. E., & Spector, P. E. (2010). What is method variance and how can we cope with it? A panel discussion. *Organizational Research Methods*, *13*(3), 407–420. https://doi.org/10.1177/1094428109360993

Carpenter, D. L., Janssen, E., Graham, C. A., Vorst, H., & Wicherts, J. (2011). The sexual inhibition/sexual excitation scales-short form. In *T. D. Fisher, C. M. Davis, W. L. Yarber, & S. L. Davis (Eds.), Handbook of sexuality-related measures* (pp. 236–239). Routledge.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's Corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(3), 471–492. https://doi.org/10.1207/s15328007sem1203_7

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Clayton, A. H., Kingsberg, S. A., & Goldstein, I. (2018). Evaluation and management of hypoactive sexual desire disorder. *Sexual Medicine*, *6*(2), 59–74. https://doi.org/10.1016/j.esxm.2018.01.004

Davies, S., Katz, J., & Jackson, J. L. (1999). Sexual desire discrepancies: Effects on sexual and relationship satisfaction in heterosexual dating couples. *Archives of Sexual Behavior*, *28*(6), 553–567.

Eisinga, R., te Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*(4), 637–642. https://doi.org/10.1007/s00038-012-0416-3

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, *8*(4), 370–378. https://doi.org/10.1177/1948550617693063

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*(6), 1011–1027. https://doi.org/10.1037/0022-3514.80.6.1011

Fleeson, W., & Jayawickreme, E. (2015). Whole Trait Theory. *Journal of Research in Personality*, *56*, 82–92. https://doi.org/10.1016/j.jrp.2014.10.009

Frankenbach, J., Weber, M., Loschelder, D. D., Kilger, H., & Friese, M. (2022). Sex drive: Theoretical conceptualization and meta-analytic review of gender differences. *Psychological Bulletin*, *148*(9–10), 621–661. https://doi.org/10.1037/bul0000366

French, B. F., & Finch, W. H. (2008). Multigroup confirmatory factor analysis: Locating the invariant referent sets. *Structural Equation Modeling: A Multidisciplinary Journal*, *15*(1), 96–113. https://doi.org/10.1080/10705510701758349

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Gailliot, M. T., & Baumeister, R. F. (2007). Self-regulation and sexual restraint: Dispositionally and temporarily poor self-regulatory abilities contribute to failures at restraining sexual behavior. *Personality and Social Psychology Bulletin*, *33*(2), 173–186. https://doi.org/10.1177/0146167206293472

Goldey, K. L., & van Anders, S. M. (2012). Sexual arousal and desire: Interrelations and responses to three modalities of sexual stimuli. *The Journal of Sexual Medicine*, *9*(9), 2315–2329. https://doi.org/10.1111/j.1743-6109.2012.02845.x

Gonzalez, O., MacKinnon, D. P., & Muniz, F. B. (2021). Extrinsic convergent validity evidence to prevent jingle and jangle fallacies. *Multivariate Behavioral Research*, *56*(1), 3–19. https://doi.org/10.1080/00273171.2019.1707061

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*(3), 213–224.

Greiff, S., & Scherer, R. (2018). Still comparing apples with oranges? Some thoughts on the principles and practices of measurement invariance testing. *European Journal of Psychological Assessment*, *34*(3), 141–144. https://doi.org/10.1027/1015-5759/a000487

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hui Xu, & Tracey, T. J. G. (2017). Use of multi-group confirmatory factor analysis in examining measurement invariance in counseling psychology research. *European Journal of Counselling Psychology*, *6*(1), 75–82. https://doi.org/10.5964/ejcop.v6i1.120

Janssen, E., Vorst, H., Finn, P., & Bancroft, J. (2002). The sexual inhibition (SIS) and sexual excitation (SES) scales: I. Measuring sexual inhibition and excitation proneness in men. *The Journal of Sex Research*, *39*(2), 114–126. https://doi.org/10.1080/00224490209552130

Jones, B. C., Hahn, A. C., Fisher, C. I., Wang, H., Kandrik, M., & DeBruine, L. M. (2018). General sexual desire, but not desire for uncommitted sexual relationships, tracks changes in women's hormonal status. *Psychoneuroendocrinology*, *88*, 153–157. https://doi.org/10.1016/j.psyneuen.2017.12.015

Kang, Y., McNeish, D. M., & Hancock, G. R. (2016). The role of measurement quality on practical guidelines for assessing measurement and structural invariance. *Educational and Psychological Measurement*, *76*(4), 533–561. https://doi.org/10.1177/0013164415603764

Kim, J. J., Muise, A., Barranti, M., Mark, K. P., Rosen, N. O., Harasymchuk, C., & Impett, E. (2021). Are couples more satisfied when they match in sexual desire? New insights from response surface analyses. *Social Psychological and Personality Science*, *12*(4), 487–496. https://doi.org/10.1177/1948550620926770

Klein, V., Schmidt, A. F., Turner, D., & Briken, P. (2015). Are sex drive and hypersexuality associated with pedophilic interest and child sexual abuse in a male community sample? *PLoS ONE*, *10*(7), 1–11. https://doi.org/10.1371/journal.pone.0129730

Levine, S. B. (2003). The nature of sexual desire: A clinician's perspective. *Archives of Sexual Behavior*, *32*(3), 279–285. https://doi.org/10.1023/A:1023421819465

Lippa, R. A. (2006). Is high sex drive associated with increased sexual attraction to both sexes? It depends on whether you are male or female. *Psychological Science*, *17*(1), 46–52. https://doi.org/10.1111/j.1467-9280.2005.01663.x

Lippa, R. A. (2009). Sex differences in sex drive, sociosexuality, and height across 53 nations: Testing evolutionary and social structural theories. *Archives of Sexual Behavior*, *38*(5), 631–651. https://doi.org/10.1007/s10508-007-9242-8

Little, T. D., Slegers, D. W., & Card, N. A. (2006). A non-arbitrary method of identifying and scaling latent variables in SEM and MACS models. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(1), 59–72. https://doi.org/10.1207/s15328007sem1301_3

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*(3), 635–694. https://doi.org/10.2466/pr0.1957.3.3.635

Luong, R., & Flake, J. K. (2022). Measurement invariance testing using confirmatory factor analysis and alignment optimization: A tutorial for transparent analysis planning and reporting. *Psychological Methods*. https://doi.org/10.1037/met0000441

Mark, K. P. (2012). The relative impact of individual sexual desire and couple desire discrepancy on satisfaction in heterosexual couples. *Sexual and Relationship Therapy*, *27*(2), 133–146. https://doi.org/10.1080/14681994.2012.678825

Mark, K. P., Toland, M. D., Rosenkrantz, D. E., Brown, H. M., & Hong, S. (2018). Validation of the sexual desire inventory for lesbian, gay, bisexual, trans, and queer adults. *Psychology*

of Sexual Orientation and Gender Diversity, *5*(1), 122–128. https://doi.org/10.1037/sgd0000260

McCrae, R. R., & Costa, P. T. (2003). *Personality in adulthood: A five-factor theory perspective*. Guilford Press.

McNeish, D., & Wolf, M. G. (2021). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*. https://doi.org/10.1037/met0000425

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134). https://doi.org/10.1098/rsif.2017.0213

Niemand, T., & Mai, R. (2018). Flexible cutoff values for fit indices in the evaluation of structural equation models. *Journal of the Academy of Marketing Science*, *46*(6), 1148–1172. https://doi.org/10.1007/s11747-018-0602-9

Nilsen, F. A., Bang, H., Boe, O., Martinsen, Ø. L., Lang-Ree, O. C., & Røysamb, E. (2020). The multidimensional self-control scale (MSCS): Development and validation. *Psychological Assessment*, *32*(11), 1057–1074. https://doi.org/10.1037/pas0000950

Nye, C. D., & Drasgow, F. (2011). Effect size indices for analyses of measurement equivalence: Understanding the practical importance of differences between groups. *Journal of Applied Psychology*, *96*(5), 966–980. https://doi.org/10.1037/a0022955

Ostovich, J. M., & Sabini, J. (2004). How are sociosexuality, sex drive, and lifetime number of sexual partners related? *Personality and Social Psychology Bulletin*, *30*(10), 1255–1266. https://doi.org/10.1177/0146167204264754

Park, Y., & MacDonald, G. (2022). Single and partnered individuals' sexual satisfaction as a function of sexual desire and activities: Results using a sexual satisfaction scale

demonstrating measurement invariance across partnership status. *Archives of Sexual Behavior*, *51*(1), 547–564. https://doi.org/10.1007/s10508-021-02153-y

Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2021). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, *54*(4), 1643–1662. https://doi.org/10.3758/s13428-021-01694-3

Penke, L., & Asendorpf, J. B. (2008). Beyond global sociosexual orientations: A more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *Journal of Personality and Social Psychology*, *95*(5), 1113–1135. https://doi.org/10.1037/0022-3514.95.5.1113

Pinkerton, S. D., Bogart, L. M., Cecil, H., & Abramson, P. R. (2003). Factors associated with masturbation in a collegiate sample. *Journal of Psychology & Human Sexuality*, *14*(2–3), 103–121. https://doi.org/10.1300/J056v14n02_07

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, *63*(1), 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Regnerus, M., Gordon, D., & Price, J. (2016). Documenting pornography use in America: A comparative analysis of methodological approaches. *Journal of Sex Research*, *53*(7), 873–881. https://doi.org/10.1080/00224499.2015.1096886

Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality*, *43*(2), 137–145. https://doi.org/10.1016/j.jrp.2008.12.015

Roney, J. R., & Simmons, Z. L. (2013). Hormonal predictors of sexual motivation in natural menstrual cycles. *Hormones and Behavior*, *63*(4), 636–645. https://doi.org/10.1016/j.yhbeh.2013.02.013

Röseler, L., Wolf, D., Leder, J., & Schütz, A. (2020). *Test-retest reliability is not a measure of reliability or stability: A friendly reminder* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/mt49r

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Rudnev, M., Lytkina, E., Davidov, E., Schmidt, P., & Zick, A. (2018). Testing measurement invariance for a second-order factor. A cross-national test of the alienation scale. *Methods, Data, Analyses: A Journal for Quantitative Methods and Survey Methodology (Mda)*, *12*(1), 47–76. https://doi.org/10.12758/MDA.2017.11

Sakaluk, J. K., Fisher, A. N., & Kilshaw, R. E. (2021). Dyadic measurement invariance and its importance for replicability in romantic relationship science. *Personal Relationships*, *28*(1), 190–226. https://doi.org/10.1111/pere.12341

Simms, L. J. (2008). Classical and modern methods of psychological scale construction. *Social and Personality Psychology Compass*, *2*(1), 414–433. https://doi.org/10.1111/j.1751-9004.2007.00044.x

Simpson, J. A., & Gangestad, S. W. (1991). Individual differences in sociosexuality: Evidence for convergent and discriminant validity. *Journal of Personality and Social Psychology*, *60*(6), 870–883. https://doi.org/10.1037/0022-3514.60.6.870

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, *68*, 69–81. https://doi.org/10.1016/j.jrp.2017.02.004

Spector, I. P., Carey, M. P., & Steinberg, L. (1996). The sexual desire inventory: Development, factor structure, and evidence of reliability. *Journal of Sex & Marital Therapy*, *22*(3), 175–190. https://doi.org/10.1080/00926239608414655

Stark, R., Kagerer, S., Walter, B., Vaitl, D., Klucken, T., & Wehrum-Osinsky, S. (2015). Trait sexual motivation questionnaire: Concept and validation. *The Journal of Sexual Medicine*, *12*(4), 1080–1091. https://doi.org/10.1111/jsm.12843

Toates, F. (2009). An integrative theoretical framework for understanding sexual motivation, arousal, and behavior. *Journal of Sex Research*, *46*(2–3), 168–193. https://doi.org/10.1080/00224490902747768

Vallejo-Medina, P., Rojas-Paoli, I., & Álvarez-Muelas, A. (2020). Validation of the sexual desire inventory in Colombia. *Journal of Sex & Marital Therapy, 46*(4), 385–398. https://doi.org/10.1080/0092623X.2020.1739181

Supplementary Online Materials (SOM)

Table S1

First Version of the TSMS (8 Items)

In a typical week:

Cognition	How often do you think about sex? How often do you have sexual fantasies? How often do you have sexual daydreams?
Behavior	How often are you sexually active (self-stimulation plus sex with another person)? How often do you engage in sexual activities (whenever you masturbate or have sex with a partner)?
Affect	How often do you feel sexual desire? How often do you feel like having sex? How often do you feel “turned on”?

Note. The first version of the TSMS used an open response format.

Table S2

Simple One-Factor Model: Model Fit in Total Sample and Subsamples

Sample	*N*	χ^2	*df*	*p*	CFI	SRMR	TLI	RMSEA
Total	658	557.89	9	< .001	.840	.067	.733	.304
Women	329	324.12	9	< .001	.797	.076	.661	.326
Men	329	243.87	9	< .001	.845	.076	.741	.282
Single	327	270.39	9	< .001	.852	.062	.753	.298
In a Relationship	331	304.93	9	< .001	.824	.074	.707	.315

Table S3

Nomological Validity of the BTSMS: Constructs, Instruments, Preregistered Predictions, and Results

Constructs	Instruments	$r_{expected}$	$r_{observed}$	t	df	p
Sex Drive	Sexual Attitudes and Feelings Scale, Subscale "Sex Drive" (SAF; Lippa, 2006)	$r \geq .70$	.68	19.67	448	< .001
Sexual Desire	Sexual Desire Inventory 2, "Total Sexual Desire" (SDI; Spector et al., 1996)	$r \geq .70$	.65	12.96	224	< .001
Sexual Restraint	Sexual Restraint Scale (SRS; Gailliot & Baumeister, 2007)	$-.40 \leq r \leq -.20$	-.23	-5.02	448	< .001
Sociosexuality	Revised Sociosexual Orientation Inventory (SOI-R, Subscale "Attitudes"; Penke & Asendorpf, 2008)	$.10 \leq r \leq .30$	.21	4.62	448	< .001
Self-Control	Brief Multidimensional Self-Control Scale (BMSCS; Nilsen et al., 2020)	$-.20 \leq r \leq .00$	-.03	-0.69	448	.493
Big Five	Big Five Inventory, short version (BFI-2-S; Soto & John, 2017)					
O: Openness		$.00 \leq r \leq .20$	.06	1.25	448	.210
C: Conscientiousness		$-.20 \leq r \leq .00$	-.12	-2.53	448	.012
E: Extraversion		$.00 \leq r \leq .20$	-.03	-0.60	448	.546
A: Agreeableness		$-.20 \leq r \leq .00$	-.05	-1.10	448	.274
N: Neuroticism		$-.20 \leq r \leq .20$	.01	0.16	448	.870

Note. *t*, *df*, and *p* refer to two-tailed tests of correlations with the TSMS against zero (i.e., no correlation).

Table S4

Criterion Validity of the BTSMS

	Pornography use			Orgasm frequency			Time spent with sexuality			First masturbation (Age)		
Model	β	R^2	ΔR^2	β	R^2	ΔR^2	β	R^2	ΔR^2	β	R^2	ΔR^2
M1. Covariate Model		.285***			.075***			.028**			.013	
Gender	1.04***			0.54***			0.33***			-0.11		
Age	-0.17***			0.01			-0.06			0.11*		
M2a. Incremental Model: BTSMS		.475***	.190a***		.300***	.225a***		.203***	.175a***		.063***	.050a***
Gender	0.69***			0.17			-0.01			0.08		
Age	-0.19***			-0.01			-0.07			0.11*		
BTSMS	0.47***			0.51***			0.45***			-0.24***		
M2b. Incremental Model: SAF		.361***	.076a***		.199***	.124a***		.185***	.157a***		.029**	.016a**
Gender	0.85***			0.31***			0.06			-0.02		
Age	-0.17***			0.02			-0.05			0.10*		
SAF	0.29***			0.37***			0.42***			-0.13**		
M3. Comparative Model		.475***	.000b/.114c***		.303***	.003b/.104c***		.231***	.028b***/.046c***		.063***	.000b/.034 c***
Gender	0.69***			0.15			-0.04			0.08		
Age	-0.19***			-0.01			-0.06			0.11*		
SAF	-0.01			0.09			0.23***			0.02		
BTSMS	0.47***			0.45***			0.30***			-0.26***		

* $p < .05$; ** $p < .01$; *** $p < .001$. [a]compared to M1; [b]compared to M2a; [c]compared to M2b. To present easily interpretable values, we standardized all continuous variables and left gender in its original metric (e.g., average change in SD units of pornography use if a person is male rather than female).

Table S5

Predictive Validity of the BTSMS: Associations with Sexual Outcomes in Everyday Life

Criterion	*B*	*SE*	$CI_{B\ 95\%}$	*z*	*p*	Effect
Event Frequencies						
Cognition	0.50	0.04	[0.43, 0.57]	13.51	< .001	1.65[a]
Affect	0.51	0.04	[0.43, 0.58]	13.00	< .001	1.66[a]
Behavior	0.54	0.06	[0.42, 0.65]	9.05	< .001	1.71[b]
Criterion Outcomes						
Pornography use	0.54	0.08	[0.39, 0.70]	6.94	< .001	1.72[b]
Time spent with sexuality	0.42	0.04	[0.34, 0.51]	9.61	< .001	0.15[c]
Further Sexual Experiences						
Self-rated sex drive	0.62	0.05	[0.53, 0.72]	12.41	< .001	0.23[c]
Sexual excitability	0.58	0.05	[0.48, 0.67]	11.89	< .001	0.20[c]

Note. Intercepts are not displayed for the sake of clarity. Behavior (0 = *no*, 1 = *yes*) contains information from solitary and dyadic events (preregistered) to prevent zero inflation. Effect sizes are [a]incidence rate ratios for event frequencies (IRR values greater than 1 indicate a positive association between BTSMS scores and event frequencies), [b]odds ratios for binary outcomes (positive association: OR > 1), and [c]$R^2_{marginal}$ (i.e., proportion of the total variance explained by the fixed effect; Nakagawa et al., 2017) for continuous outcomes (positive association: $R^2_{marginal} > 0$).

PART II

How Valid Are Self-Reports of Sexual Motivation?

Using the Item Sum Technique to Examine Self-Presentation Tendencies in Online Research

Marcel Weber

Saarland University

Helena Kilger

Saarland University

Malte Friese

Saarland University

Manuscript submitted for publication in *Motivation Science*.
(Currently under revision)

This is a manuscript submitted for publication in Motivation Science

on July 29, 2024.

How Valid Are Self-Reports of Sexual Motivation?

Using the Item Sum Technique to Examine Self-Presentation Tendencies in Online Research

Marcel Weber[1], Helena Kilger[1], and Malte Friese[1]

[1]Department of Psychology, Saarland University

Author Note

Marcel Weber: https://orcid.org/0000-0002-0507-0070

Helena Kilger: https://orcid.org/0000-0003-4165-3748

Malte Friese: https://orcid.org/0000-0003-0055-513X

Corresponding Author

Correspondence concerning this article should be addressed to Marcel Weber, Department of Psychology, Saarland University, Campus A2 4, 66123 Saarbrücken, Germany. E-mail: marcel.weber@uni-saarland.de.

Abstract

Honest self-reporting is crucial for valid conclusions about motivation, but it cannot be taken for granted. This is particularly true for the private and sensitive domain of sexual motivation, where participants may be tempted to adjust their responses to (gender-specific) social norms, with far-reaching consequences for society and science (e.g., potential overestimation of gender differences in sexual motivation). In the present research, we examined whether the hope that online data collection could help to overcome the problem of socially desirable responses to indicators of (sexual) motivation is justified. We used the *Item Sum Technique*, an indirect questioning technique that maximizes people's anonymity, to create a strong standard of comparison for standard online self-reports (N_{total} = 2,857). For several theory-driven indicators of sexual motivation, frequency estimates and gender differences were unaffected by response method, suggesting little evidence of social desirability bias in online studies. Small to no gender differences for sexual bias indicators, weak associations with social desirability scales, and near-maximal levels of self-reported honesty were secondary indicators that further support the accuracy of (gender differences in) self-reported sexual motivation in online research. We discuss these findings in light of the lively debate about the validity of gender differences in sexual motivation.

Keywords: sexual motivation, social desirability bias, Item Sum Technique, indirect questioning

Word count: 6,654 words (Abstract: 201 words)

How Valid Are Self-Reports of Sexual Motivation?

Using the Item Sum Technique to Examine Self-Presentation Tendencies in Online Research

How often do people think about, desire, or have sex? For many people, sexuality is one of the most important and private areas of life. To answer such questions, psychological scientists rely primarily on self-reports to find out what goes on in people's minds or behind closed doors. However, due to their sensitive nature, sexual self-reports can be biased in socially desirable ways, potentially compromising the validity of the conclusions drawn. Under- or overestimation of the frequency of sexual events and the magnitude of gender differences in sexual motivation are some of the many ways, in which self-presentation may lead to biased perceptions of the world. This can have profound implications for society and well-being such as the formation and perpetuation of gender stereotypes and an exaggeration of perceived sexual desire discrepancies as a possible challenge in (mixed-sex) couples.

Scientists have long recognized that online surveys are a promising candidate for combating social desirability bias (Buchanan, 2000). Compared to studies conducted in a laboratory, participants in online surveys are unsupervised and relatively anonymous. Theoretically, in the absence of another person, self-presentation is less likely to significantly bias self-reports. However, previous research has found mixed results regarding whether self-reports on socially sensitive topics are more accurate when collected online (e.g., Carlbring et al., 2007; Fogarty et al., 2013; Risko et al., 2006; for an overview, see Gnambs & Kaspar, 2017).

The present research was designed to better understand the extent to which the accuracy of sexual self-reports in online research is undermined by social desirability bias. Here, we used the *Item Sum Technique* (IST), a validated indirect questioning technique designed to minimize self-presentation tendencies, to create a strong standard of comparison. To provide a more

complete picture, this *indirect* approach was complemented by three other ways of testing the accuracy of online self-reporting: (1) a *logic* approach that examines gender differences on questions for which (pronounced) true gender differences are unlikely, (2) a *subjective* approach that analyzes participants' self-reported levels of honesty, and (3) a *control* approach that considers the associations between self-reported sexual motivation and social desirability scales. We focused on a core aspect of sexuality that is relevant to most people's lives, extensively studied, dependent on (accurate) self-report, and at high risk of social desirability bias: sexual motivation and gender differences therein.

Sexual Motivation, Sexual (Double) Standards, and Social Desirability Bias

Sexual motivation is the latent driving force behind the pursuit of sexual pleasure. It manifests itself in cognitive, affective, and behavioral sexual events such as sexual fantasies, sexual desire, and sexual self-stimulation (Frankenbach et al., 2022). Decades of research have demonstrated the interest of scientists in unraveling the mysteries of sexual motivation and its multiple individual and interpersonal implications. For instance, higher sexual motivation is associated with higher initial romantic interest among those who start dating (Eastwick et al., 2023) and with greater sexual satisfaction and relationship satisfaction among those who are in a romantic relationship (Kim et al., 2021; Mark, 2015).

However, social scientists seeking to better understand this private and sensitive area of life face a dilemma. Sexual events that are indicative of sexual motivation are rarely directly observable, taking place in privacy or in people's minds. Physiological indicators such as changes in penile circumference or vaginal blood volume are an important addition to the assessment repertoire, but they are resource-intensive and do not directly reflect the psychological elements of sexual motivation. In the absence of viable alternatives, self-report measures have become the standard way to assess sexual motivation (but see creative new developments, Hinzmann et al.,

2023; Schultheiss et al., 2023). Self-report measures allow for cost-effective assessment on a large scale, which has made the wealth of research on sexual motivation possible in the first place.

The validity of sexual self-reports, however, depends on participants' propensity to respond in an honest and unbiased way, which cannot be taken for granted. Because of the sensitive nature of sexuality, researchers must be concerned that people's responses will be systematically biased by their perceptions of what society considers appropriate, ethical, and desirable. Previous research has assumed that there is a sexual double standard in the societal evaluation of male and female sexuality: that sexual activity and permissiveness are socially rewarded for men but socially punished for women (Marks & Fraley, 2005). When people are asked to report on sexual events, gendered social norms are likely to result in "answers distorted in opposite directions for men and women such that men may be motivated to […] exaggerate the frequency and variability of their sexual encounters, whereas women may be motivated to understate theirs" (Alexander & Fisher, 2003, p. 28).

The potential consequences of this (gender-specific) social desirability bias are far-reaching. Tendencies to over- or underreport sexual events are associated with uncertainties about the true prevalence and frequency of the events under study. This makes it difficult to estimate the role that sexual motivation plays in people's daily lives. In addition, unbiased reporting is key to answering the question of whether men and women differ in their average levels of sexual motivation. If women tend to underreport sexual events and men tend to overreport, measured gender differences would be exaggerated. Since the degree of gender-specific desirable responding is unknown, dozens of previous studies showing higher male than female sexual motivation (Baumeister et al., 2001) have not been able to end debates about (the magnitude of)

gender differences (e.g., Conley et al., 2011; Schmitt et al., 2012), nor will future studies conducted in the same way.

Findings coming from laboratory research suggest that the risk of biased self-reports and conclusions is not just hypothetical but real. Meston (1998) found small to moderate associations between sexual self-reports and scales measuring impression management and self-deceptive enhancement. For women, these associations were mostly negative (i.e., higher social desirability scores associated with fewer sexual events) and more consistent than for men. Alexander and Fisher (2003) compared sexual self-reports between an exposure threat group in which participants were led to believe that the experimenter might their answers and a bogus pipeline group, in which participants were connected to a false polygraph that supposedly detects lies. Some of the gender differences were found in the exposure threat group, but not in the bogus pipeline group, in which participants may be particularly likely to tell the truth.

While creative methods such as the bogus pipeline procedure may be able to elicit (more) accurate responses, they have obvious pragmatic (e.g., time, financial resources) and ethical (e.g., deception) limitations that restrict their widespread use in sexuality research.

More Accurate Responses in Online Research due to Higher Anonymity

An alternative route to more honest responses that comes at no cost is to increase the anonymity of participants. While the bogus pipeline technique exploits the fear of detection and sanction for lying, high anonymity can facilitate honest responses by removing the fear of detection and sanction for a socially undesirable response. In online surveys, which have become the norm in recent years, high anonymity is a key feature. Online surveys do not require participants to interact with an experimenter or interviewer, thus removing the contextual factors that might motivate the self-presentation tendency. This is critically different from most laboratory studies, including those providing support for biased sexual self-reporting, in which

participants were placed in a larger group of individuals (e.g., Meston et al., 1998) or had to interact with an experimenter (e.g., Alexander & Fisher, 2003; Fisher, 2013).

Previous research examining whether self-reports of socially sensitive topics are more accurate in online surveys has yielded mixed results. A triplet of meta-analyses covering moderately sensitive topics (e.g., personality characteristics) found no effect of survey mode (Gnambs & Kaspar, 2017). In contrast, another meta-analysis revealed higher prevalence rates of socially sensitive behaviors that are viewed negatively in society in computerized, often web-based surveys compared to paper-pencil administration. These group differences were most pronounced in particularly sensitive areas such as drug use or sexuality (Gnambs & Kaspar, 2015). However, it is possible that factors such as online privacy concerns (Evans & Mathur, 2018) and self-deceptive reinforcement (Paulhus, 1988) may still prevent participants from providing fully accurate responses. Therefore, findings suggesting that online survey responses are less biased than responses from laboratory studies should not be misinterpreted as evidence of absence of bias.

Using the Item Sum Technique to Test the Accuracy of Sexual Self-Reports in Online Research

Indirect questioning techniques, such as the Item Sum Technique, can provide the strong standard of comparison needed to better estimate the accuracy of sexual self-reports in online research. Rather than asking participants to answer a sensitive question directly, indirect questioning techniques ask participants to mask their true answer using non-sensitive information that only they have access to. For example, when studying sexual motivation, a standard direct questioning (DQ) group may be asked to directly indicate how often they masturbate in a typical week. An item sum (IS) group may be presented the same sensitive question together with two-nonsensitive questions (e.g., asking for the last two digits of a friend’s phone number). Instead of

answering each question directly, the IS group is asked to report the sum of the answers to the three items as one joint response only (see Figure 1, Panel A). Since the relative contribution of the three answers to the joint response remains unknown, any participant's privacy is protected, and this process is completely transparent to the participants. Crucially, however, the mean level estimates for the sensitive question across participants in the IS group *can* be accurately computed indirectly if researchers know the sampling characteristics of the sum of the two non-sensitive questions (i.e., the mean of the sum of all three questions minus the mean estimate of the sum of the two non-sensitive questions)[1].

There is considerable evidence that indirect questioning can promote more accurate responses to sensitive questions, mostly from studies using binary questions or statements (e.g., "I make use of marijuana"; Chaudhuri & Christofides, 2007). Two meta-analyses found that estimated prevalence rates were higher when indirect rather than direct questioning was used to study socially sanctioned behaviors that people are prone to underreport (Ehler et al., 2021; Li & Van den Noortgate, 2022). The Item Sum Technique specifically was developed as an extension and advancement of the existing item count technique for non-binary questions (e.g., "How many cannabis cigarettes did you consume last year?"; Perri et al., 2018). Studies using this technique found higher frequency estimates for cannabis consumption, behaviors indicative of sexual addiction, and undeclared hours of work in the IS group compared to the DQ group (Perri et al., 2018; Trappmann et al., 2014), suggesting higher accuracy in the indirect IS group.

The Present Research

Socially desirable responding is a fundamental threat to the validity of sexual self-reports. The purpose of the present research was to better assess whether the hopes placed in online data

[1] Characteristics of the distribution of non-sensitive questions may either be retrieved conceptually (e.g., assuming a uniform distribution for phone-number digits) or empirically (e.g., collecting data from a short-list group only reporting the sum of phone-number digits).

collection for providing more accurate responses and conclusions are justified. To rigorously examine the (un)biased nature of sexual self-reporting in online research, we compared estimates of sexual motivation and gender differences therein between two groups: a direct questioning (DQ) group that answered sexual questions directly (i.e., standard online study) and an item sum (IS) group that answered the questions indirectly by providing the sum of the answers to several sexual and non-sexual questions. Based on past research on sexual (double) standards and socially desirable responding, we preregistered the following predictions.

First, *if* there are differences between the groups, we expected event frequencies indicative of sexual motivation to be higher in the IS group than in the DQ group (i.e., H1: main effect group). This should be particularly true for women, who are expected to be sexually reserved according to the traditional sexual double standard (H1a). In light of inconsistent findings regarding male sexual norms, we did not specify directed hypotheses for group differences among men (H1b).

Second, *if* there are gender differences, we expected sexual motivation to be higher for men than for women (i.e., H2: main effect gender). We expected to find this difference in the DQ group specifically (H2a) and we also explored it in the IS group (H2b). Hence, if there were biasing self-presentation tendencies in the DQ group (standard online survey), the main effects may be qualified by a group x gender interaction (i.e., stronger tendency of women to underreport sexual events).

Examining the extent to which self-reported sexual motivation and gender differences therein obtained in the DQ group (i.e., standard online survey) were different from the IS group, for which accurate responses are more likely, was our primary way of testing the accuracy of sexual self-reports in online research. To get a more complete picture, we explored three additional approaches to social desirability bias.

In what we call the *logic* approach, we used items asking about participants' frequency of sexual intercourse and number of lifetime sexual partners as sexual bias indicators. These pieces of information are (at least) as socially sensitive as the sexual motivation indicators used. Critically, however, the true population-level gender difference on the bias indicators is close to zero in heterosexual populations of similar age (Frankenbach et al., 2022). Substantial gender differences on these bias indicators may therefore indicate (gender-specific) self-presentation tendencies.

The *subjective* approach was to examine and compare participants' self-reported levels of honesty. Honesty ratings were provided after the completion of the main part of the study. Participants were reassured that admitting to biased answers to sexual questions would not affect their payment. In this way, we aimed to open the door to accurate honesty reports by removing potential social, moral, or financial doubts.

Within the control approach, we used social desirability scales to assess people's domain-general tendency for self-deceptive enhancement and impression management, which constitute distinct factors of biased responding (Paulhus, 1988). Consistent with recent recommendations for assessing and controlling for desirable responding (King, 2022), we examined associations between these measures and self-reported sexual motivation. Based on the sexual double standard, we would expect that if present, social desirability bias should manifest itself in associations between social desirability scales and sexual self-reports that are negative for women (i.e., higher social desirability scores associated with fewer reported events) and positive for men (i.e., higher social desirability scores associated with more reported events). Because differences between men and women may in part be the result of gender-specific self-presentation, we tested whether gender differences would remain unchanged when controlling for self-deceptive enhancement and impression management.

Method

Transparency Statement

We conducted two preregistered studies that featured identical main parts. All study materials (i.e., online questionnaires, codebooks), scripts, preregistration files, and data are openly available on the Open Science Framework (https://osf.io/q8bfv/?view_only=4210b0d438dd47328417aee0c3d17421). We report how we determined our sample size, all data exclusions, and all manipulations. All measures relevant to the research questions are presented in the manuscript, the full list of measures can be found in the codebooks (adapted from Simmons et al., 2012). Most of the analyses reported in this manuscript are consistent with the preregistered analytic plans, but we made some adjustments to maximize the reliability and interpretability of our results, which we transparently report in the Methods section. One deviation was that we used the total sample of all participants (i.e., Study 1 plus Study 2) whenever possible rather than analyzing both studies separately. The individual samples are likely to be strongly affected by the randomness introduced by the IS procedure, compromising the reliability of our estimates in the individual samples. Analyses based on the total sample follow the analysis plan described in the second preregistration file, which is similar to the first one but provides a greater level of detail. The research project was approved by the Institutional Review Board of Saarland University (approval number: 20-20). We used R, version 4.2.1, to analyze the data (R Core Team, 2023).

Participants and Power

We used pilot data from another project (Weber et al., 2024) in which participants responded directly to questions about sexual fantasies, desires, and sexual self-stimulation to conduct a-priori power analyses. Monte-Carlo simulations (10,000 iterations) indicated that a sample size of N = 2,500 would provide (at least) 80% power to detect group differences (i.e., IS

vs. DQ) as small as the preregistered minimum effect size of interest of 0.5 events per unit (e.g., 0.5 fantasies per day)[2].

We recruited a total of N = 2,980 participants, all living in the United Kingdom, using the online crowdsourcing platform Prolific.com. After applying the preregistered exclusion criteria (i.e., missed attention check, low self-reported data quality), a final sample of N = 2,857 participants remained ($n_{\text{Study 1}}$ = 1,325, $n_{\text{Study 2}}$ = 1,532; 53.5% male, 46.5% female; age: M = 27.34, SD = 6.37, range: 18-40 years; 53.1% in a romantic relationship; 82.1% heterosexual) who completed an online survey about sexuality created in SoSci Survey (Leiner, 2022). Participants were assigned to one of two groups: a direct questioning group (DQ group, n = 1,535) who answered the sexual items directly, and an item sum group (IS group, n = 1,322) who reported the sum of answers to sexual and innocuous questions. Both groups received a payment equivalent to U.S. $12.50/hr at the time the study was conducted.

Procedure, Instructions, and Measures

Both studies comprised of three parts: an intro (background and instructions), the main part (sexual self-reports), and the outro (reflection on study participation). Unless otherwise stated, measures were assessed in both studies.

Intro: Background and Instructions

After giving informed consent, participants answered background questions (sex, age, sexual orientation, and relationship status). Next, they received information and instructions, preparing them for the main part of the study. All participants were informed that they will answer questions that "some people may feel uncomfortable to answer honestly," but that truthful answers are of "utmost importance" for the success of the study. Participants were then asked to

[2] We initially used a different and less accurate method to estimate the standard deviation in the IS group, which resulted in an underestimation of the sample size needed to detect small differences with higher power. We therefore adjusted our analyses and pooled the two samples to obtain higher statistical power (see Methods section for details).

"answer completely honestly." Participants assigned to the IS group were further asked to select five numbers from their phone book and write down the last two digits of each (e.g., number A: 91, number B: 38). Next, participants in the IS group were introduced to the item sum principle:

> To ensure truthful responses in this study, we will employ a special response format that maximizes your anonymity. With this response format, you do not have to worry about what other people might think about your responses – because nobody will ever know your responses to any of the specific questions.

Participants assigned to the IS group generated the joint response for a fictitious person's cocaine use (times used last year: 2; phone number digits: 8, 0) and could only continue if they indicated the correct response (2 + 8 + 0 = 10) to make sure that they understood the procedure.

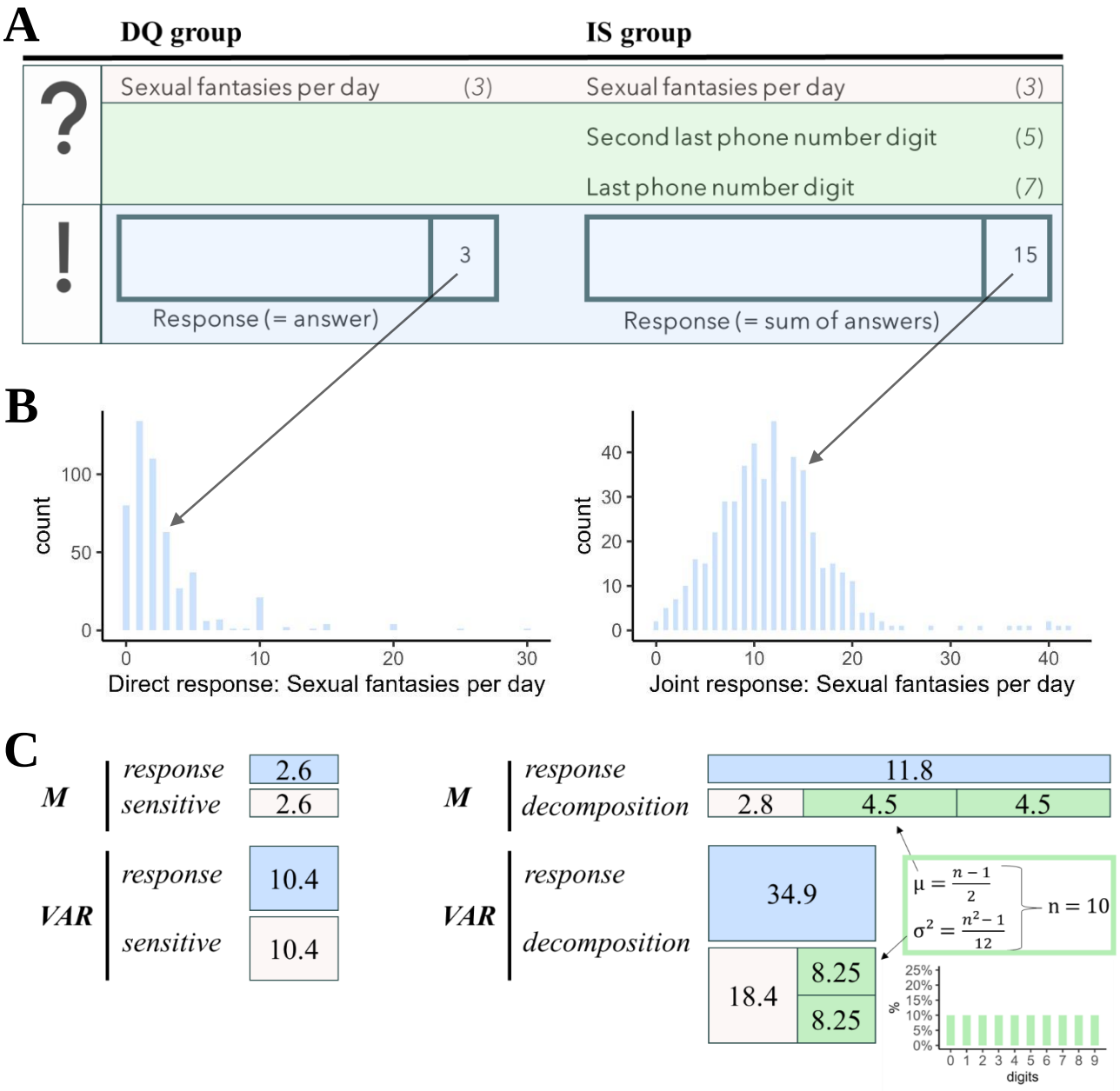
Main Part: Sexual Self-Reports

In the second part, participants provided self-reports of sexual events, presented in randomized order, using an open-response format. The DQ group answered each sexual question directly, whereas the IS group provided one joint response that was the sum of the answers to one sexual question and two non-sensitive (i.e., phone number digits) questions. This procedure is illustrated in the upper part of Figure 1.

Sexual motivation. Our primary outcome was sexual motivation, which manifests itself in the frequency of sexual cognition, sexual affect, and sexual behavior (Frankenbach et al., 2022). We assessed each facet with one item, adapting items from the Trait Sexual Motivation Scale (Weber et al., 2024; "On a typical day: How often do you have sexual fantasies?"; "On a typical day: How often do you feel sexual desire?"; "In a typical week: How often do you pleasure yourself sexually?").

Figure 1

Item Sum Technique: Response Format and Estimation Procedure



Note. Panel A illustrates the Direct Questioning (DQ) and Item Sum (IS) response format for one fictious participant whose answers are presented in parentheses. The DQ group answers the sensitive questions (e.g., number of sexual fantasies) directly, while the IS group only provides a joint response (blue-colored area), which is the sum of the answer to the same sensitive question (light red-colored area) and two non-sensitive questions (e.g., two phone number digits; green-

colored area). The histogram in Panel B shows the distribution of the responses provided by this participant (indicated by the arrow) and other fictitious participants in the two groups. For the DQ group, the distribution characteristics (e.g., mean [M], variance [VAR]) of the response and the sensitive question are identical. For the IS group, the estimated characteristics of the sensitive question can be calculated as the difference between the characteristics of the joint responses and the population characteristics of the non-sensitive responses (i.e., phone number digits), which follow approximately a uniform distribution for n = 10 possible outcomes (i.e., digits 0 to 9; see Panel C).

Sexual bias indicators. We used items assessing frequency of sexual intercourse ("In a typical week: How often are you sexually active with another person?," adapted from Weber et al., 2024) and the number of lifetime sexual partners ("In your lifetime: With how many different partners did you have sex [oral, vaginal, or anal]?) as sexual bias indicators. Unlike indicators of sexual motivation, for which the true magnitude of gender differences is unknown, the population-level gender differences for the bias indicators are known to be close to zero among heterosexual individuals.

Outro: Reflecting on Study Participation

In the final part of the study, participants reflected on their participation.

Perceived anonymity and self-reported honesty. Participants rated the degree to which they felt anonymous ("I felt anonymous while answering the questions") and to which they answered the sexual questions honestly ("I answered the questions honestly"), using 7-point unipolar scales (1 = *not at all* to 7 = *very much*). They were assured that their responses would not affect their payment.

Social desirability scales (only Study 2). In the second study, we used the Balanced Inventory of Desirable Responding short form (BIDR-16, Hart et al., 2015) to measure self-deceptive enhancement (e.g., "I never regret my decisions," $\alpha = .73$) and impression management (e.g., "I sometimes tell lies if I have to [R]," $\alpha = .71$) with eight items each (1 = *totally disagree* to 8 = *totally agree*). Consistent with theory (Paulhus, 1988), these two facets of social desirability bias were moderately positively correlated ($r = .29$).

Analytic Strategy for Sexual Self-Reports

Main Analyses

Our primary approach to studying self-presentation tendencies was to compare (gender differences in) sexual motivation between the DQ group and the IS group. While the DQ group directly responded to the sensitive questions of interest (e.g., number of sexual fantasies), the IS group provided the sum of the same sensitive question and two non-sensitive questions as their response (i.e., two phone number digits, see Figure 1, Panel A). Because the responses in the IS group contain information other than the answers to the sensitive items, sample means of the responses in the DQ group and the IS group cannot be compared (see Figure 1, Panel B). However, in the IS group, mean level estimates for the sensitive questions of interest can be calculated by subtracting the expected value for the sum of two phone number digits (i.e., two times the population mean of a uniformly distributed variable with ten options [digits 0 to 9], that is, 2 x 4.5 = 9) from the joint responses (see Figure 1, Panel C).

We then used a 2(gender: female, male) x 2(questioning group: IS, DQ) between-subjects analysis of variance framework to test our hypotheses. Omnibus tests (H1, H2) were supplemented with two kinds of planned simple contrasts: calculating group differences (IS vs. DQ) for women (H1a) and for men (H1b), and calculating gender differences (men vs. women) in

the IS group (H2a) and in the DQ group (H2b), respectively (preregistered). We will also report group × gender interactions (non-preregistered).

Power: Effect-Size Sensitivity Analyses

The central drawback of the IS procedure is that the maximum privacy protection comes at the cost of introducing random noise, which reduces the statistical power. We therefore collected large samples, used the total sample of both studies, and conducted effect-size sensitivity analyses using Monte Carlo simulations. They suggest that in the long run, samples of the size collected here provide 80% power to detect group differences (i.e., DQ versus IS) as small as 0.46 events per time period (e.g., 0.46 fantasies per day).

Effect-Size Estimates

A second consequence of the variance inflation is that it leads to a systematic underestimation of (variance-based) effect sizes (e.g., Cohen's *d*) in the IS group. For example, consider a scenario in which on average, men have one more fantasy per day than women. Further assume that all individuals' responses in both the DQ and the IS group are completely accurate and unbiased. Although the gender differences in the raw metric (e.g., events per day) would be identical, standardizing these differences using the group's standard deviation would result in an underestimation of the true effect, expressed in Cohen's *d*, in the IS group because the inflated standard deviation appears in the denominator. Thus, large sample sizes alone are not a sufficient solution to this systematic underestimation of effect sizes in the IS group. We applied two measures to arrive at meaningful effect size estimates. First, we considered the raw mean difference, unaffected by variance inflation, as the primary effect size (e.g., $\hat{\mu}_{IS} - \hat{\mu}_{DQ}$). Second, for the IS group, we used the estimated variances of the critical sensitive questions alone to compute standardized differences instead of the (systematically inflated) variances of the joint responses of three questions, as detailed below. Assuming independence of the three questions

that make up the IS response, the variance of the joint response, Var(IS), is equal to the sum of the variances of the one sensitive item (s1) and the two non-sensitive digits (d1, d2):

$$Var(IS) = Var(s1 + d1 + d2) = Var(s1) + Var(d1) + Var(d2) \quad (1)$$

For uniform distributions, such as phone-number digits, the population variance is known (where n = number of digits; Walpole et al., 1993):

$$Var(d1) = Var(d2) = \frac{n^2 - 1}{12} \quad (2)$$

Rearranging equation (1) and substituting values from equation (2), the estimated true variance of the sensitive item is given by:

$$Var(s1) = Var(IS) - Var(d1) - Var(d2) = Var(IS) - 2 \times \frac{n^2 - 1}{12} \quad (3)$$

For n = 10 (i.e., ten possible outcomes with digits from 0 to 9), this is:

$$Var(s1) = Var(IS) - 2 \times \frac{10^2 - 1}{12} = Var(IS) - 2 \times 8.25 = Var(IS) - 16.5 \quad (4)$$

(also see Figure 1, Panel C).[3]

Outlier Treatment

In the DQ group, we treated values greater than 30 (e.g., sexual fantasies per day) as outliers, which very few participants reported in an independent pilot study (i.e., 0-3%). For the IS group, we preregistered two alternative outlier criteria. The primary criterion excluded values greater than 48. This guarantees that no person with true values of 30 or lower will be excluded (i.e., DQ outlier threshold + maximum value of two phone number digits = 30 + 18 = 48). Because this carries the risk of not excluding all participants with true values greater than 30, we established a stricter secondary outlier criterion that excludes participants with values greater than 39 (i.e., DQ outlier threshold + expected value of two phone number digits = 30 + 9 = 39).

[3] Before collecting the data, we were unaware of this formula and instead preregistered to roughly estimate the true variance of the sensitive item based on the variance of the corresponding items in the DQ group. The procedure employed here is superior because it uses know population parameters rather than requiring informed guessing about the relationship between the variances in the DQ and IS groups.

Results

Manipulation Check: Perceived Anonymity

As expected, perceived anonymity (1 = *not at all* to 7 = *very much*) was high in the DQ group (M = 5.70, SD = 1.65), but even higher in the IST group (M = 6.13, SD = 1.35), $t(2855) = 2.68$, $p = .007$, $d = 0.10$ [0.03, 0.17].

Main Analyses: Sexual Motivation

Replicating past findings (Baumeister et al., 2001; Frankenbach et al., 2022), two-way analyses of variance revealed significant gender differences for all three indicators of sexual motivation. There were no significant effects of group (IS vs. DQ) and no significant group x gender interactions (Table 1). Descriptive statistics and planned simple effects were used for a closer inspection. The results are illustrated in Figure 2. We found significant gender differences across both groups and all indicators of sexual motivation. Effect sizes were medium to large ($0.53 \leq d \leq 0.83$; see Lovakov & Agadullina, 2021) and of comparable size in the DQ and the IS group (see Table 2). Event frequencies were descriptively higher in the IS group across both genders and all indicators of sexual motivation. However, these group effects were small ($0.10 \leq d \leq 0.15$; see Table 3) and not significant except for male fantasies. When using the alternative outlier criterion, group differences were even smaller and all non-significant. The significance and effect sizes of the gender differences did not change depending on the outlier criterion. Results of the sensitivity analyses that are based on the alternative outlier criterion are presented in the Supplementary Online Materials (SOM, Tables S1-S3).

Table 1

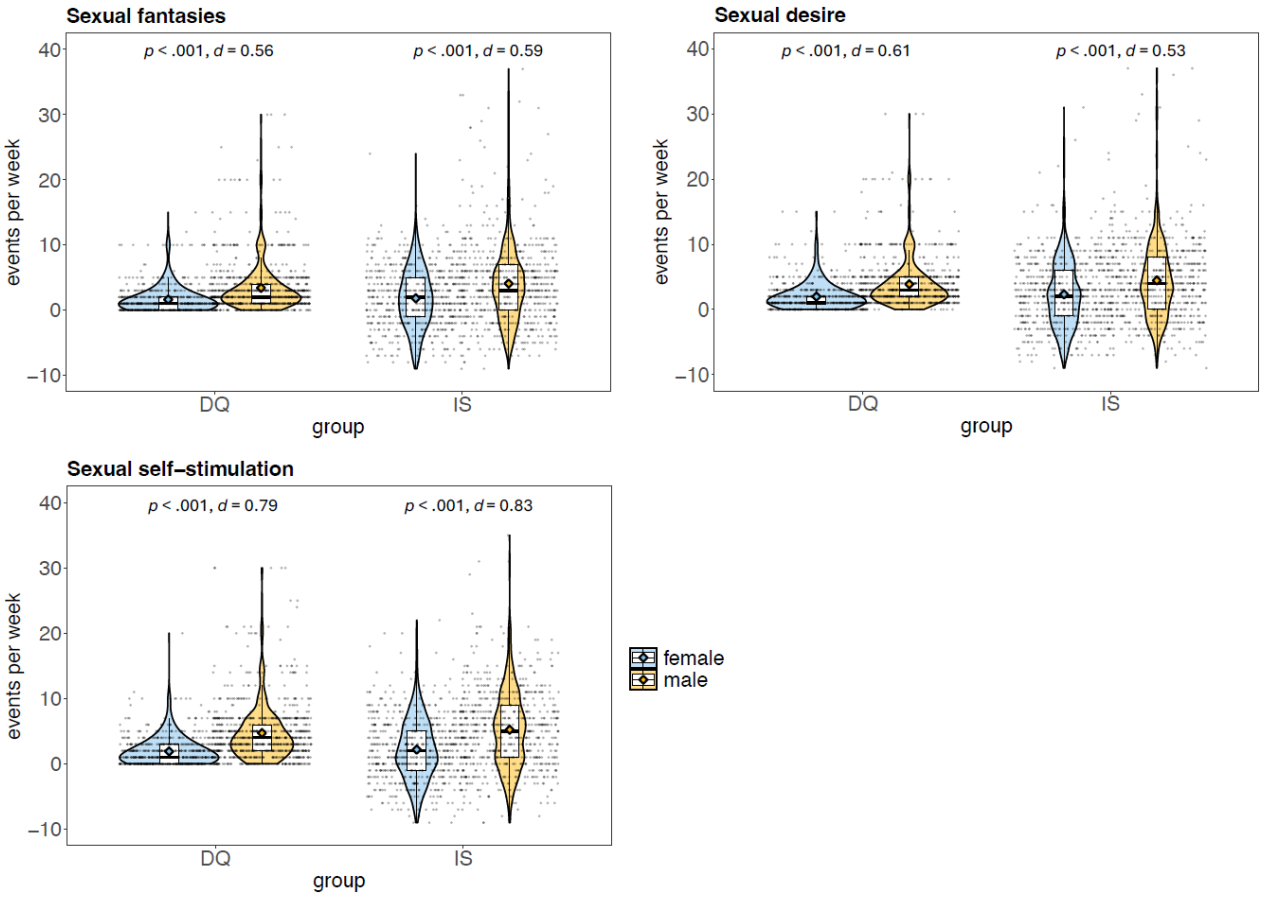
ANOVA Results: Omnibus Test

	Indicators of sexual motivation						Sexual bias indicators			
	Sexual fantasies		Sexual desire		Self-stimulation		Sexual intercourse		Sexual partners	
	F(1, df2)	*p*	*F*(1, df2)	*p*	*F*(1, df2)	*p*	*F*(1, df2)	*p*	*F*(1, df2)	*p*
gender	61.92	< .001	68.05	< .001	145.93	< .001	5.29	.022	0.43	.511
group	0.67	.414	2.07	.150	1.28	.259	0.23	.631	1.19	.276
group x gender	1.93	.165	0.40	.527	0.43	.514	6.43	.011	1.35	.245

Note. *df* = 2839 (sexual fantasies), 2833 (sexual desire), 2840 (sexual self-stimulation), 2330 (sexual intercourse), 2241 (sexual partners). The results presented for sexual intercourse and sexual partners are based on the subset of individuals who self-identify as heterosexual, as bias indicator logic applies only to mixed-gender activities.

Figure 2

Estimated Event Frequencies for the Indicators of Sexual Motivation



Note. The violin plots illustrate the distribution of the estimated event frequencies separately for each group (DQ, IS) x gender (male, female) combination. The dots represent individual values (jittered for illustration purposes), the white-colored boxes show the interquartile ranges (25th to 75th percentile), and the diamonds symbolize the sample means. The statistics presented show that we found significant gender differences for all indicators of sexual motivation that were of similar magnitude in the DQ group and the IS group. Negative estimates of event frequencies for some individuals of the IS group are a byproduct of the subtracting the population mean for the non-sensitive information (i.e., 9) which is necessary to allow for the critical mean-level comparisons of sexual motivation between the groups.

Table 2

Simple Effects: Gender Differences in the DQ Group and the IS Group

	M_{men}	SD_{men}	M_{wom}	SD_{wom}	*raw*	*df*	*t*	*p*	*d* [$CI_{95\%}$]
Sexual fantasies									
DQ	3.41	3.87	1.61	1.98	1.80	2839	7.87	< .001	0.56 [0.46, 0.67]
IST	4.08	6.35 (4.88)	1.81	4.63 (2.22)	2.26	2839	9.21	< .001	0.59 [0.48, 0.70]
Sexual desire									
DQ	3.87	3.75	1.96	2.05	1.91	2833	8.25	< .001	0.61 [0.51, 0.71]
IST	4.44	6.12 (4.58)	2.32	5.19 (3.23)	2.12	2833	8.55	< .001	0.53 [0.42, 0.64]
Sexual self-stimulation									
DQ	4.72	4.36	1.91	2.11	2.82	2840	12.08	< .001	0.79 [0.69, 0.90]
IST	5.23	6.07 (4.51)	2.19	4.77 (2.50)	3.04	2840	12.14	< .001	0.83 [0.72, 0.94]
Sexual intercourse									
DQ	1.61	1.97	1.16	1.58	0.45	2330	2.30	.098	0.25 [0.14, 0.36]
IST	0.98	4.55 (2.04)	1.26	4.93 (2.80)	-0.28	2330	-1.33	.544	-0.12 [-0.24, 0.00]
Number of sexual partners									
DQ	6.06	7.18	6.37	7.31	-0.31	2241	-0.66	.913	-0.04 [-0.16, 0.07]
IST	6.31	9.87 (9.00)	5.81	8.34 (7.28)	0.50	2241	-0.98	.763	0.06 [-0.06, 0.18]

Note. We calculated pairwise comparisons using Tukey's corrections. In the IS group, inferential statistics (*t, p*) are based on the standard deviations of the joint responses (i.e., sum of three items). For effect-size estimates that are meaningful and comparable to those in the DQ group, standardized mean differences in the IS group were calculated based on the standard deviations of the estimated true variance of the sensitive items only (reported in parenthesis here; for details on the rationale, please consult the Methods section). The results presented for sexual intercourse and sexual partners are based on the subset of individuals who self-identify as heterosexual as the bias indicator logic is only applicable for mixed-gender activities.

Table 3

Simple Effects: Group Differences for Men and Women

	M_{IS}	SD_{IS}	M_{DQ}	SD_{DQ}	*raw*	*df*	*t*	*p*	*d* [$CI_{95\%}$]
Sexual fantasies									
m	4.08	6.35 (4.88)	3.41	3.88	0.67	2839	2.90	.020	0.15 [0.05, 0.25]
f	1.81	4.63 (2.22)	1.61	1.98	0.20	2839	0.82	.846	0.10 [-0.01, 0.20]
Sexual desire									
m	4.44	6.12 (4.58)	3.87	3.75	0.57	2833	2.45	.068	0.14 [0.04, 0.24]
f	2.32	5.19 (3.23)	1.96	2.05	0.36	2833	1.44	.474	0.13 [0.02, 0.24]
Sexual self-stimulation									
m	5.23	6.07 (4.51)	4.72	4.36	0.51	2840	2.16	.136	0.11 [0.01, 0.22]
f	2.19	4.77 (2.50)	1.91	2.11	0.28	2840	1.13	.671	0.12 [0.01, 0.23]
Sexual intercourse									
m	0.98	4.55 (2.04)	1.61	1.97	-0.63	2330	-3.26	.006	-0.32 [-0.43, -0.20]
f	1.26	4.93 (2.80)	1.16	1.58	0.10	2330	0.48	.964	0.05 [-0.08, 0.17]
Number of sexual partners									
m	6.31	9.87 (9.00)	6.06	7.18	0.25	2241	0.53	.952	0.03 [-0.08, 0.14]
f	5.81	8.34 (7.28)	6.37	7.31	-0.57	2241	-1.09	.696	-0.08 [-0.20, 0.05]

Note. We calculated pairwise comparisons using Tukey's corrections. Standardized mean differences were calculated based on the observed standard deviations (DQ group) and the estimated true variance of the sensitive items only (IS group; reported in parenthesis here; for details on the rationale, please consult the Methods section). The results presented for sexual intercourse and sexual partners are based on the subset of individuals who self-identify as heterosexual as the bias indicator logic is only applicable for mixed-gender activities.

Complementary Analyses

Biased responding in online research may also be reflected in pronounced gender differences in sexual intercourse and the number of sexual partners when in reality appreciable differences are unlikely (logic approach), in low self-reported levels of honesty (subjective approach), and in associations between social desirability scales and sexual event frequencies that are positive for men and negative for women (control approach).

Sexual Bias Indicators (Logic Approach)

In stark contrast to the indicators of sexual motivation, gender differences for the sexual bias indicators (i.e., sexual intercourse, number of sexual partners) were weak to nonexistent. Significantly higher values for men than for women were only found for sexual intercourse in the total sample, but the gender difference was considerably smaller than those for sexual fantasies, desire, and self-stimulation. Follow-up analyses did not reveal significant gender differences in the DQ group or in the IS group.

Self-Reported Honesty (Subjective Approach)

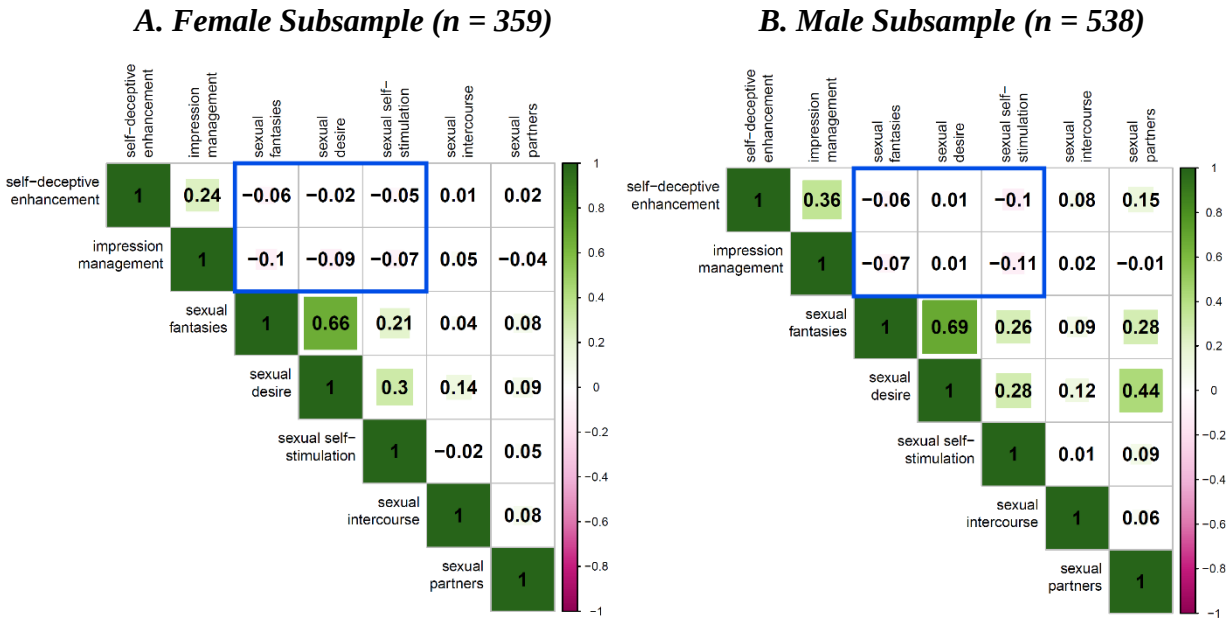
The average level of self-reported honesty approached the maximum of 7 in both the IS group (M = 6.90, SD = 0.39) and the DQ group (M = 6.86, SD = 0.45). Large sample sizes and very small standard deviations resulted in a significant, yet negligible group difference ($M_{\text{difference}}$ = 0.04, $t(2855) = 2.88$, $p = .004$, $d = 0.11$ [0.03, 0.18].

Social Desirability Scales (Control Approach, Study 2)

Figure 3 illustrates the associations between the social desirability scales and sexual self-reported sexual motivation. The correlations were very small ($-0.11 \leq r \leq 0.05$) and non-significant. Descriptively, most of the associations were negative for men and for women. Controlling for the social desirability scores did not change any of the conclusions regarding gender differences in sexual motivation (ps < .001).

Figure 3

Associations between Social Desirability Scales and Sexual Self-Reports



Note. $N_{\text{Study 2}}$ = 897. The values are bivariate correlations. Larger and darker colored squared areas indicate stronger positive (green) and negative (red) correlation. The area within the blue rectangle shows the associations between the social desirability scales and the self-reported sexual motivation.

Discussion

Valid assessment of motivation can be challenging. This is especially true the socially sensitive topic of sexual motivation, where any large-scale assessment depends on people's honest self-reporting, which cannot be taken for granted. In many situations, people may want to present themselves in a socially favorable way. This self-presentation tendency may also differ between men and women, consistent with the notion of a sexual double standard. Not knowing the extent to which outcomes are affected by social desirability bias has tangible consequences, such as uncertainty about the magnitude of gender differences in sexual motivation. Because of

the anonymity of the Internet, online data collection holds the promise of less biased responses, but the extent to which this hope is realized is unclear. In the present research, we combined four approaches to obtain a clear picture of the (un)biased nature of self-reported sexual motivation in online research. Taken together, they provide little indication that social desirability bias is a substantial threat to the validity of self-reported sexual motivation in online research. Furthermore, the combined results provide no evidence that social desirability bias is a major factor accounting for gender differences in sexual motivation.

First, we used the Item Sum Technique and compared frequency estimates for theory-driven indicators of sexual motivation (i.e., sexual fantasies, sexual desire, masturbation; Frankenbach et al., 2022) and gender differences therein between a standard direct questioning (DQ) group and an indirect item sum (IS) group. Previous use of the Item Sum Technique with topics of moderate to high sensitivity (e.g., cannabis use, undeclared work, sexual addiction) revealed somewhathigher frequency estimates for socially undesirable behaviors in the IS group in laboratory research and computer-assisted telephone interviews, demonstrating the propensity of method to reveal biased responding when present (Perri et al., 2018; Trappmann et al., 2014).

In the present research, frequency estimates for the indicators of sexual motivation were descriptively higher in the IS group, but group differences were not significant and small in magnitude. On average, men's self-reported sexual motivation was higher than women's. These effects were moderate to large and did not differ between the two groups.

The use of sexual bias indicators was as a second, logic approach to help determine the extent to which gender differences in sexual motivation are likely to reflect true differences or social desirability bias. Since population-level gender differences among heterosexual individuals should be close to zero for these bias indicators, small and inconsistent gender differences would be expected if participants' reports were truthful. This is what we found. There was no evidence

for the large and consistent (i.e., men > women) gender differences. which would support that gender differences are primarily a product of social desirability bias. Moreover, the pattern of results found here conceptually replicates a recent meta-analytic finding (Frankenbach et al., 2022).

In addition, the associations between scales assessing self-deceptive enhancement and self-presentation on the one hand and sexual self-reports on the other hand were weak. Controlling for these two factors of social desirability (Paulhus, 1988) did not change the differences between men and women. This is in contrast to previous, mostly non-online research, which has repeatedly documented associations between social desirability scales and sexual self-reports (for a review, see King, 2022). Finally, participants' honesty ratings were near the maximum not only in the IS group but also in the DQ group, further supporting the high accuracy of sexual self-reports in online research.

These findings are an important contribution to the ongoing debate about gender differences in sexual motivation. Higher average male than female self-reported sexual motivation has been reliably found across multiple indicators (Baumeister et al., 2001), countries (Lippa, 2009), and thousands of participants (Frankenbach et al., 2022). However, the validity of these findings has been questioned for several plausible reasons (Conley et al., 2011; Conley & Klein, 2022). A critical one is that different social norms for men and women may lead to gender-specific socially desirable responding (Alexander & Fisher, 2003; Conley et al., 2011; Touraille & Ågmo, 2024). At least for online studies, our findings provide little support for the assumption that socially desirable responding is a major driver of reported gender differences in sexual motivation. This is consistent with meta-analytical results of Frankenbach and colleagues (2022), who found that gender differences in sexual motivation were considerably larger than gender differences in sexual bias indicators.

A critical inconsistency that needs to be addressed is that other measures of sexual motivation (e.g., implicit measures, physiological measures, a casually valid picture story; Schultheiss et al., 2023; Touraille & Ågmo, 2024) have found no or only small gender differences. We suspect that these heterogeneous findings can be integrated into a larger conceptual framework, which is an interesting avenue for future research. One idea is that if any two people (e.g., a male and a female) are in a situation that offers them the same low (e.g., absence of stimulating cues) or high incentive value (i.e., a similarly stimulating cue such as their relationship partner), they are likely to experience similar levels of (state-level) sexual motivation (Ågmo & Laan, 2022; Schultheiss et al., 2023). However, some people (e.g., on average, men) have higher trait levels of sexual motivation, which may be expressed in a wider range of stimuli that can elicit sexual motivation, resulting in a greater frequency of events indicative of sexual motivation (Frankenbach et al., 2022; Weber et al., 2024).

Finally, the fact that we found no compelling evidence of a strong social desirability bias, even for such a sensitive topic as sexuality, might raise doubts about the appropriateness of the item sum technique approach. However, it is important to note that sensitivity is not primarily a characteristic of a topic, but rather a function of the topic and the situational context in which it is presented (Krumpal, 2013; Lee, 1993). The lower the risks and losses associated with (admitting to) a behavior in a specific context, the more likely it is that self-reports will be accurate (Krumpal, 2013; Rasinski et al., 1999). For online surveys, this means that the anonymity provided is likely to reduce participants' motivation for desirable responding to a minimum, even for very intimate topics such as sexual motivation.

Strengths, Limitations, and Future Research

Each of the methods used to prevent, test, and control for socially desirable responding that we used in the present investigation is associated with specific limitations. By design, the

item sum procedure introduces random noise. We developed a strategy to still perform high-powered tests and calculate unbiased effect sizes but estimates of event frequencies and gender differences are still likely less precise in the IS group than in the DQ group. Due to natural sampling variation, there may be (heterosexual) samples for which true gender differences in the bias indicators are different from zero. Therefore, finding more sexual partners for men than for women does not necessarily indicate biased responding. Likewise, gender differences close to zero do not necessarily indicate the absence of biased responding (e.g., samples with a higher number of sexual partners among women which disappear due to socially desirable responding). In addition, some researchers have argued that social desirability scales may not (only) measure response bias but (also) personality characteristics (Lanz et al., 2022) and have questioned the usefulness of self-reported honesty ratings (Vésteinsdóttir et al., 2019). While none of these (and other) measures are perfect indicators of social desirability bias, we view their collective assessment as a particular strength of the present research. Finding that different measures with different strengths and weaknesses arrive at a similar conclusion increases confidence in the validity of the main findings.

Another point to keep in mind when interpreting the present findings is that collecting data online was not the only aspect realized in the present study that may have worked against the manifestation of self-presentation tendencies. For instance, consistent with much other online research (Blais et al., 2023; Petsnik & Vorauer, 2020), we reminded participants of their anonymity before presenting the critical questions. This may have facilitated honest responses in addition to the survey mode alone. How these and other features interact to shape participants' self-reporting was not a goal of the present research and may be something that future research could look at more closely. As a cautious interim conclusion, we may venture to say that online

data collection, combined with simple measures to further enhance the trustworthiness of the study, is likely to largely buffer against potential tendencies toward sexual self-presentation.

Conclusion

Self-report measures are an indispensable part of the methodological toolbox for the scientific study of sexual motivation, but concerns about social desirability bias are pervasive. We combined four approaches, including an application of the item sum technique, to examine social desirability bias in online surveys. We found little evidence that the validity of self-reported sexual motivation in online studies is significantly reduced by social desirability bias. In particular, our results do not support the hypothesis that social desirability bias is a major driver of gender differences in sexual motivation.

References

Ågmo, A., & Laan, E. (2022). Sexual incentive motivation, sexual behavior, and general arousal: Do rats and humans tell the same story? *Neuroscience & Biobehavioral Reviews*, *135*, Article 104595. https://doi.org/10.1016/j.neubiorev.2022.104595

Alexander, M. G., & Fisher, T. D. (2003). Truth and consequences: Using the bogus pipeline to examine sex differences in self-reported sexuality. *Journal of Sex Research*, *40*(1), 27–35. https://doi.org/10.1080/00224490309552164

Baumeister, R. F., Catanese, K. R., & Vohs, K. D. (2001). Is there a gender difference in strength of sex drive? Theoretical views, conceptual distinctions, and a review of relevant evidence. *Personality and Social Psychology Review*, *5*(3), 242–273. https://doi.org/10.1207/S15327957PSPR0503_5

Blais, J., Fazaa, G. R., & Mungall, L. R. (2023). A pre-registered examination of the relationship between psychopathy, boredom-proneness, and university-level cheating. *Psychological Reports*, 00332941231184385. https://doi.org/10.1177/00332941231184385

Buchanan, T. (2000). *Potential of the Internet for personality research* (M. H. Birnbaum, Ed.; pp. 121–140). Academic Press. https://westminsterresearch.westminster.ac.uk/item/94427/potential-of-the-internet-for-personality-research

Carlbring, P., Brunt, S., Bohman, S., Austin, D., Richards, J., Öst, L.-G., & Andersson, G. (2007). Internet vs. Paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. *Computers in Human Behavior*, *23*(3), 1421–1434. https://doi.org/10.1016/j.chb.2005.05.002

Chaudhuri, A., & Christofides, T. C. (2007). Item count technique in estimating the proportion of people with a sensitive feature. *Journal of Statistical Planning and Inference, 137*(2), 589–593. https://doi.org/10.1016/j.jspi.2006.01.004

Conley, T. D., & Klein, V. (2022). Women get worse sex: A confound in the explanation of gender differences in sexuality. *Perspectives on Psychological Science*, 1–19.

Conley, T. D., Moors, A. C., Matsick, J. L., Ziegler, A., & Valentine, B. A. (2011). Women, men, and the bedroom: Methodological and conceptual insights that narrow, reframe, and eliminate gender differences in sexuality. *Current Directions in Psychological Science, 20*(5), 296–300. https://doi.org/10.1177/0963721411418467

Eastwick, P. W., Joel, S., Carswell, K. L., Molden, D. C., Finkel, E. J., & Blozis, S. A. (2023). Predicting romantic interest during early relationship development: A preregistered investigation using machine learning. *European Journal of Personality, 37*(3), 276–312. https://doi.org/10.1177/08902070221085877

Ehler, I., Wolter, F., & Junkermann, J. (2021). Sensitive questions in surveys: A comprehensive meta-analysis of experimental survey studies on the performance of the item count technique. *Public Opinion Quarterly*, *85*(1), 6–27. https://doi.org/10.1093/poq/nfab002

Evans, J. R., & Mathur, A. (2018). The value of online surveys: A look back and a look ahead. *Internet Research*, *28*(4), 854–887. https://doi.org/10.1108/IntR-03-2018-0089

Fisher, T. D. (2013). Gender roles and pressure to be truthful: The bogus pipeline modifies gender differences in sexual but not non-sexual behavior. *Sex Roles, 68*(7–8), 401–414. https://doi.org/10.1007/s11199-013-0266-3

Fogarty, T. J., Jonas, G. A., & Parker, L. M. (2013). The medium is the message: Comparing paper-based and web-based course evaluation modalities. *Journal of Accounting Education, 31*(2), 177–193. https://doi.org/10.1016/j.jaccedu.2013.03.002

Frankenbach, J., Weber, M., Loschelder, D. D., Kilger, H., & Friese, M. (2022). Sex drive: Theoretical conceptualization and meta-analytic review of gender differences. *Psychological Bulletin*, *148*(9–10), 621–661. https://doi.org/10.1037/bul0000366

Gnambs, T., & Kaspar, K. (2015). Disclosure of sensitive behaviors across self-administered survey modes: A meta-analysis. *Behavior Research Methods*, *47*(4), 1237–1259. https://doi.org/10.3758/s13428-014-0533-4

Gnambs, T., & Kaspar, K. (2017). Socially desirable responding in web-based questionnaires: A meta-analytic review of the candor hypothesis. *Assessment*, *24*(6), 746–762. https://doi.org/10.1177/1073191115624547

Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The Balanced Inventory of Desirable Responding Short Form (BIDR-16). *SAGE Open*, *5*(4), 215824401562111. https://doi.org/10.1177/2158244015621113

Hinzmann, J., Khalaidovski, K., Kullmann, J. S., Brummer, J., Braun, S., Kurz, P., Wagner, K. M., & Schultheiss, O. C. (2023). Developing a causally valid picture-story measure of sexual motivation: I Effects of priming. *Motivation Science*. https://doi.org/10.1037/mot0000300.supp

Kim, J. J., Muise, A., Barranti, M., Mark, K. P., Rosen, N. O., Harasymchuk, C., & Impett, E. (2021). Are couples more satisfied when they match in sexual desire? New insights from response surface analyses. *Social Psychological and Personality Science*, *12*(4), 487–496. https://doi.org/10.1177/1948550620926770

King, B. M. (2022). The influence of social desirability on sexual behavior surveys: A review. *Archives of Sexual Behavior*, *51*(3), 1495–1501. https://doi.org/10.1007/s10508-021-02197-0

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, *47*(4), 2025–2047. https://doi.org/10.1007/s11135-011-9640-9

Lanz, L., Thielmann, I., & Gerpott, F. H. (2022). Are social desirability scales desirable? A meta-analytic test of the validity of social desirability scales in the context of prosocial behavior. *Journal of Personality*, *90*(2), 203–221. https://doi.org/10.1111/jopy.12662

Lee, R. M. (1993). The problem of researching sensitive topics: An overview and introduction. *The Problems of Researching Sensitive Topics*.

Leiner, D. J. (2022). *SoSci Survey (Version 3.2.31) [Computer software]*. Available at https://www.soscisurvey.de

Li, J., & Van den Noortgate, W. (2022). A meta-analysis of the relative effectiveness of the item count technique compared to direct questioning. *Sociological Methods & Research*, *51*(2), 760–799. https://doi.org/10.1177/0049124119882468

Lippa, R. A. (2009). Sex differences in sex drive, sociosexuality, and height across 53 nations: Testing evolutionary and social structural theories. *Archives of Sexual Behavior*, *38*(5), 631–651. https://doi.org/10.1007/s10508-007-9242-8

Lovakov, A., & Agadullina, E. R. (2021). Empirically derived guidelines for effect size interpretation in social psychology. *European Journal of Social Psychology*, *51*(3), 485–504. https://doi.org/10.1002/ejsp.2752

Mark, K. P. (2015). Sexual desire discrepancy. *Current Sexual Health Reports*, *7*(3), 198–202. https://doi.org/10.1007/s11930-015-0057-7

Marks, M. J., & Fraley, R. C. (2005). The sexual double standard: Fact or fiction? *Sex Roles*, *52*(3), 175–186. https://doi.org/10.1007/s11199-005-1293-5

Meston, C. M., Heiman, J. R., Trapnell, P. D., & Paulhus, D. L. (1998). Socially desirable responding and sexuality self-reports. *The Journal of Sex Research, 35*(2), 148–157. https://doi.org/10.1080/00224499809551928

Paulhus, D. L. (1988). Balanced inventory of desirable responding (BIDR). *Acceptance and Commitment Therapy. Measures Package*, *41*, 79586–79587.

Perri, P. F., Cobo Rodríguez, B., & Rueda García, M. D. M. (2018). A mixed-mode sensitive research on cannabis use and sexual addiction: Improving self-reporting by means of indirect questioning techniques. *Quality & Quantity*, *52*(4), 1593–1611. https://doi.org/10.1007/s11135-017-0537-0

Petsnik, C., & Vorauer, J. D. (2020). Do dominant group members have different emotional responses to observing dominant-on-dominant versus dominant-on-disadvantaged ostracism? Some evidence for heightened reactivity to potentially discriminatory ingroup behavior. *PLoS ONE, 15*(6), e0234540–e0234540. https://doi.org/10.1371/journal.pone.0234540

R Core Team. (2023). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. https://www.R-project.org/

Rasinski, K. A., Willis, G. B., Baldwin, A. K., Yeh, W., & Lee, L. (1999). Methods of data collection, perceptions of risks and losses, and motivation to give truthful answers to sensitive survey questions. *Applied Cognitive Psychology*, *13*(5), 465–484. https://doi.org/10.1002/(SICI)1099-0720(199910)13:5<465::AID-ACP609>3.0.CO;2-Y

Risko, E. F., Quilty, L. C., & Oakman, J. M. (2006). Socially desirable responding on the web: Investigating the candor hypothesis. *Journal of Personality Assessment*, *87*(3), 269–276. https://doi.org/10.1207/s15327752jpa8703_08

Schmitt, D. P., Jonason, P. K., Byerley, G. J., Flores, S. D., Illbeck, B. E., O'Leary, K. N., & Qudrat, A. (2012). A reexamination of sex differences in sexuality: New studies reveal old truths. *Current Directions in Psychological Science*, *21*(2), 135–139. https://doi.org/10.1177/0963721412436808

Schultheiss, O. C., Hinzmann, J., Bergmann, S., Matthes, M., Beyer, B., & Janson, K. T. (2023). Developing a causally valid picture-story measure of sexual motivation: II Effects of film clips. *Motivation Science*, *9*(4), 272–287. https://doi.org/10.1037/mot0000301.supp

Simmons, J., P., Nelson, L. D., & Simonsohn, U. (2012). *A 21 word solution*. Retrieved from http://dx.doi.org/10.2139/ssrn.2160588

Touraille, P., & Ågmo, A. (2024). Sex differences in sexual motivation in humans and other mammals: The role of conscious and unconscious processes. *Behavioral Sciences*, *14*(4), 277. https://doi.org/10.3390/bs14040277

Trappmann, M., Krumpal, I., Kirchner, A., & Jann, B. (2014). Item sum: A new technique for asking quantitative sensitive questions. *Journal of Survey Statistics and Methodology*, *2*(1), 58–77. https://doi.org/10.1093/jssam/smt019

Vésteinsdóttir, V., Joinson, A., Reips, U.-D., Danielsdottir, H. B., Thorarinsdottir, E. A., & Thorsdottir, F. (2019). Questions on honest responding. *Behavior Research Methods*, *51*(2), 811–825. https://doi.org/10.3758/s13428-018-1121-9

Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (1993). *Probability and statistics for engineers and scientists* (Vol. 5). Macmillan New York.

Weber, M., Reis, D., & Friese, M. (2024). Development and validation of the trait sexual motivation scale (TSMS). *Journal of Personality Assessment*, *106*(2), 267–282. https://doi.org/10.1080/00223891.2023.2206896

Supplementary Online Materials (SOM)

Purpose

We preregistered the use of an alternative outlier removal criterion to test the robustness of our results. The results are presented in the supplementary online materials.

Table of Contents

Table S1

Sensitivity Analyses (Alternative Outlier Criterion): ANOVA Results

	Indicators of sexual motivation						Bias indicators			
	Sexual fantasies		Sexual desire		Self-stimulation		Sexual intercourse		Sexual partners	
	F(1, df2)	*p*	*F*(1, df2)	*p*	*F*(1, df2)	*p*	*F*(1, df2)	*p*	*F*(1, df2)	*p*
Gender	67.89	< .001	76.18	< .001	151.55	< .001	5.65	.018	0.50	.480
Group	0.73	.392	1.78	.183	1.32	.250	0.01	.939	2.79	.095
group x gender	0.40	.527	0.00	.968	0.09	.766	4.81	.028	0.00	.972

Note. *df* = 2833 (sexual fantasies), 2826 (sexual desire), 2837 (sexual self-stimulation), 2328 (sexual intercourse), 2218 (sexual partners). The results presented for sexual intercourse and sexual partners are based on the subset of individuals who self-identify as heterosexual as the bias indicator logic is only applicable for mixed-gender activities.

Table S2

Sensitivity Analyses (Alternative Outlier Criterion): Simple Effects (Gender Differences)

	M_{men}	SD_{men}	M_{wom}	SD_{wom}	*raw*	*df*	*t*	*p*	*d* [$CI_{95\%}$]
Sexual fantasies									
DQ	3.41	3.88	1.61	1.98	1.80	2833	8.24	< .001	0.56 [0.46, 0.67]
IST	3.81	5.75 (4.06)	1.81	4.63 (2.22)	2.00	2833	8.51	< .001	0.61 [0.50, 0.72]
Sexual desire									
DQ	3.87	3.75	1.96	2.05	1.91	2826	8.73	< .001	0.61 [0.51, 0.71]
IST	4.17	5.44 (3.62)	2.28	5.07 (3.03)	1.89	2826	8.06	< .001	0.57 [0.46, 0.68]
Sexual self-stimulation									
DQ	4.72	4.36	1.91	2.11	2.82	2837	12.31	< .001	0.79 [0.69, 0.90]
IST	5.11	5.80 (4.14)	2.19	4.77 (2.50)	2.92	2837	11.87	< .001	0.85 [0.74, 0.96]
Sexual intercourse									
DQ	1.61	1.97	1.16	1.58	0.45	2328	2.38	.082	0.25 [0.14, 0.36]
IST	0.98	4.55 (2.04)	1.14	4.56 (2.08)	-0.16	2328	-0.80	.857	-0.08 [-0.20, 0.04]
Number of sexual partners									
DQ	6.06	7.18	6.37	7.31	-0.31	2218	-0.71	.895	-0.04 [-0.16, 0.07]
IST	5.22	8.22 (7.15)	5.56	7.92 (6.80)	0.34	2218	-0.70	.899	-0.05 [-0.17, 0.08]

Note. We calculated pairwise comparisons using Tukey's corrections. In the IS group, inferential statistics (*t, p*) are based on the standard deviations of the joint responses (i.e., sum of three items). For effect-size estimates that are meaningful and comparable to those in the DQ group, standardized mean differences in the IS group were calculated based on the standard deviations of the estimated true variance of the sensitive items only (reported in parenthesis here; for details on the rationale, please consult the Methods section). The results presented for sexual intercourse and sexual partners are based on the subset of individuals who self-identify as heterosexual as the bias indicator logic is only applicable for mixed-gender activities.

Table S3

Sensitivity Analyses (Alternative Outlier Criterion): Simple Effects (Group Differences)

	M_{IS}	SD_{IS}	M_{DQ}	SD_{DQ}	*raw*	*df*	*t*	*p*	*d* [$CI_{95\%}$]
Sexual fantasies									
m	3.81	5.75 (4.06)	3.41	3.88	0.40	2833	1.83	.258	0.10 [0.00, 0.20]
f	1.81	4.63 (2.22)	1.61	1.98	0.20	2833	0.86	.827	0.10 [-0.01, 0.20]
Sexual desire									
m	4.17	5.44 (3.62)	3.87	3.75	0.30	2826	1.35	.529	0.08 [-0.02, 0.18]
f	2.28	5.07 (3.03)	1.96	2.05	0.31	2826	1.33	.542	0.12 [0.01, 0.23]
Sexual self-stimulation									
m	5.11	5.80 (4.14)	4.72	4.36	0.38	2837	1.66	.346	0.09 [-0.01, 0.19]
f	2.19	4.77 (2.50)	1.91	2.11	0.28	2837	1.15	.658	0.12 [0.01, 0.23]
Sexual intercourse									
m	0.98	4.55 (2.04)	1.61	1.97	-0.63	2328	-3.37	.004	-0.32 [-0.43, -0.20]
f	1.14	4.56 (2.08)	1.16	1.58	0.02	2328	0.08	1.00	-0.01 [-0.13, 0.11]
Number of sexual partners									
m	5.22	8.22 (7.15)	6.06	7.18	-0.83	2218	-1.90	.230	0.03 [-0.08, 0.14]
f	5.56	7.92 (6.80)	6.37	7.31	-0.81	2218	-1.67	.339	-0.08 [-0.20, 0.05]

Note. We calculated pairwise comparisons using Tukey's corrections. Standardized mean differences were calculated based on the observed standard deviations (DQ group) and the estimated true variance of the sensitive items only (IS group; reported in parenthesis here; for details on the rationale, please consult the Methods section). The results presented for sexual intercourse and sexual partners are based on the subset of individuals who self-identify as heterosexual as the bias indicator logic is only applicable for mixed-gender activities.

PART III

Sexual (Double) Standards Revisited:

Similarities and Differences in the Societal Evaluation of Male and Female Sexuality

Marcel Weber

Saarland University

Malte Friese

Saarland University

Manuscript accepted for publication in *Social Psychological and Personality Science* (in press).

This is an Accepted Manuscript of an article accepted for publication in

Social Psychological and Personality Science, first published online on April 11, 2024,

available at https://doi.org/10.1177/19485506241237288

Sexual (Double) Standards Revisited:

Similarities and Differences in the Societal Evaluation of Male and Female Sexuality

Marcel Weber[1] and Malte Friese[1]

[1]Department of Psychology, Saarland University

Author Note

Marcel Weber: https://orcid.org/0000-0002-0507-0070

Malte Friese: https://orcid.org/0000-0003-0055-513X

Corresponding Author

Correspondence concerning this article should be addressed to Marcel Weber, Department of Psychology, Saarland University, Campus A2 4, 66123 Saarbrücken, Germany. E-mail: marcel.weber@uni-saarland.de.

Abstract

Past research has been inconclusive regarding the continued existence of the sexual double standard (SDS)—that is, differential expectations and evaluations of sexual activity for men (rewarded for sexual activity) and women (punished for sexual activity). Here, we present the similarities and differences (S&D) model of sexual standards, which significantly qualifies the traditional SDS by highlighting both similarities and differences between standards applied to women and men. Across two samples (student/community sample, crowdsourcing sample; N_{total} = 342) and seven sexual outcomes, high sexual activity was rated more favorably in men than in women (replicating previous research), and the opposite was true for low sexual activity (extending previous research). Importantly, moderate (not extremely low or high) sexual activity was rated most favorably in both genders, suggesting similar and curvilinear intragender trajectories. These findings illustrate a distinctly different perspective on male and female sexuality and open avenues for new research.

Keywords: sexual double standard, sexual norms, sexuality

Sexual (Double) Standards Revisited:

Similarities and Differences in the Societal Evaluation of Female and Male Sexuality

To what extent do men and women face different versus similar sexual norms? The seminal sexual double standard (SDS) proposes that evaluations of sexual activities depend on gender.[1] Under the umbrella of the SDS, multiple assumptions about the gendered nature of sexual norms have been summarized. First, a traditional SDS should manifest in more positive evaluations of (a) sexually active men and of (b) sexually *in*active women compared to an equally (in)active individual of the opposite gender (Crawford & Popp, 2003; Endendijk et al., 2020). Second, the SDS has repeatedly been introduced as (c) men being socially rewarded for sexual activity and (d) women being socially punished for sexual activity (Marks & Fraley, 2005; Wesche et al., 2021). Together, these *inter*gender effects (i.e., same level of sexual activity, different genders) and *intra*gender effects (i.e., same gender, different levels of sexual activity) would cumulate in the crossover pattern of a strong sexual double standard (Figure 1A, adapted from Marks & Fraley, 2005), suggesting that male and female sexual norms could hardly be more different.

Empirically, a meta-analysis found that, "[f]or men, frequent sexual activity was more expected, and evaluated more positively, than for women" (d = 0.25; Endendijk et al., 2020, p. 163). These *inter*gender differences were stronger in studies that operationalized the SDS as differences in societal expectations and respondents' perceived societal evaluations than as differences in respondents' personal evaluations. However, a more positive evaluation of high male than high female sexual activity is not equivalent to men being increasingly socially rewarded and women being increasingly socially punished for higher levels of sexual activity.

[1] The SDS does not specify expectations regarding nonbinary persons. We therefore follow previous research and focus on male and female targets.

The few studies that included multiple levels of sexual activity inconsistently found that both men and women were evaluated less favorably as the level of sexual activity increased, but these *intra*gender effects were heterogeneous across outcomes and studies, with higher sexual activity sometimes being evaluated more, equally, and less favorably than lower sexual activity (Marks et al., 2019; Marks & Fraley, 2005, 2007). In addition, little research exists comparing how men and women are evaluated for low sexual activity (Endendijk et al., 2020). To date, there is uncertainty about the (continued) existence of the SDS, with the tentative conclusion being that there may be a rather weak SDS for some types of sexual behavior, while for others there is a single standard for men and women (Bordini & Sperb, 2013; Endendijk et al., 2020; see Figure 1B-C, adapted from Marks & Fraley, 2005).

Similarities and Differences: The S&D Model

Here, we present the similarities and differences model (*S&D model*) of sexual (double) standards, which offers a critically different perspective on male and female sexual norms. The model is based on two central assumptions. First, the S&D model proposes that the associations between levels of sexual activity and evaluations are curvilinear (i.e., nonmonotonic) rather than linear (i.e., monotonic). According to the model, increasing sexual activity is increasingly socially expected and positively evaluated up to a certain maximum point: the ideal level of sexual activity (ILSA). Sexual activity and permissiveness above the ILSA are no longer viewed as positively, but are increasingly devalued. This curvilinearity along the continuum of very low to very high sexual activity is expected to occur for both target genders, suggesting a striking similarity between male and female sexual norms.

Importantly, the prediction of similar and curvilinear *intra*gender associations between sexual activity and evaluations does not imply the absence of differences. Instead, the second assumption of the S&D model is that the locations of the predicted curves differ between the

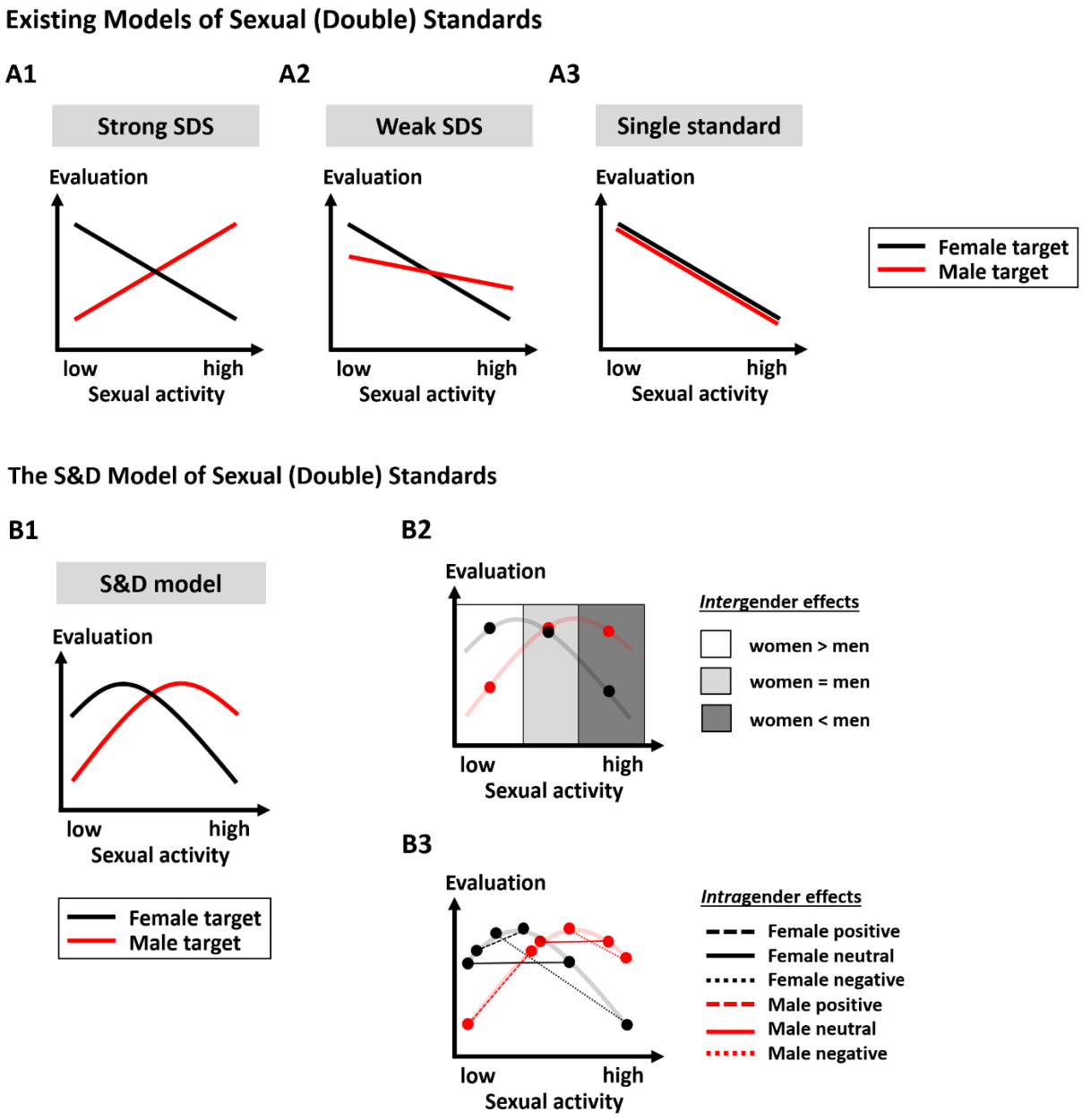
genders. Specifically, we predict that for a range of sexual activities, the ILSA will be located at higher levels of sexual activity for men compared to women. In other words, the S&D model predicts that the devaluation of sexuality will set in earlier for women. This assumption thus emphasizes the differences in sexual norms applied to the genders.

Figure 1 shows a schematic representation of the S&D model (Panel B1), which proposes the coexistence of (*inter*gender) differences and (*intra*gender) similarities. The model may explain part of the heterogeneity in previous research: Depending on the level(s) of sexual activity examined in a study, *inter*gender comparisons for a given level may indicate more favorable evaluations of female targets (area colored white), similar evaluations of female and male targets (area colored light gray), or more favorable evaluations of male targets (area colored dark gray; Panel B2), and (linear) *intra*gender effects may be positive (dashed lines), neutral (solid lines), or negative (dotted lines; Panel B3).

Preliminary qualitative and quantitative evidence supports the S&D model. First, in an interview study, single women reported feeling torn between appearing too prude and too permissive (Pickens & Braun, 2018), both of which were perceived as undesirable extremes compared to more moderate levels of sexual activity. Second, a survey by the medical service ZAVA found that male and female targets who had never had sex and those who had multiple sexual partners were considered unattractive. Between these more extreme levels of sexual activity, there was an ideal zone of the number of sexual partners, which varied according to the age and gender of the target person (Zava, 2021).

Figure 1

Sexual (Double) Standards: Previous Models (A) and the Novel S&D Model (B)



Note. The *inter*gender and *intra*gender assumptions summarized under the SDS umbrella would cumulate into a strong SDS (Panel A1). Previous research has been more consistent with a weak SDS for some types of sexual behavior (Panel A2) and a sexual single standard for other behaviors (Panel A3), but the results have been heterogeneous. The S&D model presented here posits that male and female sexual norms are defined equally by *inter*gender differences and

*intra*gender similarities (Panel B1). The model may explain some of the heterogeneity in previous research by predicting that *inter*gender effects (Panel B2) and *intra*gender effects (Panel B3) will vary in magnitude and direction depending on the level(s) of sexual activity examined in the study.

Theoretically, the S&D model can be derived from the same theoretical accounts used to explain differences between male and female sexual norms. From an evolutionary perspective, society is likely to reward levels of sexual (in)activity that maximize reproductive success (Zaikman & Marks, 2017). Because unwanted pregnancy is associated with greater risk for women than for men due to women's greater minimum parental investment and lower reproductive capacity (Trivers, 1972), sexually permissive behavior (e.g., having numerous sexual partners) is likely to be more negatively evaluated for women than for men (i.e., *inter*gender difference). However, this does not mean that women who are completely sexually absent (and therefore not considered a suitable mating option) or men who have hundreds of sexual partners (e.g., risk of contracting STDs, impaired long-term mate value; Buss & Schmitt, 1993) have maximum success (and therefore the best evaluations). Instead, for both genders, the most positive evaluations may occur at moderate rather than very low or high levels of sexual activity (i.e., *intra*gender similarity).

Social role theory posits that biological differences between the genders have led to the division of labor, which in turn has led to different gender role expectations for men and women (Eagly & Wood, 2012; Wood & Eagly, 2002). High sexual permissiveness is more consistent with the agentic role assigned to men (e.g., being active and dominant) than with the communal role typically assigned to women (e.g., being passive and caring). Through the mechanisms of direct and indirect learning specified in cognitive social learning theory (Bandura, 1986), people

internalize that role-consistent sexual behaviors are socially rewarded and that role-inconsistent behaviors are socially punished, which explains the differences between male and female sexual norms (Zaikman & Marks, 2017). This does not mean, however, that (people learn that) maximum sexual restraint on the part of women or maximum sexual permissiveness on the part of men is socially expected and rewarded. Stigmas surrounding female virginity (Gesselman et al., 2017) and the predominant use of negative terminology to describe both male and (even more so) female individuals with multiple sexual partners (Milhausen & Herold, 2002) are consistent with the coexistence of similarities (i.e., moderate levels of sexual activity are evaluated most favorably) and differences (e.g., high levels of sexual activity are evaluated more negatively for women) between male and female sexual norms.

The Present Research

In the present research, we empirically tested the coexistence of the (*intra*gender) similarities and (*inter*gender) differences specified in the S&D model proposed here across two samples and seven types of sexual activity. We focused on perceived societal evaluations (i.e., participants' perceptions of how society would view people) rather than actual personal evaluations to maximize the internal validity (e.g., personal evaluations are more likely to be influenced by social desirability bias) and external validity of our research (e.g., people's feelings and behaviors are more likely to be influenced by their perceptions of the world than by the world as it is). Confirmatory *inter*gender predictions were that high sexual activity will be evaluated more favorably for men than for women (*H1*, empirical replication) and that low sexual activity will be evaluated more favorably for women than for men (*H2*, empirical extension). Confirmatory *intra*gender predictions were that within the same gender, nonmonotonic, curvilinear, rather than monotonic, linear trajectories will best describe the data for men (*H3*) and women (*H4*)—a similarity in *intra*gender trajectories across the sexual activity continuum that is

uniquely predicted by the S&D model. We also tested whether the ideal level of sexual activity is higher for men than for women (*H5*). Finally, we examined a potential asymmetry: whether sexual activity above the ILSA is punished more severely for women than for men (*H6a*), and whether sexual activity below the ILSA is punished more severely for men than for women (*H6b*).

Method

Transparency and Openness

All study materials (i.e., online questionnaires, codebooks), scripts, and data are openly available on the Open Science Framework (https://osf.io/wyknx/). We preregistered our research goals, hypotheses, exclusion criteria, and analytic strategies for one sample and applied the preregistered operations to both samples used in the present research. We transparently report non-preregistered analyses and deviations from the preregistered plan. The project was approved by the Institutional Review Board of Saarland University. We used R, version 4.2.1, to analyze the data (R Core Team, 2023).

Samples and Power Considerations

We collected data from young adults who reported how they thought society would view a 25-year-old target person who was roughly their age to ensure that the evolutionary, social role, and cognitive social learning mechanisms that may underlie sexual (double) standards were in place (i.e., target of reproductive age, participants familiar with their social roles and standards). In light of discussions about the advantages and disadvantages of student samples and online crowdsourcing samples (Goodman et al., 2013) and about the replicability of psychological research (Open Science Collaboration, 2015), we tested our hypotheses across two samples.

Sample 1 was a mixed student and community sample recruited through flyers on a medium-sized German university and social media (n = 190; age: M = 22.04, SD = 2.97, range:

19-41; 37.9% male, 61.6% female, 0.5% nonbinary). Participants were rewarded with partial course credit or entry into a €50 gift card lottery. Sample 2 was a crowdsourcing sample of German residents recruited through the platform Clickworker (n = 170; age: M = 30.34, SD = 5.81, range: 19-40; 51.8% male, 48.2% female). Participants were paid €4.40 (i.e., €10.55/hr, equivalent to U.S. $5.20 and $12.50/hr at the time the study was launched). We excluded 14 participants from Sample 1 and four participants from Sample 2 because they failed an attention check or because they indicated that the quality of their data was comprised after completing the survey (preregistered exclusion criteria).

All data were collected online using SoSci survey (Leiner, 2022). A priori power analyses suggested that when assuming a correlation of r = .5 between evaluations of male and female targets, a sample size of N = 101 (139) participants would be required to replicate the meta-analytic *inter*gender effect (d = 0.25; Endendijk et al., 2020) with 80% (90%) power. Following recent recommendations, we report effect-size sensitivity analyses for all confirmatory analyses (i.e., H1-H4) in the Results section (Giner-Sorolla et al., 2019).

Design, Procedure, and Measures

Participants read scenarios of fictional male and female targets who exhibited one of seven different levels of low to high sexual activity across seven different sexual outcomes (presented in a fixed order, see Table 1). The sexual outcomes were sexual behaviors used in previous SDS research, plus sexual cognition and sexual affect. To denote the low and high ends of the sexual activity continuum, we used natural null points (e.g., 0 sexual partners), operationalizations used in previous research (e.g., 12 sexual partners; Zaikman et al., 2016), and evidence-based ranges of typical event frequencies (e.g., 6 desires/day; Weber et al., 2024).

Table 1

Sexual Outcomes and Levels of Sexual Activity

Sexual outcomes	Metric	Level1	Level2	Level3	Level4	Level5	Level6	Level7
Sex partners	Absolute number in life	0 (0)	1 (2)	2 (4)	3 (6)	4 (8)	5 (10)	6 (12)
Casual sex partners	Absolute number in life	0	1	2	3	4	5	6
Sexual debut	Age in years	24	22	20	18	16	14	12
Sexual intercourse	Frequency in a typical week	0	1	2	3	4	5	6
Masturbation	Frequency in a typical week	0	1	2	3	4	5	6
Sexual desire	Frequency on a typical day	0	1	2	3	4	5	6
Sexual fantasies	Frequency on a typical day	0	1	2	3	4	5	6

Note. During data collection in the community sample, some participants indicated that they felt that six sexual partners was too low of an upper limit. Therefore, we increased the range in the crowdsourcing sample (in parentheses).

In this within-participant experimental design, participants indicated how society would view the target on a 13-point rating scale ("How do you think society would view a young man/woman [25 years old] who..."; -6 = *very negatively* to 6 = *very positively*) for each of these 98 scenarios (i.e., 2 [target genders] x 7 [activity levels] x 7 [sexual outcomes]). For simplicity, we refer to these perceived societal evaluations as "evaluations" below. Participants provided all 14 evaluations (2 target genders and 7 activity levels) per sexual outcome in the same visual display. The slider bars were preset to the center of the scale (i.e., value 0) and could be adjusted by dragging or clicking on the marker for the respective gender. To distinguish between intended and unintended neutral responses, participants received a warning message if they did not move the sliders. Figure 2 illustrates this procedure for the sexual outcome "casual sex partners".

In the final part of the survey, we collected demographic information (e.g., gender: "What is your gender?", 1 = female, 2 = male, 3 = other; age: "How old are you?") and self-reported data quality ("Data quality can be compromised for a variety of reasons (e.g., rushed, distracted, not focused, not honest). Please indicate your self-perceived data quality," 1 = my data is okay, 2 = I am not sure if my data is okay, 3 = I am sure that my data is not okay).

Figure 2

Perceived Societal Evaluations of Sexual Outcomes: Response Format

Note. Participants evaluated male and female targets for seven levels of sexual activity in the same visual display. This example for "casual sex partners" shows selected values for the first two levels of sexual activity.

Data-Analytic Strategy and Presentation of Results

To test H1 and H2, we computed a 2 (target gender: male vs. female) x 2 (activity level: lowest vs. highest) within-subjects ANOVA on the evaluations for each of the seven outcomes. We were primarily interested in the interaction and the simple contrasts between target genders for low (H1) and high (H2) sexual activity, respectively. These analyses encompass those used in

previous research comparing evaluations of male and female targets with high levels of sexual activity were compared (Endendijk et al., 2020).

The within-gender hypotheses (i.e., different activity levels, same gender; H3, H4) were tested using a multilevel framework (i.e., seven levels of sexual activity nested within participants). We centered the factor level at its grand mean before specifying multilevel orthogonal polynomial contrasts separately for male and female targets. We included linear and quadratic terms and allowed intercepts to vary randomly across participants. Superiority of the S&D model over the SDS was inferred when the quadratic terms were significantly negative and the trajectories were nonmonotonic (i.e., an inverted U-shaped pattern).

To examine whether devaluation sets in earlier for female targets than for male targets, we compared the level(s) of sexual activity most favorably evaluated using paired *t*-tests (H5). Finally, we examined whether devaluation for exceeding the ILSA was more pronounced for women than for men (H6a), whereas devaluation for falling below the ILSA was more pronounced for men than for women (H6b). Using reduced data sets containing only activity levels at and (a) above or (b) below the ILSA, we examined the target gender × level interactions within a multilevel framework (random intercept, fixed slope; person-mean centered to examine pure within-person effects).

All inferential statistics are supplemented with effect size measures; qualitative descriptions of effect sizes (e.g., “small”, “medium”) follow recent suggestions by Field (2013) and Funder and Ozer (2019). Our analyses of Sample 1 and Sample 2 yielded highly comparable results. Including participant gender and its interactions with the focal predictors in our confirmatory analyses did not change any of the conclusions. To improve readability, statistics, tables, and figures are based on the total sample and the set of focal predictors. Only for “sexual partners”, for which the response options differed between the two samples, are the results

presented separately for Sample 1 and Sample 2. We refer readers interested in the robustness of our analyses to the supplementary online materials [SOM], where results are presented separately for both samples (Tables S1-S6) and for models including participant gender (Tables S7-S10).

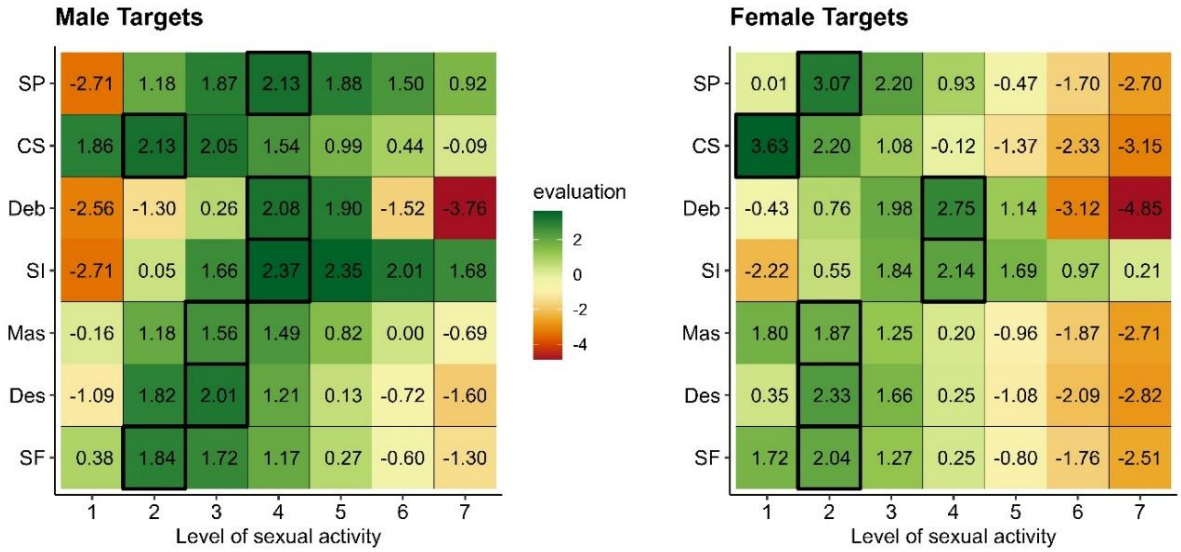
Results

Descriptive Statistics

Figure 3 visualizes the mean estimates of societal evaluations of the seven activity levels for each sexual outcome separately for female and male targets. This plot illustrates some basic findings: First, the most positive estimated societal evaluations tend to occur for low to moderate levels of sexual activity, not for the lowest level or for the highest level. Second, overall, lower levels of sexual activity tend to be evaluated more positively for women than for men, and higher levels of sexual activity tend to be evaluated more positively for men than for women; the ILSA (indicated by black borders across the respective squares) also tends to occur at higher levels of activity for men than for women.

Figure 3

Perceived Societal Evaluations: Mean Estimates Across Outcomes, Targets, and Activity Levels



Note. Perceived societal evaluations were rated on a 13-point scale, ranging from -6 = *very negatively* to 6 = *very positively*. SP = sex partners; CS = casual sex partners; Deb = sexual debut; SI = sexual intercourse; Mas = masturbation; Des = sexual desire; SF = sexual fantasies. Shown in black are the highest rated levels of sexual activity across all participants (ILSA).

Intergender Differences (Preregistered Confirmatory)

Effect-Size Sensitivity Analyses

The S&D model predicts that high sexual activity is evaluated more favorably for men than for women (H1), whereas low sexual activity is evaluated more favorably for women than for men (H2). Effect-size sensitivity analyses indicated 80% (90%) power to detect small to medium *inter*gender differences of d_{rm} = .13 (.16) in the total sample.

Intergender Differences: Male versus Female Targets (H1, H2)

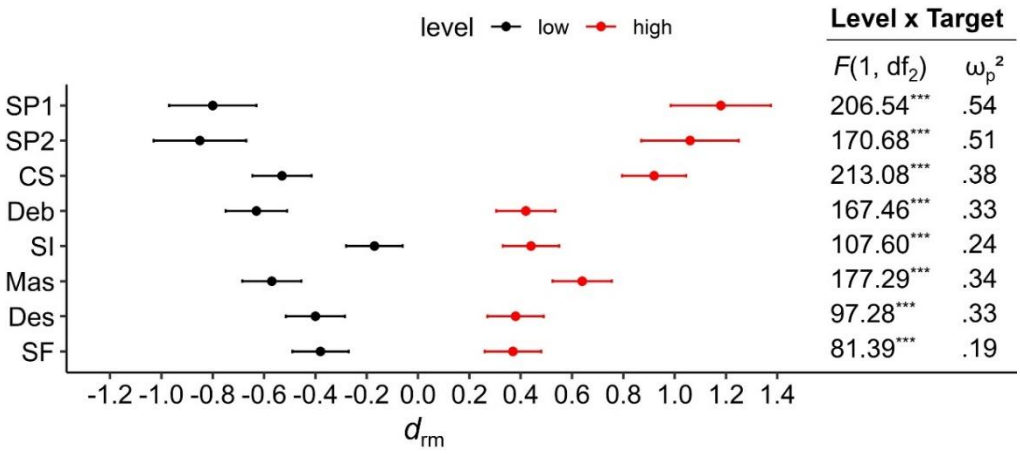
In support of H1 and H2, two-way interactions between target gender and activity level were significant across all seven sexual outcomes, with medium to large effect sizes by convention. Conceptually replicating previous research (Endendijk et al., 2020), high sexual activity was consistently evaluated more positively for men than for women across all outcomes. Extending previous research, low sexual activity was consistently evaluated more positively for women than for men across all outcomes. The results are summarized in Figure 4.

Intergender Differences for All Levels of Sexual Activity (Non-Preregistered)

Exploratory comparisons revealed significant *inter*gender differences for most levels of sexual activity. In general, (very) low levels of sexual activity were evaluated more favorably for women than for men, whereas high(er) levels were evaluated more favorably for men than for women (see Table S3 for an overview).

Figure 4

Intergender Comparisons for Low and High Sexual Activity



Note. Shown are point estimates and 95% noncentral confidence intervals for d_{rm} (i.e., repeated-measures equivalent to Cohen's *d* controlling for the correlation between the evaluations of male and female targets; see Lakens, 2013). Values less than 0 indicate better evaluations of female targets than of male targets; values greater than 0 indicate better evaluations of male targets than of female targets. df_2 = 175 (SP1), 165 (SP2), or 361 (all other outcomes); SP = sex partners; CS = casual sex partners; Deb = sexual debut; SI = sexual intercourse; Mas = masturbation; Des = sexual desire; SF = sexual fantasies. *** $p < .001$.

Intragender Trajectories (Preregistered Confirmatory)

Effect-Size Sensitivity Analyses

The strong SDS predicts that higher levels of sexual activity are increasingly rewarded for men, but increasingly punished for women (i.e., monotonic, linear associations). In contrast, the S&D model predicts that for both men (H3) and women (H4), sexual activity is increasingly rewarded up to the ILSA and increasingly punished when exceeding the ILSA (i.e., curvilinear, quadratic associations). We conducted simulation-based effect-size sensitivity analyses using the *simr* package (Green & MacLeod, 2016) to determine the increment in the within-person

variance explained by the fixed effects of the quadratic over the linear effect of sexual activity level ($\Delta R_w^{2\,(f1)}$; Rights & Sterba, 2019) that can be detected with a power of ≥ 80%. These effects were very small to large for female targets (H3: $.01 \leq \Delta R_w^{2\,(f1)} \leq .12$) and very small to medium for male targets (H4: $.01 \leq \Delta R_w^{2\,(f1)} \leq .07$).

Intragender Trajectories: Monotonic versus Nonmonotonic Associations (H3, H4)

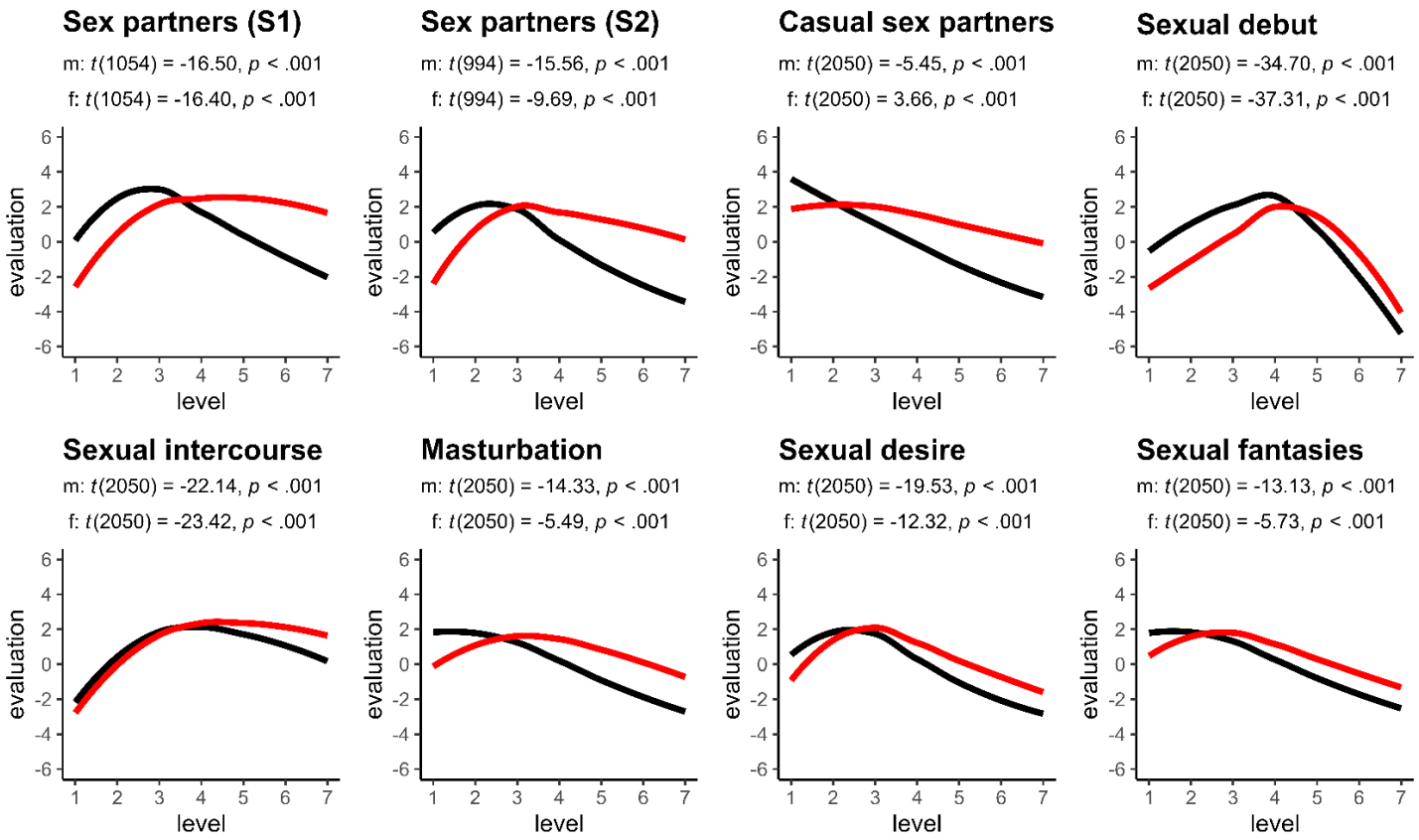
For male targets (H3), *intra*gender trajectories were curvilinear and nonmonotonic across all seven sexual outcomes. Multilevel orthogonal polynomial analyses consistently revealed incremental effects of the quadratic model over the linear model (*p*s < .001, $.01 \leq \Delta R_w^{2\,(f1)} \leq .33$). For female targets (H4), almost all effects were again curvilinear and nonmonotonic, with incremental quadratic effects (*p*s < .001, $.01 \leq \Delta R_w^{2\,(f1)} \leq .32$). Only for casual sex partners did we find monotonically negative trajectories (i.e., more is worse). Figure 5 illustrates estimated societal evaluations for female (black lines) and male (red lines) targets along the continuum of low to high sexual activity. Overall, *intra*gender trajectories were mostly curvilinear and nonmonotonic for both genders, consistent with the S&D model but inconsistent with previous models of sexual (double) standards.

Do Ideal Levels of Sexual Activity Differ Between the Genders? (H5, Non-Preregistered Exploratory)

The aggregated data depicted in Figure 5 reveal striking similarities between the trajectories of evaluations for men and women, but they also suggest that the sexual activity level that is rated highest on average, the ILSA, is higher for men than for women. Paired *t* tests statistically supported this observation for all seven outcomes (*p*s < .001; $0.39 \leq d_{rm} \leq 1.27$). Although very low and very high levels of sexual activity are not particularly valued for either gender, the most positive evaluations are found at higher levels of sexual activity for men than for women (see Table 2).

Figure 5

Perceived Societal Evaluations as a Function of Target Gender and Activity Level



Note. The graphs show average perceived societal evaluations of male (*red*) and female (*black*) targets, except for "sex partners," for which the level operationalizations differed between Sample 1 and Sample 2 (see Method section) and that are therefore presented separately. Statistics refer to the quadratic component for male (m) and female (f) targets.

Table 2

Comparisons of Male and Female ILSAs

Outcome	M_m	SD_m	M_f	SD_f	M_{dif}	t(df)	p	$CI_{95\%}$	d_{rm}
Sex partners (S1)	4.57	1.70	2.71	1.18	1.86	12.65	< .001	[1.57, 2.15]	1.27
Sex partners (S2)	3.83	1.69	2.31	1.13	1.52	10.02	< .001	[1.22, 1.82]	1.05
Casual sex partners	2.95	1.96	1.68	1.15	1.27	11.79	< .001	[1.06, 1.49]	0.77
Sexual debut	3.24	1.09	2.66	1.11	0.58	9.03	< .001	[0.45, 0.71]	0.53
Intercourse	4.54	1.64	4.05	1.58	0.48	6.89	< .001	[0.35, 0.62]	0.30
Masturbation	3.32	1.63	2.34	1.39	0.98	10.97	< .001	[0.80, 1.16]	0.64
Sexual desire	3.08	1.47	2.48	1.25	0.60	7.39	< .001	[0.44, 0.76]	0.44
Sexual fantasies	2.91	1.58	2.32	1.45	0.59	6.37	< .001	[0.41, 0.77]	0.39

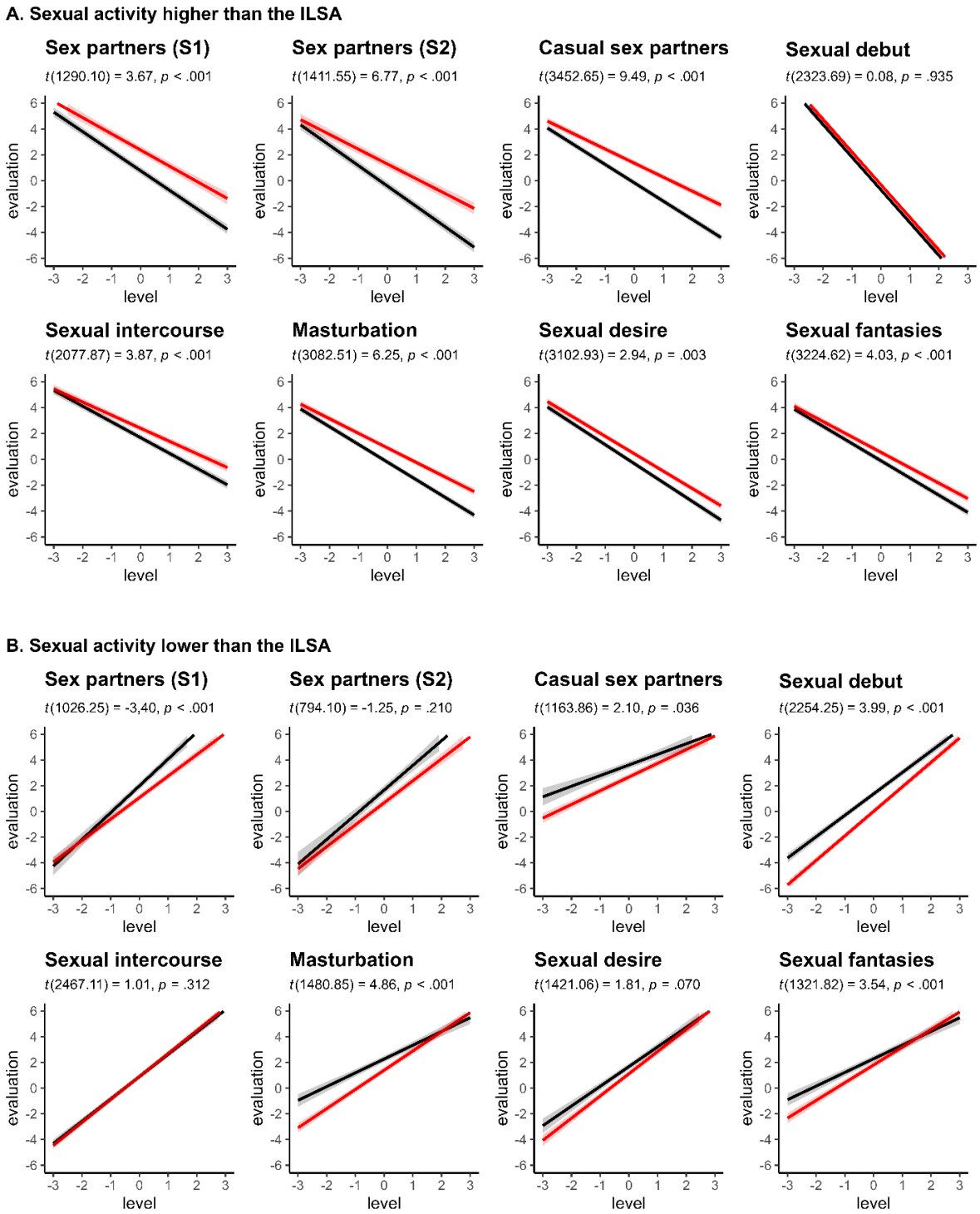
Note. Means (*M*) and standard deviations (*SD*) of the ILSA for male (m) and female (f) targets. df = 361 (total sample), 175 (S1), or 165 (S2). Positive mean differences (M_{dif}) and effect sizes indicate a higher ILSA for male targets than for female targets.

Asymmetry in the Devaluation of Male and Female Targets (H6, Non-Preregistered Exploratory)

Women and men who deviate from their (gender-specific) ILSA to the same extent may not be evaluated similarly. We speculated that the negative effects of exceeding the ILSA would be more pronounced for women than for men, consistent with a particularly negative evaluation of high female sexual activity. Exploratory multilevel analyses including only levels of sexual activity equal to or greater than the ILSA supported this assumption (see Figure 6, Panel A). Parallel analyses including activity levels equal to or below the ILSA inconsistently supported that the devaluation of levels considered “too low” was more pronounced for men than for women (Figure 6, Panel B).

Figure 6

Asymmetry in the Devaluation of Values Deviating from the ILSA



Note. Perceived societal devaluation of male (*red*) and female (*black*) targets for values greater than the ILSA (Panel A) or for values lower than the ILSA (Panel B). We centered level at its

person mean, so that 0 denotes the midpoint between the ILSA and the highest or lowest values, respectively (Enders & Tofighi, 2007). The statistics presented refer to the interaction between sexual activity level and target, shaded areas indicate values within the 95% confidence interval.

Discussion

How different or similar are the sexual standards that are applied to men and women? High-powered analyses across seven outcomes consistently showed that participants perceived that society evaluates highly sexually active men and highly sexually inactive women more positively than equally (in)active individuals of the opposite gender. In addition, the "ideal" level of sexual activity (ILSA) was higher for men than for women. People also perceived that some deviations from the ILSA are more consequential than others: activity levels above the ILSA were more devalued for women than for men, while activity levels below the ILSA tended to be more devalued for men than for women. These results support the continued existence of different sexual norms for men and women.

However, this pattern of pronounced gender differences does not imply an absence of similarities, nor does it imply that sexual activity is generally socially denigrated or rewarded in either gender. The associations between sexual activity and perceived societal evaluations were nonmonotonic and curvilinear for both male and female targets. Participants perceived that the most valued levels of sexual activity were neither the lowest for women nor (and especially not) the highest for men. Rather, the most positive evaluations were found for low to moderate levels of sexual activity.

In sum, our findings provide a perspective on sexual norms that differs critically from previously discussed models. We robustly found differences between male and female sexual norms for common sexual event types and frequencies that are prevalent in everyday life. These

findings do not support the notion that there is a single standard for most sexual event types and that the SDS today would be limited to a few types of sexual behavior that are uncommon (e.g., threesomes) or extremely different from the median level of sexual activity in these samples (e.g., 50 sexual partners; Jonason, 2008; Jonason & Marks, 2009). In addition, curvilinear *intra*gender associations significantly qualify the assumption that higher sexual activity is generally socially rewarded or socially punished for either gender, as expressed in existing models of strong and weak double standards. Instead, male and female sexual norms are equally characterized by similarities and differences, a pattern uniquely predicted by the S&D model. This may further explain some of the heterogeneity of *inter*- and *intra*gender effects in previous research: if male and female sexual norms along the continuum from very low to very high sexual activity are best represented by two intersecting curved lines, then the direction and magnitude of gender differences will depend on which level(s) of sexual activity are examined in the study.

The coexistence of similarities and differences is consistent with theories used to explain the SDS. Evolutionary theories suggest that the ideal level of sexual activity should be higher for men than for women because of women's greater investment in producing and raising offspring (Trivers, 1972), but also emphasize the challenges associated with very low and very high levels of sexual activity (Buss & Schmitt, 1993). A higher male than female ILSA is also consistent with different gender role expectations for men and women (Wood & Eagly, 2002), which are internalized based on personal experience and model learning (Bandura, 1986). However, in modern Western societies, individuals do not learn that behaviors associated with maximum sexual restraint or with maximum sexual permissiveness are part of the gender roles assigned to men and women.

Implications for Society and Science

For decades, the SDS has shaped researchers' and laypeople's thinking about the standards that society presumably applies to men's and women's sexuality (Crawford & Popp, 2003; Reiss, 1956). The belief that there is a strong SDS may pressure women to be sexually passive and men to be sexually active beyond their natural inclinations (Crawford & Popp, 2003; Wesson, 2022) and may explain the more negative emotional reactions following sexual activity for women than for men (McKeen et al., 2022). Therefore, communicating the similarities between male and female sexual standards alongside the differences may not only contribute to a more accurate understanding of sexual norms, but may also work against the perpetuation of established stereotypes and help to increase sexual well-being more generally.

The observation that, even for men, moderate levels of sexual activity are more positively evaluated than high levels is particularly noteworthy because the event frequencies representing the high end of the sexual activity continuum were rather moderate, not extreme (e.g., masturbating 6 times per week). Thus, the devaluation of higher levels of sexual activity was not due to the presentation of hypothetical, clearly extreme manifestations of sexual activity, but occurred for manifestations that are common in everyday life (Haversath et al., 2017; Weber et al., 2024).

Finally, several statements about the differential nature of male and female sexual norms have been collected under the umbrella of the traditional SDS—most prominently, that (high) sexual activity is evaluated more favorably for men than for women (e.g., Endendijk et al., 2020; Zaikman & Marks, 2017), that men are granted more sexual freedom than women (e.g., Bordini & Sperb, 2013; Milhausen & Herold, 2002), and that sexual activity is socially rewarded for men but socially punished for women (e.g., Marks et al., 2019; Marks & Fraley, 2005). In the present research, we found robust evidence for *inter*gender differences (i.e., different evaluations of the

same level of sexual activity depending on gender), but little evidence that very high or very low sexual activity is particularly rewarded for either gender. We believe that future research would benefit from carefully distinguishing between assumptions that have received substantial empirical support and those that do not. In addition, the extent to which men subjectively perceive a higher ILSA as more sexual freedom rather than as pressure to be sexually active is an interesting question to explore in future research.

Limitations

All participants were German residents. As sexual (double) standards differ across countries (Endendijk et al., 2020; Sprecher & Hatfield, 1996), our findings may not generalize to other societies. Based on the S&D model, we speculate that cultural differences may be reflected in the exact shapes of the predicted curves. For example, in liberal countries with high levels of gender equality, female and male ILSAs may be close at higher levels, and societal devaluation for exceeding the ILSA may be less pronounced.

In addition, we focused on global societal evaluations ("How would society view…?"). These represent holistic perceptions of societal pressures that people face in their daily lives. Other researchers have examined specific dimensions on which people may be judged (e.g., intelligence, power, or popularity; Marks et al., 2019; Marks & Fraley, 2005). Examining the extent to which curvilinear associations occur on these specific dimensions is an avenue for future research.

Finally, we focused on participants' perceived societal evaluations because these are likely to be particularly powerful in influencing how people think, feel, and behave in their everyday lives. Preliminary analyses using exploratorily measured personal evaluations ("How would *you* judge…?") in our own data suggest that the reported coexistence of similarities and differences may in part generalize to personal standards, but that *inter*gender differences may be

less consistent and weaker than those found for societal evaluations, underscoring the importance of distinguishing between personal and perceived societal evaluations (Milhausen & Herold, 2002).

Conclusion

There are both similarities and differences in the societal evaluation of male and female sexuality. People perceive that high levels of sexual activity are evaluated more positively for men than for women, while low levels of sexual activity are evaluated more positively for women than for men. However, contrary to common assumptions, moderate rather than extremely low or extremely high levels of sexual activity are most valued for both genders.

Acknowledgments

We thank Paula Laute and Tobias Ludwig for their valuable assistance in creating the online surveys and preparing the manuscript.

References

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*.

Bordini, G. S., & Sperb, T. M. (2013). Sexual double standard: A review of the literature between 2001 and 2010. *Sexuality and Culture*, *17*(4), 686–704. https://doi.org/10.1007/s12119-012-9163-0

Buss, D. M., & Schmitt, D. P. (1993). Sexual Strategies Theory: An evolutionary perspective on human mating. *Psychological Review*, *100*(2), 204–232.

Crawford, M., & Popp, D. (2003). Sexual double standards: A review and methodological critique of two decades of research. *Journal of Sex Research*, *40*(1), 13–26. https://doi.org/10.1080/00224490309552163

Eagly, A. H., & Wood, W. (2012). Social role theory. In P. Van Lange, A. Kruglanski, & E. Higgins, *Handbook of Theories of Social Psychology* (pp. 458–476). SAGE Publications Ltd. https://doi.org/10.4135/9781446249222.n49

Endendijk, J. J., van Baar, A. L., & Deković, M. (2020). He is a stud, she is a slut! A meta-analysis on the continued existence of sexual double standards. *Personality and Social Psychology Review*, *24*(2), 163–190. https://doi.org/10.1177/1088868319891310

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, *12*(2), 121–138. https://doi.org/10.1037/1082-989X.12.2.121

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. sage.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Gesselman, A. N., Webster, G. D., & Garcia, J. R. (2017). Has virginity lost its virtue? Relationship stigma associated with being a sexually inexperienced adult. *Journal of Sex Research*, *54*(2), 202–213. https://doi.org/10.1080/00224499.2016.1144042

Giner-Sorolla, R., Aberson, C. L., Bostyn, D. H., Carpenter, T., Conrique, B. G., Lewis, N. A., & Soderberg, C. (2019). *Power to detect what? Considerations for planning and evaluating sample size*.

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*(3), 213–224.

Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Haversath, J., Gärttner, K. M., Kliem, S., Vasterling, I., Strauss, B., & Kröger, C. (2017). Sexualverhalten in Deutschland. *Deutsches Ärzteblatt*, *114*(33–34), 545–550. https://doi.org/10.3238/arztebl.2017.0545

Jonason, P. K. (2008). A mediation hypothesis to account for the sex difference in reported number of sexual partners: An intrasexual competition approach. *International Journal of Sexual Health*, *19*(4), 41–49. https://doi.org/10.1300/J514v19n04_05

Jonason, P. K., & Marks, M. J. (2009). Common vs. uncommon sexual acts: Evidence for the sexual double standard. *Sex Roles*, *60*(5–6), 357–365. https://doi.org/10.1007/s11199-008-9542-z

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00863

Leiner, D. J. (2022). *SoSci Survey (Version 3.2.31) [Computer software]*. Available at https://www.soscisurvey.de

Marks, M. J., & Fraley, R. C. (2005). The sexual double standard: Fact or fiction? *Sex Roles*, *52*(3), 175–186. https://doi.org/10.1007/s11199-005-1293-5

Marks, M. J., & Fraley, R. C. (2007). The impact of social interaction on the sexual double standard. *Social Influence*, *2*(1), 29–54. https://doi.org/10.1080/15534510601154413

Marks, M. J., Young, T. M., & Zaikman, Y. (2019). The sexual double standard in the real world: Evaluations of sexually active friends and acquaintances. *Social Psychology*, *50*(2), 67–79. https://doi.org/10.1027/1864-9335/a000362

McKeen, B. E., Anderson, R. C., & Mitchell, D. A. (2022). Was it good for you? Gender differences in motives and emotional outcomes following casual sex. *Sexuality & Culture*, *26*(4), 1339–1359. https://doi.org/10.1007/s12119-022-09946-w

Milhausen, R. R., & Herold, E. S. (2002). Reconceptualizing the sexual double standard. *Journal of Psychology & Human Sexuality*, *13*(2), 63–83. https://doi.org/10.1300/J056v13n02_05

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Pickens, C., & Braun, V. (2018). "Stroppy bitches who just need to learn how to settle"? Young single women and norms of femininity and heterosexuality. *Sex Roles, 79*(7–8), 431–448. https://doi.org/10.1007/s11199-017-0881-5

R Core Team. (2023). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria*. https://www.R-project.org/

Reiss, I. L. (1956). The double standard in premarital sexual intercourse—A neglected concept. *Social Forces*, *34*(1). https://www.proquest.com/docview/1290943608/citation/37BB566E87074941PQ/1

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*, *24*(3), 309–338. https://doi.org/10.1037/met0000184

Sprecher, S., & Hatfield, E. (1996). Premarital sexual standards among U.S. college students: Comparison with Russian and Japanese students. *Archives of Sexual Behavior*, *25*(3), 261–288. https://doi.org/10.1007/BF02438165

Trivers, R. L. (1972). Parental investment and sexual selection. *In B. Campbell (Ed.), Sexual Selection and the Descent of Man 1871-1971 (Pp. 136–207). Aldine de Gruyter.*

Weber, M., Reis, D., & Friese, M. (2024). Development and validation of the trait sexual motivation scale (TSMS). *Journal of Personality Assessment*, *106*(2), 267–282. https://doi.org/10.1080/00223891.2023.2206896

Wesche, R., Claxton, S. E., & Waterman, E. A. (2021). Emotional outcomes of casual sexual relationships and experiences: A systematic review. *Journal of Sex Research*, *58*(8), 1069–1084. https://doi.org/10.1080/00224499.2020.1821163

Wesson, C. (2022). A mixed-methods exploration of young adults' views of the sexual double standard. *Sexuality & Culture*, *26*(4), 1314–1325. https://doi.org/10.1007/s12119-022-09944-y

Wood, W., & Eagly, A. H. (2002). A cross-cultural analysis of the behavior of women and men: Implications for the origins of sex differences. *Psychological Bulletin*, *128*(5), 699–727. https://doi.org/10.1037/0033-2909.128.5.699

Zaikman, Y., & Marks, M. J. (2017). Promoting theory-based perspectives in sexual double standard research. *Sex Roles*, *76*(7–8), 407–420. https://doi.org/10.1007/s11199-016-0677-z

Zaikman, Y., Marks, M. J., Young, T. M., & Zeiber, J. A. (2016). Gender role violations and the sexual double standard. *Journal of Homosexuality*, *63*(12), 1608–1629. https://doi.org/10.1080/00918369.2016.1158007

Zava (2021). *Lying between the sheets: Is my partner honest about the sexual past?* https://www.zavamed.com/uk/lying-between-the-sheets.html

Supplementary Online Materials (SOM)

Table S1

Intergender Effects: Omnibus Tests

	Sample 1				Sample 2			
Outcome	Effect	$F(1, 175)$	p	ω_p	Effect	$F(1, 165)$	p	ω_p
Sex partners	Level	22.63	< .001	.11	Level	2.47	.118	.02
	Target	14.18	< .001	.07	Target	3.66	.058	.02
	Level x Target	206.55	< .001	.54	Level x Target	170.68	< .001	.51
Casual sex partners	Level	256.41	< .001	.59	Level	144.42	< .001	.46
	Target	31.64	< .001	.15	Target	7.08	.009	.04
	Level x Target	139.96	< .001	.44	Level x Target	78.62	< .001	.32
Sexual debut	Level	125.60	< .001	.41	Level	97.84	< .001	.37
	Target	24.46	< .001	.12	Target	11.18	.001	.06
	Level x Target	110.76	< .001	.38	Level x Target	63.91	< .001	.27
Sexual intercourse	Level	325.33	< .001	.65	Level	19.53	< .001	.10
	Target	27.26	< .001	.13	Target	8.30	.004	.04
	Level x Target	51.32	< .001	.22	Level x Target	56.26	< .001	.25
Masturbation	Level	56.96	< .001	.24	Level	53.85	< .001	.24
	Target	0.00	1.00	.00	Target	0.19	.661	.00
	Level x Target	129.39	< .001	.42	Level x Target	56.28	< .001	.25
Sexual desire	Level	17.89	< .001	.09	Level	56.12	< .001	.25
	Target	0.20	.656	.00	Target	1.50	.223	.00
	Level x Target	64.39	< .001	.26	Level x Target	35.02	< .001	.17
Sexual fantasies	Level	69.99	< .001	.28	Level	81.40	< .001	.33
	Target	0.19	.666	.00	Target	2.45	.119	.01
	Level x Target	66.77	< .001	.27	Level x Target	20.64	< .001	.11

Table S2

Descriptive Statistics and Intergender Effects for All Levels of Sexual Activity

	Sample 1						Sample 2				
Outcome	Level	M_m (SD_m)	M_f (SD_f)	$t(175)$	p	d_{rm}	Level	M_m (SD_m)	M_f (SD_f)	$t(165)$	p	d_{rm}
Sex partners	1	-2.76 (3.04)	-0.20 (3.31)	-10.46	< .001	-0.80	1	-2.65 (3.10)	0.23 (3.63)	-9.57	< .001	-0.85
	2	1.00 (3.36)	3.25 (2.52)	-9.27	< .001	-0.75	2	1.36 (3.19)	2.88 (2.53)	-6.18	< .001	-0.52
	3	1.91 (2.87)	2.72 (2.58)	-3.78	< .001	-0.29	3	1.82 (2.60)	1.64 (2.70)	0.78	.438	0.07
	4	2.56 (2.49)	1.76 (2.75)	4.51	< .001	0.30	4	1.67 (2.79)	0.04 (2.92)	6.56	< .001	0.57
	5	2.48 (2.55)	0.29 (2.87)	10.36	< .001	0.80	5	1.25 (2.98)	-1.28 (3.12)	9.30	< .001	0.83
	6	2.18 (2.77)	-0.90 (2.90)	13.58	< .001	1.08	6	0.78 (3.26)	-2.55 (2.97)	11.93	< .001	1.07
	7	1.67 (3.24)	-2.02 (3.02)	13.11	< .001	1.18	7	0.13 (3.66)	-3.42 (2.94)	11.80	< .001	1.06
Casual sex partners	1	2.22 (3.42)	4.02 (2.49)	-7.29	< .001	-0.59	1	1.49 (3.86)	3.21 (3.25)	-6.33	< .001	-0.48
	2	2.43 (2.37)	2.02 (2.60)	2.22	.028	0.17	2	1.82 (3.06)	2.40 (2.91)	-2.92	.004	-0.20
	3	2.41 (2.29)	0.80 (2.84)	8.30	< .001	0.62	3	1.67 (2.93)	1.37 (2.92)	1.55	.124	0.10
	4	1.88 (2.53)	-0.54 (2.87)	11.65	< .001	0.89	4	1.18 (2.98)	0.32 (3.04)	4.32	< .001	0.29
	5	1.28 (2.93)	-1.68 (2.87)	12.24	< .001	1.02	5	0.67 (3.24)	-1.05 (3.05)	7.62	< .001	0.55
	6	0.68 (3.14)	-2.59 (2.97)	12.35	< .001	1.07	6	0.17 (3.29)	-2.06 (3.07)	8.33	< .001	0.70
	7	0.18 (3.52)	-3.30 (3.06)	11.82	< .001	1.08	7	-0.37 (3.58)	-2.91 (2.96)	8.55	< .001	0.77
Sexual debut	1	-2.74 (3.21)	-0.49 (3.51)	-10.08	< .001	-0.67	1	-2.38 (3.22)	-0.36 (3.56)	-8.00	< .001	-0.59
	2	-1.35 (3.17)	0.85 (2.95)	-11.00	< .001	-0.71	2	-1.25 (3.00)	0.66 (3.03)	-8.84	< .001	-0.64
	3	0.24 (3.07)	2.14 (2.59)	-9.80	< .001	-0.66	3	0.28 (3.01)	1.80 (2.81)	-7.53	< .001	-0.52
	4	2.16 (2.73)	3.00 (2.47)	-4.30	< .001	-0.32	4	1.99 (2.56)	2.49 (2.65)	-3.03	.003	-0.19
	5	2.20 (2.47)	1.40 (2.62)	4.32	< .001	0.32	5	1.58 (2.77)	0.87 (3.04)	3.81	< .001	0.25
	6	-1.32 (2.86)	-3.10 (2.19)	8.95	< .001	0.69	6	-1.72 (2.87)	-3.14 (2.65)	6.65	< .001	0.51
	7	-4.05 (2.56)	-5.12 (1.50)	6.42	< .001	0.47	7	-3.46 (3.07)	-4.56 (2.52)	4.89	< .001	0.39
Sexual intercourse	1	-3.88 (2.79)	-3.47 (2.83)	-4.07	< .001	-0.15	1	-1.69 (3.67)	-0.90 (3.67)	-4.99	< .001	-0.22
	2	-0.65 (3.25)	-0.11 (3.16)	-4.06	< .001	-0.17	2	0.80 (3.29)	1.25 (3.19)	-2.92	.004	-0.14
	3	1.51 (2.80)	1.67 (2.66)	-1.43	.155	-0.06	3	1.83 (2.73)	2.02 (2.68)	-1.75	.082	-0.07
	4	2.59 (2.23)	2.39 (2.25)	1.83	.069	0.09	4	2.13 (2.56)	1.88 (2.68)	1.89	.060	0.10
	5	2.91 (2.11)	2.15 (2.45)	4.97	< .001	0.33	5	1.75 (2.70)	1.20 (2.83)	2.94	.004	0.20
	6	2.69 (2.45)	1.65 (2.78)	5.66	< .001	0.40	6	1.30 (2.76)	0.25 (3.01)	5.17	< .001	0.36
	7	2.45 (2.89)	0.97 (3.34)	6.92	< .001	0.45	7	0.86 (3.09)	-0.60 (3.38)	6.72	< .001	0.45

Table S2 [*continued*]

		Sample 1						Sample 2				
Outcome	Level	M_m (SD_m)	M_f (SD_f)	t(175)	p	d_{rm}	Level	M_m (SD_m)	M_f (SD_f)	t(165)	p	d_{rm}
Masturbation	1	-0.68 (3.59)	1.83 (3.17)	-8.93	< .001	-0.74	1	0.39 (3.54)	1.77 (3.42)	-6.56	< .001	-0.40
	2	1.06 (3.07)	1.81 (2.59)	-3.04	.003	-0.26	2	1.30 (2.87)	1.95 (2.65)	-3.11	.002	-0.24
	3	1.61 (2.66)	1.10 (2.64)	2.25	.026	0.20	3	1.51 (2.50)	1.41 (2.56)	0.52	.606	0.04
	4	1.71 (2.27)	0.02 (2.71)	8.41	< .001	0.67	4	1.25 (2.50)	0.40 (2.66)	4.57	< .001	0.33
	5	1.03 (2.45)	-1.19 (2.68)	10.20	< .001	0.87	5	0.60 (2.82)	-0.72 (2.86)	6.08	< .001	0.46
	6	0.17 (2.85)	-2.15 (2.72)	10.55	< .001	0.83	6	-0.18 (3.19)	-1.58 (3.19)	5.66	< .001	0.44
	7	-0.56 (3.14)	-3.06 (2.78)	10.67	< .001	0.84	7	-0.84 (3.39)	-2.33 (3.31)	5.95	< .001	0.45
Sexual desire	1	-1.55 (3.59)	0.01 (3.54)	-7.53	< .001	-0.44	1	-0.60 (3.58)	0.72 (3.61)	-5.86	< .001	-0.37
	2	1.90 (2.76)	2.38 (2.52)	-2.55	.012	-0.18	2	1.75 (2.92)	2.28 (2.70)	-2.85	.005	-0.19
	3	2.38 (2.15)	1.80 (2.29)	3.63	< .001	0.26	3	1.62 (2.49)	1.52 (2.61)	0.56	.574	0.04
	4	1.65 (2.15)	0.35 (2.44)	6.78	< .001	0.56	4	0.73 (2.73)	0.14 (2.90)	3.06	.003	0.21
	5	0.52 (2.65)	-1.09 (2.51)	7.51	< .001	0.62	5	-0.27 (3.04)	-1.08 (3.01)	3.71	< .001	0.27
	6	-0.42 (2.93)	-2.12 (2.75)	6.99	< .001	0.60	6	-1.04 (3.30)	-2.06 (3.05)	4.38	< .001	0.32
	7	-1.43 (3.15)	-2.88 (2.89)	5.93	< .001	0.48	7	-1.78 (3.45)	-2.77 (3.26)	3.97	< .001	0.29
Sexual fantasies	1	-0.06 (3.71)	1.53 (3.35)	-6.84	< .001	-0.45	1	0.84 (3.53)	1.91 (3.47)	-4.96	< .001	-0.30
	2	1.65 (2.71)	1.82 (2.60)	-0.95	.343	-0.07	2	2.04 (2.65)	2.27 (2.56)	-1.29	.198	-0.09
	3	1.79 (2.25)	0.97 (2.64)	4.40	< .001	0.33	3	1.64 (2.61)	1.60 (2.59)	0.28	.783	0.02
	4	1.36 (2.35)	-0.03 (2.64)	6.97	< .001	0.56	4	0.97 (2.73)	0.55 (2.67)	2.33	.021	0.16
	5	0.53 (2.63)	-1.11 (2.63)	7.55	< .001	0.62	5	0.01 (2.98)	-0.47 (3.02)	2.31	.022	0.16
	6	-0.46 (2.82)	-2.11 (2.81)	7.19	< .001	0.59	6	-0.75 (3.27)	-1.32 (3.28)	2.78	.006	0.19
	7	-1.19 (3.02)	-2.89 (2.94)	7.05	< .001	0.57	7	-1.42 (3.49)	-2.11 (3.49)	2.86	.005	0.20

Note. Means (*M*) and standard deviations (*SD*) for male (m) and female targets (f). Positive effects in paired sample *t*-tests indicate a more positive evaluation of male targets compared to female targets.

Table S3

Intragender Trajectories: Multilevel Polynomial Regression Analyses

	Male targets						Female targets					
Sample 1 (*n* = 176)												
	B	*SE*	*df*	*t*	*p*	$R_w^{2\,(f1)}$	*B*	*SE*	*df*	*t*	*p*	$R_w^{2\,(f1)}$
Sex partners						.301						.311
Intercept	2.67	0.16	324.40	16.76	< .001	-	1.99	0.17	283.56	11.85	< .001	-
Level	3.48	0.22	1054.00	16.07	< .001	.143	-3.47	0.20	1054.00	-16.92	< .001	.156
Level²	-12.37	0.75	1054.00	-16.50	< .001	.158	-11.64	0.71	1054.00	-16.40	< .001	.154
Casual sex partners						.113						.625
Intercept	1.90	0.17	268.59	11.21	< .001	-	-0.57	0.18	226.76	-3.22	.001	-
Level	-2.30	0.20	1054.00	-11.79	< .001	.100	-7.26	0.16	1054.00	-44.87	< .001	.613
Level²	-2.85	0.68	1054.00	-4.22	< .001	.013	3.41	0.56	1054.00	6.08	< .001	.012
Sexual debut						.375						.543
Intercept	1.68	0.15	374.52	11.03	< .001	-	2.22	0.14	376.13	16.07	< .001	-
Level	-0.41	0.23	1054.00	-1.80	.073	.002	-4.83	0.21	1054.00	-23.33	< .001	.191
Level²	-21.38	0.79	1054.00	-27.12	< .001	.373	-21.72	0.72	1054.00	-30.29	< .001	.352
Sexual intercourse						.506						.380
Intercept	2.56	0.14	321.30	17.666	< .001	-	2.35	0.15	308.40	15.17	< .001	-
Level	5.80	0.20	1054.00	29.66	< .001	.345	3.71	0.20	1054.00	18.34	< .001	164
Level²	-13.19	0.68	1054.00	-19.47	< .001	.161	-14.36	0.70	1054.00	-20.47	< .001	.217
Masturbation						.108						.389
Intercept	1.62	0.15	337.42	10.53	< .001	-	0.05	0.16	287.52	0.31	.760	-
Level	-0.43	0.22	1054.00	-2.00	.046	.003	-5.33	0.19	1054.00	-27.72	< .001	.381
Level²	-8.98	0.75	1054.00	-12.05	< .001	.105	-2.55	0.67	1054.00	-3.83	< .001	.008
Sexual desire						.215						.310
Intercept	1.87	0.15	341.04	12.24	< .001	-	0.63	0.15	336.11	4.19	< .001	-
Level	-1.31	0.22	1054.00	-6.10	< .001	.023	-4.40	0.21	1054.00	-21.01	< .001	.245
Level²	-12.94	0.75	1054.00	-17.34	< .001	.192	-7.67	0.73	1054.00	-10.56	< .001	.065
Sexual fantasies						.152						.371
Intercept	1.41	0.16	298.13	8.91	< .001	-	0.04	0.16	264.94	0.23	.821	-
Level	-1.90	0.20	1054.00	-9.46	< .001	.061	-4.97	0.19	1054.00	-26.60	< .001	.362
Level²	-7.98	0.70	1054.00	-11.48	< .001	.092	-2.67	0.65	1054.00	-4.12	< .001	.009

Table S3 [*continued*]

	Male targets						Female targets					
Sample 2 (*n* = 166)												
	B	*SE*	*df*	*t*	*p*	$R_w^{2\,(f1)}$	*B*	*SE*	*df*	*t*	*p*	$R_w^{2\,(f1)}$
Sex partners						.195						.368
Intercept	1.98	0.18	271.09	10.78	< .001	-	0.47	0.18	265.88	2.59	.010	-
Level	1.42	0.23	994.00	6.25	< .001	.026	-5.30	0.22	994.00	-24.16	< .001	.315
Level²	-12.22	0.79	994.00	-15.56	< .001	.169	-7.36	0.76	994.00	-9.69	< .001	.054
Casual sex partners						.084						.482
Intercept	1.24	0.20	240.04	6.17	< .001	-	0.22	0.19	230.08	1.15	.251	-
Level	-2.11	0.22	994.00	-9.69	< .001	.074	-6.37	0.19	994.00	-32.83	< .001	.482
Level²	-2.65	0.76	994.00	-3.51	< .001	.010	-0.31	0.67	994.00	-0.47	.642	.000
Sexual debut						.297						.438
Intercept	1.33	0.16	372.01	8.48	< .001	-	1.71	0.16	330.64	10.61	< .001	-
Level	-0.62	0.24	994.00	-2.57	.010	.004	-4.53	0.23	994.00	-19.60	< .001	.179
Level²	-18.33	0.83	994.00	-22.01	< .001	.293	-18.25	0.80	994.00	-22.81	< .001	.259
Sexual intercourse						.179						.152
Intercept	2.11	0.17	272.92	12.12	< .001	-	1.91	0.18	255.32	10.34	< .001	-
Level	1.84	0.22	994.00	8.52	< .001	.050	-0.41	0.21	994.00	-1.90	.058	.003
Level²	-10.04	0.75	994.00	-13.41	< .001	.128	-10.59	0.74	994.00	-14.28	< .001	.149
Masturbation						.099						.308
Intercept	1.22	0.18	252.08	6.80	< .001	-	0.44	0.18	254.06	2.47	.014	-
Level	-1.61	0.21	994.00	-7.83	< .001	.047	-4.60	0.21	994.00	-22.42	< .001	.299
Level²	-5.82	0.71	994.00	-8.16	< .001	.052	-2.80	0.71	994.00	-3.93	< .001	.009
Sexual desire						.159						.301
Intercept	0.96	0.18	269.50	5.24	< .001	-	0.40	0.18	265.66	2.20	.028	-
Level	-2.36	0.22	994.00	-10.50	< .001	.079	-4.65	0.22	994.00	-21.30	< .001	.272
Level²	-8.09	0.78	994.00	-10.41	< .001	.080	-5.19	0.76	994.00	-6.86	< .001	.029
Sexual fantasies						.187						.310
Intercept	1.04	0.19	240.52	5.51	< .001	-	0.65	0.19	243.26	3.51	< .001	-
Level	-3.00	0.20	994.00	-14.71	< .001	.151	-4.59	0.20	994.00	-22.50	< .001	.300
Level²	-5.02	0.71	994.00	-7.10	< .001	.036	-2.81	0.71	994.00	-3.98	< .001	.010

Note. Intercepts are model-implied estimates for the average evaluation of moderate sexual activity (i.e., level 4). The polynomials are orthogonal, so that *Level* can be interpreted as the linear effect and $Level^2$ as the (incremental) quadratic effect of sexual activity.

Table S4

ILSA Comparison between Male and Female Targets

Outcome	M_m	SD_m	M_f	SD_f	M_{dif}	t(df)	p	$CI_{95\%}$	d_{rm}
Sample 1									
Sex partners	3.57	1.70	1.71	1.18	1.86	12.65	< .001	[1.57, 2.15]	1.27
Casual sex partners	1.91	1.96	0.55	1.08	1.36	8.75	< .001	[1.05, 1.67]	0.85
Sexual debut	3.38	1.01	2.70	1.06	0.67	7.42	< .001	[0.49, 0.85]	0.65
Intercourse	3.98	1.51	3.48	1.54	0.50	5.55	< .001	[0.32, 0.68]	0.33
Masturbation	2.49	1.63	1.17	1.29	1.32	10.17	< .001	[1.06, 1.58]	0.89
Sexual desire	2.13	1.44	1.44	1.18	0.68	6.06	< .001	[0.46, 0.91]	0.52
Sexual fantasies	1.98	1.56	1.16	1.27	0.82	6.57	< .001	[0.57, 1.07]	0.57
Sample 2									
Sex partners	2.83	1.69	1.31	1.13	1.52	10.02	< .001	[1.22, 1.82]	1.05
Casual sex partners	2.00	1.97	0.82	1.22	1.18	7.90	< .001	[0.89, 1.48]	0.70
Sexual debut	3.10	1.15	2.62	1.16	0.48	5.33	< .001	[0.30, 0.66]	0.42
Intercourse	3.06	1.64	2.60	1.50	0.46	4.27	< .001	[0.25, 0.67]	0.29
Masturbation	2.14	1.62	1.52	1.47	0.62	5.33	< .001	[0.39, 0.85]	0.40
Sexual desire	2.02	1.50	1.52	1.33	0.50	4.37	< .001	[0.28, 0.73]	0.36
Sexual fantasies	1.83	1.60	1.49	1.60	0.35	2.56	0.011	[0.08, 0.61]	0.22

Note. Means (*M*) and standard deviations (*SD*) of the ILSA for male (m) and female (f) targets.

df = 175 (Sample 1) or 165 (Sample 2). Positive effects indicate a higher ILSA for male targets than for female targets.

Table S5

Devaluation of Levels of Sexual Activity Greater than the ILSA

	Sample 1				Sample 2			
Outcome	Effect	*t*	*df*	*p*	Effect	*t*	*df*	*p*
Sex partners	Intercept	5.16	198.84	< .001	Intercept	-2.50	182.73	.013
	Level	-40.71	1290.10	< .001	Level	-42.99	1411.55	< .001
	Target	16.40	1311.55	< .001	Target	17.36	1430.62	< .001
	Level x Target	3.67	1290.10	< .001	Level x Target	6.77	1411.55	< .001
Casual sex partners	Intercept	-2.52	195.89	.013	Intercept	0.47	183.51	.642
	Level	-48.53	1809.34	< .001	Level	-43.54	1640.20	< .001
	Target	22.81	1829.21	< .001	Target	11.70	1653.54	< .001
	Level x Target	6.91	1809.34	< .001	Level x Target	6.59	1640.20	< .001
Sexual debut	Intercept	-6.44	282.75	< .001	Intercept	-5.36	235.06	< .001
	Level	-48.57	1177.28	< .001	Level	-35.71	1145.01	< .001
	Target	3.83	1196.51	< .001	Target	3.08	1158.96	.002
	Level x Target	-1.99	1177.28	.047	Level x Target	1.36	1145.01	.174
Sexual intercourse	Intercept	13.05	197.00	< .001	Intercept	6.94	188.31	< .001
	Level	-23.44	909.93	< .001	Level	-26.60	1166.59	< .001
	Target	9.38	916.03	< .001	Target	6.49	1174.87	< .001
	Level x Target	2.25	909.93	.025	Level x Target	3.12	1166.59	.002
Masturbation	Intercept	-2.87	198.73	.005	Intercept	-0.07	179.42	.944
	Level	-43.05	1588.91	< .001	Level	-38.51	1492.04	< .001
	Target	16.42	1604.91	< .001	Target	8.71	1499.57	< .001
	Level x Target	3.23	1588.91	.001	Level x Target	5.42	1492.04	< .001
Sexual desire	Intercept	-2.53	206.09	.012	Intercept	-1.87	185.07	.063
	Level	-44.07	1596.25	< .001	Level	-36.87	1504.39	< .001
	Target	12.61	1609.01	< .001	Target	5.15	1514.84	< .001
	Level x Target	2.27	1596.25	.023	Level x Target	1.87	1504.39	.061
Sexual fantasies	Intercept	-2.89	193.69	.004	Intercept	1.31	182.48	.193
	Level	-43.11	1670.14	< .001	Level	-39.78	1551.34	< .001
	Target	12.47	1681.18	< .001	Target	3.43	1561.33	< .001
	Level x Target	2.97	1670.14	.003	Level x Target	2.86	1551.34	.004

Note. Positive interaction coefficients indicate that the perceived societal devaluation of levels greater than the ILSA was less pronounced for male targets (coded as 1) than for female (coded as 0) targets.

Table S6

Devaluation of Levels of Sexual Activity Lower than the ILSA

	Sample 1				Sample 2			
Outcome	Effect	*t*	*df*	*p*	Effect	*t*	*df*	*p*
Sex partners	Intercept	13.79	341.92	< .001	Intercept	9.43	352.30	< .001
	Level	18.98	1026.25	< .001	Level	12.84	794.10	< .001
	Target	-6.96	1068.48	< .001	Target	-5.92	833.79	< .001
	Level x Target	-3.40	1026.25	< .001	Level x Target	-1.25	794.10	.210
Casual sex partners	Intercept	22.05	303.33	< .001	Intercept	15.35	255.03	< .001
	Level	3.33	567.96	< .001	Level	7.28	589.09	< .001
	Target	-6.24	596.06	< .001	Target	-6.26	606.42	< .001
	Level x Target	3.18	567.96	.002	Level x Target	0.31	589.09	.757
Sexual debut	Intercept	9.09	209.98	< .001	Intercept	7.13	199.86	< .001
	Level	28.57	1198.25	< .001	Level	21.48	1052.81	< .001
	Target	-14.67	1205.08	< .001	Target	-11.86	1060.29	< .001
	Level x Target	1.96	1198.25	.050	Level x Target	3.57	1052.81	< .001
Sexual intercourse	Intercept	4.14	232.28	< .001	Intercept	7.44	207.84	< .001
	Level	39.07	1421.07	.180	Level	21.93	1048.33	< .001
	Target	1.34	1429.84	< .001	Target	-1.22	1059.64	.221
	Level x Target	0.57	1421.07	.571	Level x Target	1.13	1048.33	.258
Masturbation	Intercept	13.42	288.60	< .001	Intercept	12.09	209.16	< .001
	Level	9.67	757.62	< .001	Level	11.45	716.99	< .001
	Target	-7.25	788.37	< .001	Target	-5.96	727.98	< .001
	Level x Target	3.25	757.62	.001	Level x Target	3.05	716.99	.002
Sexual desire	Intercept	9.98	298.62	< .001	Intercept	9.12	239.26	< . 001
	Level	15.07	731.61	< .001	Level	10.67	686.22	< .001
	Target	-3.42	758.59	< .001	Target	-3.78	702.03	< .001
	Level x Target	0.01	731.61	.990	Level x Target	2.06	686.22	.040
Sexual fantasies	Intercept	12.19	243.02	< .001	Intercept	12.48	217.01	< .001
	Level	12.56	657.52	< .001	Level	9.95	660.75	< .001
	Target	-3.87	676.33	< .001	Target	-3.26	675.55	.001
	Level x Target	2.27	657.52	.024	Level x Target	1.53	660.75	.127

Note. Positive interaction coefficients indicate that the perceived societal devaluation of levels below the ILSA was more pronounced for male targets (coded as 1) than for female (coded as 0) targets.

Table S7

Intergender Effects: Robustness Analyses Including Participant Gender

Outcome	Effect	$F(1, df_2)$	p	ω_p
Sex partners (S1)	Participant gender	3.97	.048	.02
	Level	20.80	< .001	.10
	Target	11.58	< .001	.06
	Participant gender x Level	0.00	.958	.00
	Participant gender x Target	0.69	.406	.00
	Level x Target	187.12	< .001	.51
	Participant gender x Level x Target	0.20	.653	.00
Sex partners (S2)	Participant gender	2.40	.123	.01
	Level	2.36	.126	.01
	Target	3.69	.057	.02
	Participant gender x Level	0.74	.391	.00
	Participant gender x Target	0.15	.703	.00
	Level x Target	169.84	< .001	.50
	Participant gender x Level x Target	1.55	.216	.00
Casual sex partners	Participant gender	0.79	.375	.00
	Level	378.09	< .001	.52
	Target	34.05	< .001	.09
	Participant gender x Level	0.02	.887	.00
	Participant gender x Target	0.02	.902	.00
	Level x Target	205.62	< .001	.37
	Participant gender x Level x Target	0.73	.394	.00
Sexual debut	Participant gender	6.56	.011	.02
	Level	221.05	< .001	.39
	Target	32.45	< .001	.08
	Participant gender x Level	0.79	.375	.00
	Participant gender x Target	0.34	.559	.00
	Level x Target	163.16	< .001	.32
	Participant gender x Level x Target	0.15	.697	.00

Table S7 [*continued*]

Outcome	Effect	$F(1, df_2)$	p	ω_p
Sexual intercourse	Participant gender	3.00	.084	.01
	Level	192.37	< .001	.36
	Target	28.26	< .001	.07
	Participant gender x Level	3.08	.080	.01
	Participant gender x Target	8.05	.005	.02
	Level x Target	101.91	< .001	.23
	Participant gender x Level x Target	3.81	.052	.01
Masturbation	Participant gender	0.01	.941	.00
	Level	108.47	< .001	.24
	Target	0.01	.904	.00
	Participant gender x Level	0.17	.678	.00
	Participant gender x Target	2.30	.130	.00
	Level x Target	168.18	< .001	.33
	Participant gender x Level x Target	12.12	< .001	.03
Sexual desire	Participant gender	0.20	.652	.00
	Level	64.63	< .001	.16
	Target	1.28	.259	.00
	Participant gender x Level	0.02	.888	.00
	Participant gender x Target	0.08	.775	.00
	Level x Target	91.17	< .001	.21
	Participant gender x Level x Target	3.98	.047	.01
Sexual fantasies	Participant gender	0.00	.978	.00
	Level	146.59	< .001	.30
	Target	0.58	.446	.00
	Participant gender x Level	1.15	.284	.00
	Participant gender x Target	0.31	.580	.00
	Level x Target	74.48	< .001	.18
	Participant gender x Level x Target	16.63	< .001	.04

Note. df_2 = 173 (Sex partners, S1), 164 (Sex partners, S2), or 339 (all other outcomes).

Table S8

Intragender Effects: Robustness Analyses Including Participant Gender

	Female targets						Male targets					
	B	*SE*	*df*	*t*	*p*	$R_w^{2\,(f1)}$	*B*	*SE*	*df*	*t*	*p*	$R_w^{2\,(f1)}$
Sex partners (S1)						.31						.30
Intercept	0.77	0.15	173	4.96	< .001		1.32	0.14	173	9.33	< .001	
Level	-3.39	0.21	1046	-15.93	< .001		3.50	0.23	1046	15.55	< .001	
Level²	-11.50	0.74	1046	-15.58	< .001		-12.18	0.78	1046	-15.62	< .001	
Gender	0.25	0.17	173	1.63	.105		0.11	0.14	173	0.77	.443	
Level x Gender	0.30	0.21	1046	1.42	.156		0.08	0.23	1046	0.35	.723	
Level² x Gender	0.31	0.74	1046	0.42	.676		0.47	0.78	1046	0.61	.545	
Sex partners (S2)						.37						.20
Intercept	-0.35	0.16	164	-2.17	.032		0.63	0.16	164	3.84	< .001	
Level	-5.29	0.22	992	-24.17	< .001		1.41	0.23	992	6.24	< .001	
Level²	-7.43	0.76	992	-9.80	< .001		-12.26	0.79	992	-15.60	< .001	
Gender	-0.01	0.16	164	-0.05	.964		0.03	0.16	164	0.16	.870	
Level x Gender	-0.28	0.22	992	1.29	.197		0.10	0.23	992	0.40	.689	
Level² x Gender	1.92	0.76	992	2.53	.012		1.19	0.79	992	1.51	.131	
Casual sex partners						.55						.20
Intercept	-0.02	0.12	339	-0.20	.845		1.25	0.12	339	10.38	< .001	
Level	-6.81	0.13	2042	-53.27	< .001		-2.22	0.15	2042	-15.06	< .001	
Level²	1.63	0.44	2042	3.69	< .001		-2.64	0.51	2042	-5.17	< .001	
Gender	-0.09	0.12	339	-0.75	.457		-0.22	0.12	339	-1.81	.072	
Level x Gender	0.16	0.13	2042	1.22	.224		-0.04	0.15	2042	-0.30	.766	
Level² x Gender	0.08	0.44	2042	0.17	.863		0.76	0.51	2042	1.48	.138	
Sexual debut						.49						.34
Intercept	-0.25	0.09	339	-2.85	.005		-0.69	0.09	339	-7.74	< .001	
Level	-4.70	0.16	2042	-30.09	< .001		-0.55	0.17	2042	-3.27	.001	
Level²	-19.82	0.54	2042	-36.67	< .001		-19.79	0.58	2042	-34.24	< .001	
Gender	0.05	0.09	339	0.61	.542		0.18	0.09	339	2.01	.045	
Level x Gender	-0.17	0.16	2042	-1.09	.276		-0.38	0.17	2042	-2.27	.024	
Level² x Gender	1.71	0.54	2042	3.16	.002		0.58	0.58	2042	1.54	.124	

Table S8 [*continued*]

	Female targets						Male targets					
	B	*SE*	*df*	*t*	*p*	$R_w^{2\,(f1)}$	*B*	*SE*	*df*	*t*	*p*	$R_w^{2\,(f1)}$
Sexual intercourse						.22						.33
Intercept	0.74	0.11	339	7.00	< .001		1.04	0.10	339	10.53	< .001	
Level	1.71	0.16	2042	10.99	< .001		3.81	0.15	2042	24.90	< .001	
Level²	-12.33	0.54	2042	-22.89	< .001		11.55	0.53	2042	-21.81	< .001	
Gender	0.07	0.11	339	0.66	.508		-0.02	0.10	339	-0.22	.828	
Level x Gender	0.11	0.16	2042	0.69	.489		-0.46	0.15	2042	-2.98	.003	
Level² x Gender	1.15	0.54	2042	2.73	.006		0.80	0.53	2042	1.52	.129	
Masturbation						.36						.10
Intercept	-0.04	0.11	339	-0.36	.722		0.59	0.10	339	5.68	< .001	
Level	-4.90	0.14	2042	-34.75	< .001		-1.07	0.15	2042	-7.08	< .001	
Level²	-2.78	0.49	2042	-5.69	< .001		-7.43	0.52	2042	-14.20	< .001	
Gender	0.22	0.11	339	2.14	.033		-0.06	0.10	339	-0.56	.577	
Level x Gender	0.63	0.14	2042	4.46	< .001		-0.38	0.15	2042	-2.50	.013	
Level² x Gender	-1.11	0.49	2042	-2.27	.024		-0.06	0.52	2042	-0.11	.910	
Sexual desire						.31						.18
Intercept	-0.18	0.10	339	-1.90	.058		0.24	0.10	339	2.34	.020	
Level	-4.47	0.15	2042	-29.29	< .001		-1.86	0.16	2042	-11.83	< .001	
Level²	-6.52	0.53	2042	-12.34	< .001		-10.45	0.55	2042	-19.14	< .001	
Gender	0.05	0.10	339	0.50	.615		-0.08	0.10	339	-0.80	.426	
Level x Gender	0.43	0.15	2042	2.81	.005		-0.35	0.16	2042	-2.21	.027	
Level² x Gender	-0.26	0.53	2042	-0.49	.622		0.86	0.55	2042	1.58	.115	
Sexual fantasies						.35						.16
Intercept	0.06	0.11	339	0.52	.601		0.50	0.11	339	4.57	< .001	
Level	-4.71	0.14	2042	-34.03	< .001		-2.49	0.15	2042	-17.16	< .001	
Level²	-2.89	0.48	2042	-6.02	< .001		-6.55	0.50	2042	-13.05	< .001	
Gender	0.21	0.11	339	1.89	.059		0.01	0.11	339	0.07	.948	
Level x Gender	0.66	0.14	2042	4.77	< .001		-0.36	0.15	2042	-2.47	.014	
Level² x Gender	-1.26	0.48	2042	-2.64	.008		-0.26	0.50	2042	-0.51	.607	

Note. Gender = participant gender; Level² = quadratic term.

Table S9

ILSA Comparison between Male and Female Targets: Robustness Analyses Including Participant Gender

Outcome	Effect	$F(1, df_2)$	p	ω_p
Sex partners (S1)	Participant gender	1.17	.282	.00
	Target	145.55	< .001	.45
	Participant gender x Target	0.15	.700	.00
Sex partners (S2)	Participant gender	0.01	.921	.00
	Target	99.38	< .001	.37
	Participant gender x Target	0.58	.448	.00
Casual sex partners	Participant gender	1.02	.313	.00
	Target	134.29	< .001	.28
	Participant gender x Target	0.47	.491	.00
Sexual debut	Participant gender	9.28	.003	.02
	Target	79.87	< .001	.19
	Participant gender x Target	0.03	.874	.00
Sexual intercourse	Participant gender	0.05	.831	.00
	Target	43.14	< .001	.11
	Participant gender x Target	5.00	.026	.01
Masturbation	Participant gender	1.65	.200	.00
	Target	112.31	< .001	.25
	Participant gender x Target	14.11	< .001	.04
Sexual desire	Participant gender	0.34	.558	.00
	Target	50.78	< .001	.13
	Participant gender x Target	2.44	.119	.00
Sexual fantasies	Participant gender	5.59	.019	.01
	Target	35.50	< .001	.09
	Participant gender x Target	18.14	< .001	.05

Note. df_2 = 173 (Sex partners, S1), 164 (Sex partners, S2), or 339 (all other outcomes).

Table S10

Asymmetry in the Devaluation of Values Deviating from the ILSA: Robustness Analyses Including Participant Gender

	Levels greater than the ILSA				Levels lower than the ILSA			
Outcome	Effect	*t*	*df*	*p*	Effect	*t*	*df*	*p*
Sex partners (S1)	Intercept	5.35	197.55	< .001	Intercept	13.54	331.86	< .001
	Target	15.57	1298.77	< .001	Target	-6.97	1055.91	< .001
	Level	-37.96	1279.06	< .001	Level	18.39	1018.01	< .001
	Gender	1.50	197.55	.135	Gender	1.05	331.86	.293
	Target x Level	3.56	1279.06	< .001	Target x Level	-3.19	1018.01	.002
	Target x Gender	-0.58	1298.77	.562	Target x Gender	-0.89	1055.91	.376
	Level x Gender	1.23	1279.06	.221	Level x Gender	-1.67	1018.01	.095
	Target x Level x Gender	0.34	1279.06	.737	Target x Level x Gender	1.16	1018.01	.247
Sex partners (S2)	Intercept	-2.49	181.55	.014	Intercept	9.40	348.59	< .001
	Target	17.32	1427.43	< .001	Target	-5.90	830.18	< .001
	Level	-42.88	1408.43	< .001	Level	12.98	791.27	< .001
	Gender	-0.11	181.55	.916	Gender	1.62	348.59	.107
	Target x Level	6.78	1408.61	< .001	Target x Level	-1.42	791.27	.155
	Target x Gender	1.09	1427.43	.278	Target x Gender	-1.51	830.18	.132
	Level x Gender	0.71	1408.61	.478	Level x Gender	-1.74	791.27	.082
	Target x Level x Gender	-0.05	1408.61	.956	Target x Level x Gender	1.00	791.27	.320
Casual sex partners	Intercept	-1.51	379.27	.131	Intercept	25.46	548.20	< .001
	Target	23.82	3473.84	< .001	Target	-8.67	1200.80	< .001
	Level	-63.87	3439.60	< .001	Level	7.84	1157.80	< .001
	Gender	-0.95	379.27	.343	Gender	-1.24	548.20	.217
	Target x Level	9.49	3439.60	< .001	Target x Level	1.88	1157.80	.060
	Target x Gender	-1.92	3473.84	.055	Target x Gender	0.65	1200.80	.514
	Level x Gender	0.00	3439.60	.998	Level x Gender	-0.71	1157.80	.477
	Target x Level x Gender	1.02	3439.60	.309	Target x Level x Gender	-0.00	1157.80	.997
Sexual debut	Intercept	-8.09	506.22	< .001	Intercept	11.42	412.56	< .001
	Target	4.81	2347.99	< .001	Target	-18.14	2259.44	< .001
	Level	-58.39	2314.71	< .001	Level	34.45	2243.52	< .001
	Gender	0.84	506.22	.404	Gender	0.93	412.56	.353
	Target x Level	-0.04	2314.71	.970	Target x Level	4.00	2243.52	< .001
	Target x Gender	0.37	2347.99	.714	Target x Gender	1.91	2259.44	.056
	Level x Gender	4.99	2314.71	< .001	Level x Gender	-0.32	2243.52	.748
	Target x Level x Gender	0.41	2314.71	.684	Target x Level x Gender	0.23	2243.52	.821

Table S10 [*continued*]

	Levels greater than the ILSA				Levels lower than the ILSA			
Outcome	Effect	*t*	*df*	*p*	Effect	*t*	*df*	*p*
Sexual intercourse	Intercept	13.72	381.13	< .001	Intercept	8.28	426.84	< .001
	Target	10.39	2084.25	< .001	Target	0.08	2475.12	.936
	Level	-34.82	2070.79	< .001	Level	43.49	2455.60	< .001
	Gender	0.90	381.13	.367	Gender	1.23	426.84	.221
	Target x Level	3.69	2070.79	< .001	Target x Level	0.98	2455.60	.329
	Target x Gender	-3.45	2084.25	< .001	Target x Gender	-0.44	2475.12	.660
	Level x Gender	1.96	2070.79	.050	Level x Gender	-2.26	2243.52	.024
	Target x Level x Gender	-0.65	2070.79	.517	Target x Level x Gender	0.64	2243.52	.523
Masturbation	Intercept	-1.67	376.47	.095	Intercept	17.74	475.21	< .001
	Target	16.90	3093.51	< .001	Target	-9.13	1507.98	< .001
	Level	-56.18	3071.77	< .001	Level	14.72	1472.14	< .001
	Gender	2.20	376.47	.028	Gender	-1.08	475.21	.280
	Target x Level	6.11	3071.77	< .001	Target x Level	4.57	1472.14	< .001
	Target x Gender	-5.54	3093.51	< .001	Target x Gender	1.93	1507.98	.054
	Level x Gender	0.34	3071.77	.736	Level x Gender	-1.16	1472.14	.248
	Target x Level x Gender	-0.72	3071.77	.473	Target x Level x Gender	0.67	1472.14	.505
Sexual desire	Intercept	-2.95	388.44	.003	Intercept	13.19	527.44	< .001
	Target	11.79	3112.81	< .001	Target	-4.84	1454.30	< .001
	Level	-55.83	3090.91	< .001	Level	17.86	1414.00	< .001
	Gender	0.52	388.44	.601	Gender	-0.15	527.44	.882
	Target x Level	2.76	3090.91	.006	Target x Level	1.71	1414.00	.089
	Target x Gender	-3.55	3112.81	< .001	Target x Gender	1.10	1454.30	.271
	Level x Gender	1.04	3090.91	.301	Level x Gender	0.19	1414.00	.846
	Target x Level x Gender	-1.16	3090.91	.244	Target x Level x Gender	-0.62	1414.00	.537
Sexual fantasies	Intercept	-0.60	377.06	.551	Intercept	17.36	455.98	< .001
	Target	10.52	3232.31	< .001	Target	-4.92	1347.64	< .001
	Level	-56.91	3212.98	< .001	Level	15.51	1315.31	< .001
	Gender	2.18	377.06	.030	Gender	-0.75	455.98	.457
	Target x Level	4.00	3212.98	< .001	Target x Level	2.33	1315.31	.020
	Target x Gender	-3.90	3232.31	< .001	Target x Gender	2.02	1347.64	.044
	Level x Gender	-1.18	3212.98	.237	Level x Gender	-3.12	1315.31	.002
	Target x Level x Gender	-0.63	3212.98	.530	Target x Level x Gender	1.67	1315.31	.096

Note. Positive Target x Level interactions indicate that compared to female targets (coded as 0), the perceived societal devaluation of male targets (coded as 1) was (a) less pronounced for activity level greater than the ILSA and (b) more pronounced for activity levels below the ILSA.

General Discussion

Sexuality is one of the most universally significant aspects of life, with sexual motivation playing a crucial role in shaping people's sexual experiences and well-being. Self-report measures are the primary way to study sexual motivation and have contributed greatly to a better understanding of human sexuality, but their validity has been questioned. The present research aimed to promote valid measurement of sexual motivation by identifying and examining major threats to validity. In Part I, my colleagues and I developed and validated the Trait Sexual Motivation Scale (TSMS) based on a theoretical conceptualization of sexual motivation (Frankenbach et al., 2022). In four preregistered studies, we have rigorously validated the TSMS in accordance with recommended practices for scale validation (Loevinger, 1957; Simms, 2008). These theoretical and empirical steps aimed to ensure that we measure what we intend to measure, thereby counteracting the *defining threat* to validity. We further tested measurement invariance across gender and relationship status, thereby addressing the *delicate threat* that differences between groups (e.g., gender differences) may arise from different psychometric properties for these groups (i.e., "comparing apples with oranges"). Part II focused on how social desirability bias—people's tendency to present themselves in ways that are valued by others—undermines the valid measurement of sexual motivation. Although online surveys that do not require direct human interaction have become the norm in today's assessment of sexuality, evidence for social desirability bias in sexual self-reports and gender differences therein comes mostly from laboratory studies in which other people were present. We have rigorously examined this *social threat* to validity using the Item Sum Technique and three additional approaches. Part III followed up on this by examining the social norms that are thought to give rise to social desirability bias. The strong sexual double standard, which posits opposing social norms for men and women, pervades academic and everyday discussions about sexuality, but it has received

little support. We proposed and tested a distinctly different perspective on sexual double standards, predicting that male and female norms are characterized by both similarities and differences.

Summary of the Main Results

In Part I, the TSMS showed good model fit, correlated with other scales in expected ways, and predicted sexual criteria both cross-sectionally and prospectively. It was internally consistent and stable over four weeks and three months. Scalar measurement invariance for gender and relationship status indicates that the TSMS can be used with and compared between (1) men and women, and (2) people who are single and romantically involved. In summary, the economic and theory-driven TSMS emerges as a reliable and valid measure of sexual motivation. Using the TSMS, we found that on average, men report higher trait sexual motivation than women.

In Part II, self-reported sexual motivation and gender differences in sexual motivation did not differ significantly between a standard direct questioning (DQ) online group and an item sum (IS) group for which honest responses are particularly likely (*indirect* approach). While gender differences in sexual motivation were consistent and moderate to large, gender differences in sexual bias indicators were inconsistent and small (*logic* approach). Honesty ratings were close to maximum in both the DQ group and the IS group (*subjective* approach). Controlling for associations with social desirability scale scores, which were close to zero, did not affect gender differences in sexual motivation (*control* approach). In conclusion, these results provide little evidence that self-reported sexual motivation in general, and gender differences in particular, are substantially affected by social desirability bias in online surveys.

In Part III, we found robust evidence for *inter*gender differences and for *intra*gender similarities in perceived societal norms. Replicating and generalizing previous research, participants perceived that high levels of sexual activity are evaluated more favorably for men

than for women. Extending previous research, participants perceived that low levels of sexual activity are evaluated more favorably for women than for men. At the same time, however, participants perceived that society most values moderate sexual activity in both men and women. This coexistence of similarities and differences is not consistent with any of the models proposed in the past. It is a pattern that is uniquely predicted by the S&D model proposed here.

Are Gender Differences Real? Integrating the Results into the Literature

Understanding whether true gender differences exist has important implications for everyday life. People living in monogamous relationships face a sexual interdependence dilemma when they are discrepant in their sexual motivation: There is a mutual dependency in the satisfaction of interpersonal desires, with at least one partner having to put their needs second. This may contribute to lower sexual satisfaction, as not all couples find ways to successfully manage these discrepancies (Day et al., 2015). The greater the overall difference in sexual motivation, the more likely it is that sexual interdependence dilemmas are a common and significant challenge for monogamous male-female relationships.

The literature is clear: on average, men report higher sexual motivation than women. What is controversial, however, is the critical question of whether these measured differences are real or the product of bias. The present research adds to this lively debate by addressing critical threats to validity. We found pronounced gender differences, with men consistently reporting more frequent sexual fantasies, desires, and self-stimulation. These gender differences occurred in Part I, where we used the comprehensively validated TSMS that has equivalent psychometric properties for men and women. Gender differences were also found in Part II, both in the standard online survey and in the item sum group, in which honest responding is particularly likely. Results from the additional approaches to validity employed in Part II support the

conclusion that gender differences are unlikely to be the result of the major threats to validity examined in this dissertation.

Different Approaches, Divergent Results: Discussing Salient Inconsistencies

This conclusion stands in stark contrast to research suggesting that gender differences in sexual motivation are largely driven by biases, particularly social desirability bias. For example, a recent article concluded that "[t]he fact that self-report studies consistently find that women have lower sexual motivation than men can be explained by conscious manipulations of self-perceived motivation according to sexual scripts" (Touraille & Ågmo, 2024).

Indeed, gender differences were less pronounced and less consistent when other methodological approaches were used, such as implicit measures or physiological measures (e.g., changes in penile circumference or vaginal blood volume; see Touraille & Ågmo, 2024). How do these findings fit together? In the following, I'd like to offer some thoughts on how methodological approaches can be evaluated and how conflicting results from different methodological approaches can be explained and, to some extent, reconciled.

First, while it is essential to examine bias in sexual self-report, other methods may also be susceptible to threats to validity. The three threats identified in this thesis can serve as a valuable blueprint for demonstrating this point. Physiological measures, for example, may be unaffected (or little affected) by the *social threat* of altering responses to conform to societal norms. However, physiological measures neglect the psychological element (i.e., the feeling of wanting) that is considered a defining element of the construct and distinguishes it from (genital) sexual arousal (Levine, 2003; Mitchell et al., 2014). This introduces the *defining threat* of not measuring the construct in accordance with the theory. In addition, anatomical and physiological differences complicate the comparison of male and female genital responses. Thus, any attempt to compare bodily responses is likely to suffer from the *delicate threat* of comparing apples with oranges.

Because the true magnitude of gender differences is unknown, the validity of an approach and its findings cannot be inferred from the empirical results (e.g., whether or not we find gender differences). Instead, validity should be inferred from a critical reflection on whether the measured phenomena are likely to reflect the truth. Although not exhaustive, the criteria derived from the threats to validity examined in this dissertation may provide a solid foundation for further exploration.

Second, different approaches that meet all these (and other) criteria for valid measurement may still lead to varying conclusions. This occurs because validity is not an inherent property but is evaluated within the context of the questions that researchers ask and the assumptions that they make. For instance, a picture-story measure of sexual motivation has recently been developed and thoroughly validated as an alternative to sexual self-report scales (Hinzmann et al., 2023; Schultheiss et al., 2023). Different from our results, the authors did not find higher male sexual motivation. As briefly discussed in Part II, these findings are only discrepant at first sight. The measure was built on a different approach to validity, which suggests that a measure is valid if it is responsive to experimental variation of the underlying construct (Borsboom et al., 2004, 2009). For the picture-story measure, this was demonstrated in showing that participants' stories contained more sexual content when they were exposed to sexual erotic prime pictures than to neutral prime pictures (Hinzmann et al., 2023). These pictures were specifically selected to ensure that they are equally sexually stimulating to all (male and female) participants.

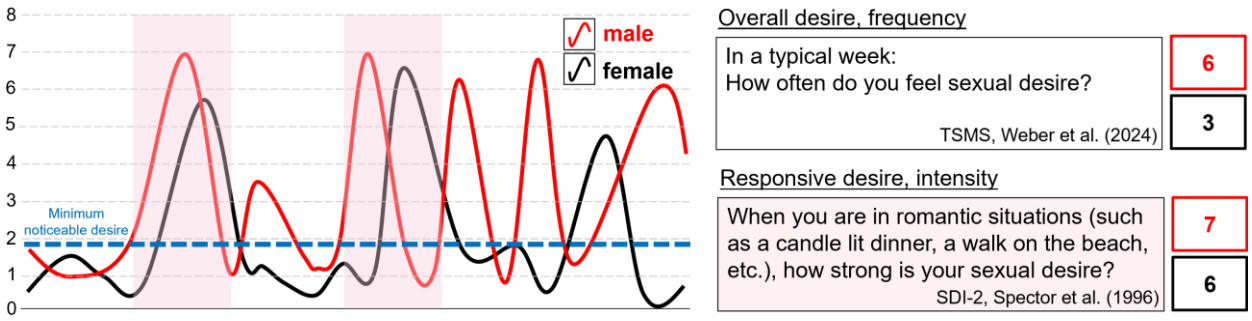
I believe that the absence of gender differences under these conditions and the pronounced and robust findings that we found are perfectly reconcilable. When several people of any gender are in situations that they all find equally sexually stimulating, they are likely to show similar responsive sexual desire. At the same time, people higher and lower in their overall level of trait sexual motivation may likely differ in how frequently they seek or encounter situations

that elicit sexual motivation. Within such a framework, men's higher average trait sexual motivation may be reframed as a broader range or frequency of stimuli that are sexually relevant (Ågmo & Laan, 2022; Toates, 2009).

This integrated perspective may help explain why more pronounced gender differences were found in the frequency rather than the intensity of sexual events (Frankenbach et al., 2022). While frequency counts typically include all events (e.g., "In a typical week: How often do you feel sexual desire?," Weber et al., 2024), intensity ratings are often conditional, tied to specific situations such as the presence of a relationship partner (e.g., "When you are in romantic situations [such as a candle lit [*sic*] dinner, a walk on the beach, etc.], how strong is your sexual desire?", Spector et al., 1996). These situations are likely to be particularly stimulating for both men and women, and therefore may result in smaller gender differences—similar to when participants are experimentally exposed to equally sexually stimulating stimuli (Schultheiss et al., 2023). Figure 1 illustrates this dissociation between overall frequency and responsive intensity.

Figure 1

Gender Differences in Sexual Motivation: Overall Frequency and Responsive Intensity



Note. Levels of sexual desire for a fictional mixed-gender romantic couple over the course of one week. Gender differences are more pronounced in the overall frequency of noticeable sexual desire (6 vs. 3 times a week) than in the intensity of responsive sexual desire in romantic situations (7 vs. 6 on a scale ranging from 0 to 8).

Imagine Tila and Lorne are the average mixed-gender couple. If we could reliably track their sexual desire as an indicator of sexual motivation over a typical week, the gender differences in their responses would likely depend on the researchers' focus. If the focus is on overall trait sexual motivation, using the average frequency of sexual desire as a valid indicator (Frankenbach et al., 2022), men are likely to report higher levels than women. However, if the focus is on responsive sexual motivation in romantic situations, the gender differences are likely to be less pronounced (Dawson & Chivers, 2014).

Different Study Characteristics, Similar Results: Discussing Surprising Consistencies

The previous section sought to explain how different approaches shape different gender differences. At second glance, the consistency of gender differences across studies using the same approach (i.e., self-reports on overall trait sexual motivation) is no less surprising. In particular, the gender differences we obtained under conditions designed to maximize the validity of the findings (i.e., measures reflect the construct, scalar measurement invariance for gender, high anonymity) were within the range of gender differences found in previous literature, which also includes effects from less "ideal" contexts (Baumeister et al., 2001; Frankenbach et al., 2022).

One interpretation of these findings is that these biases could threaten the validity of measured gender differences in theory, but that the contextual features that are needed for substantial self-presentation to occur (e.g., low anonymity, high risk of social backlash) are the exception rather than the norm in scientific studies. The exposure threat condition in Fisher and colleagues' seminal Bogus pipeline series, in which participants were led to believe that an experimenter next door might see their responses (Alexander & Fisher, 2003; Fisher, 2013; Fisher & Brunell, 2014), may be such an exception.

A second interpretation is that a significant portion of the literature *is* influenced by social desirability bias, yet self-presentation may be less gendered than commonly assumed. In what

follows, I will integrate the findings from Parts I trough III into a new model of flexible self-presentation. This model postulates a dualism of competing motivations and offers an explanation for why differences in perceived societal norms for men and women do not necessarily lead to gendered self-presentation tendencies.

Scientific Implications: Towards a Model of Flexible Self-Presentation

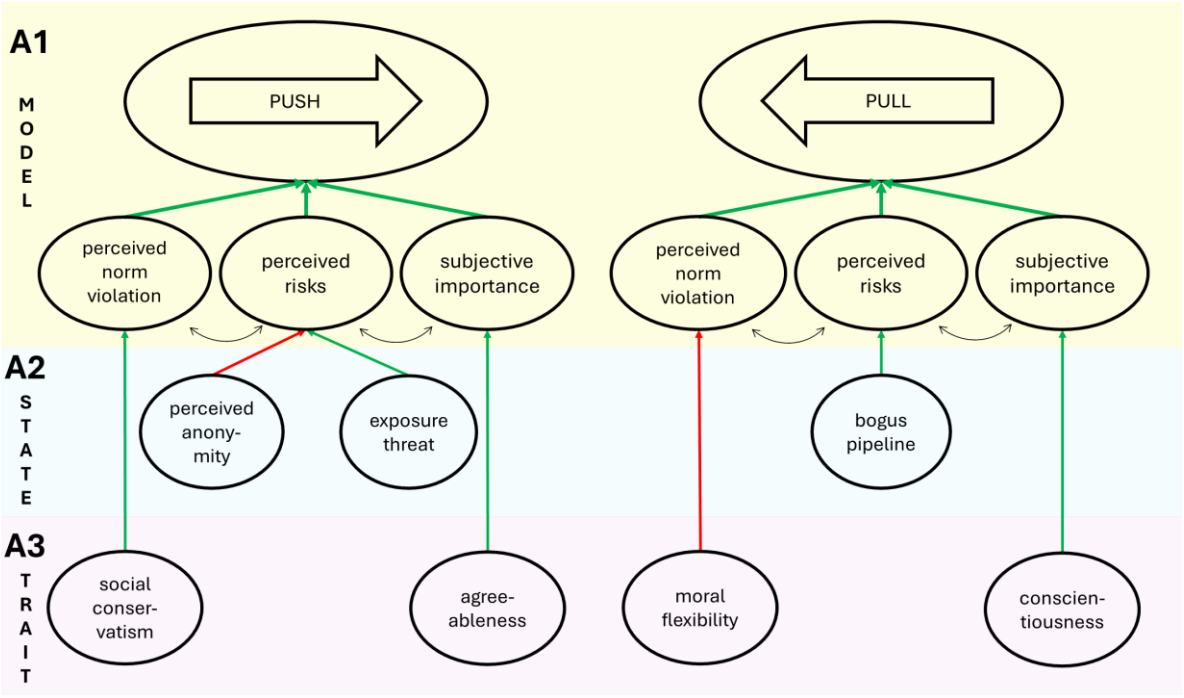
Human beings are deeply social creatures. Being valued and included by others are universal human needs, and their fulfillment is critical to well-being (Baumeister & Leary, 2007; Deci & Ryan, 2012). Therefore, people are motivated to avoid behaviors that might cast doubt on their suitability as trustworthy companions, as these behaviors can lead to rejection and social exclusion (Baumeister et al., 2007). Violations of written or unwritten norms for socially sensitive behaviors are particularly critical threats to a person's integrity, with sexuality being at the forefront of these sensitive matters. For example, individuals with many and varied sexual partners might be perceived as untrustworthy and low in self-control, reducing their social value, especially as potential long-term relationship partners. Even non-behavioral precursors to these activities, such as high and non-partner-specific sexual desire, may raise concerns and lead to social rejection.

It therefore comes as no surprise that people may not always tell the truth when asked sensitive questions. If they perceive that their true response might violate social norms and lead to social sanctions, they may consider reporting a different value. I suspect that the extent to which they conform their responses to social norms depends on the relative strength of two competing motivational forces: one pushing them away from their true response, facilitating self-presentation ("push motivation"), and one pulling them towards their true response, counteracting self-presentation ("pull motivation"). This dualism forms the core of my working model of flexible self-presentation, which is illustrated in Figure 2.

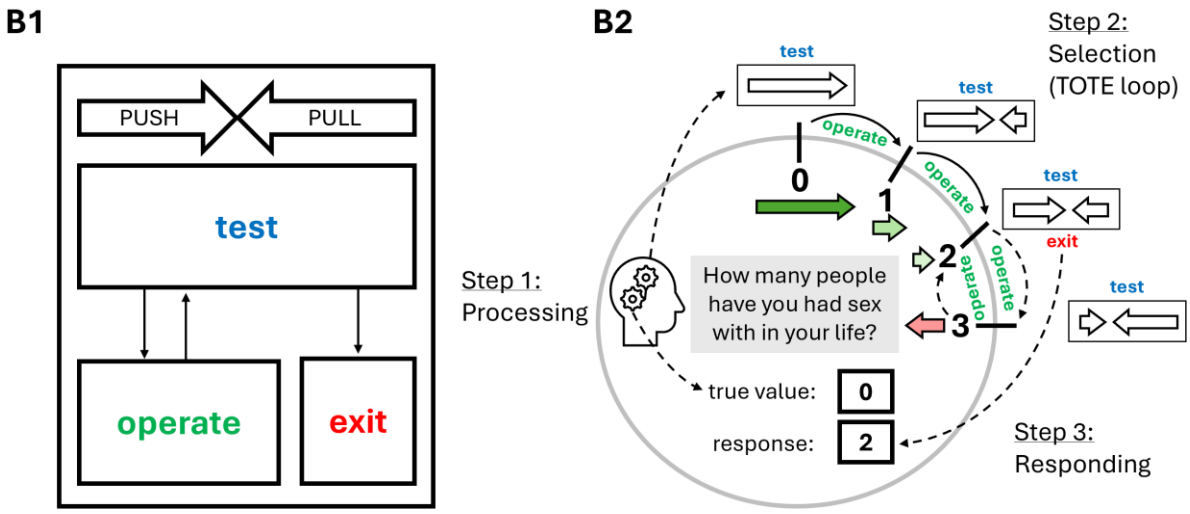
Figure 2

Model of Flexible Self-Presentation

A. Conceptual Model (A1) and Potential Predictors at the State (A2) and Trait Level (A3)



B. Process Model: General Concept (B1) and Demonstration for One Participant (B2)



Note. The conceptual part of the model (Figure 2A) postulates that self-presentation for sensitive questions (i.e., whether and how far people adjust away from their true value) depends on the strength of two competing motivations: one pushing them away from their true response, thereby

facilitating self-presentation (“push motivation) and one pulling them towards their true response, thereby counteracting self-presentation (“pull motivation”). Each three antecedents determine the strength of these motivations: perceived norm violation, perceived risks, and subjective importance (Panel A1). These antecedents are affected by state-level contextual factors (Panel A2) and trait-level dispositional factors (Panel A3).

The process of how people select their responses is shown in Figure 2B. It is assumed that response selection follows a test-operate-test-exit circle as specified in cybernetic models of motivation and behavior (Panel B1). Panel B2 illustrates the three-step process for Robert, a fictional participant who is asked a sensitive question (“How many people have you had sex with in your life?”). In a first processing step, Robert searches his memory for the true value, coactivating knowledge of perceived societal norms to prevent deprivation of social needs (e.g., the need to belong). The critical second step of response selection follows a consecutive loop of comparisons (“test”), starting with a comparison of the strength of the push motivation against zero for the true value. Because his push motivation is positive, he moves to the next value (“operate”) for comparison (“1”). For non-true values, a pull motivation is defined, and the comparison is between the push and pull motivation for that value. The test-operate loop continues as long as the net motivation (i.e., push motivation minus pull motivation) remains positive (green arrows in Panel B2, values 0 to 2). A negative net motivation (red arrow in Panel B2, value 3) signals that the loop should be ended, and that the participant should select the last value for which the net motivation has been positive (“exit”). This value is then reported in the third step.

The conceptual part of this model outlines the antecedents likely to determine the strength of each motivation (Panel A1). It assumes that the push motivation is stronger when a person (1)

perceives their potential response as being evaluated negatively by others (i.e., perceived norm violation), (2) believes that their response could actually lead to negative consequences such as social backlash (i.e., perceived risks), and (3) considers the consequences to be important (i.e., subjective importance). For instance, let's return to Lorne and Tila who participate in a survey in which they are asked to report how often they feel sexual desire in a typical week. The initial push motivation to indicate a value other than the true one is not determined by the true value per se but by the combination of these antecedents. A multiplicative function is logical, since the push motivation should be zero when any of these components is zero.

For example, suppose both Lorne and Tila care about making a good impression and believe that their true values are not perfectly consistent with the societal ideal for their gender. The interplay of these factors causes discomfort in reporting their true values, leading to a motivation for adjustment. However, a strong push motivation will only emerge if they believe that being honest could result in negative consequences. Tila feels completely anonymous in the study, so her perception of risk—and thus her push motivation—is minimal. Lorne, on the other hand, *is* concerned about what the researchers might think of him, so his push motivation is strong. Does this mean he will likely report a value other than the truth, perhaps the one he believes the researchers will most approve of?

Not necessarily, because not telling the truth carries its own personal and social costs. Lying is a significant norm violation that can diminish an individual's social value and conflict with a person's (desired) self-concept of being an honest person (Mazar et al., 2008). I therefore propose that a pull motivation counteracts self-presentation, likely influenced by factors similar to those driving the push motivation. Specifically, the pull motivation might be stronger if a person (1) perceives their potential response as conflicting with their own norms or those of others (i.e., perceived norm violation), (2) believes that their response could lead to negative

consequences such as social backlash (i.e., perceived risks), and (3) places significant importance on these consequences (i.e., subjective importance). For instance, Lorne believes that society disapproves of lies, values being seen as an honest person, and views lying as a critical threat to his self-concept as an authentic individual and a good participant. Because his pull motivation outweighs his push motivation, he reports his true value (six desires per week) in the study.

I believe this model has several important strengths. First, it is kept simple, with just two competing motivations, and parsimonious, with three primary antecedents for each motivation. Second, the idea of competing motivations may explain why participants often report true values or values close to their true ones in surveys rather than reporting the perceived ideal value. Third, while the model is inspired by my research interest in sexual motivation, it is neither limited to research contexts nor to the domain of sexuality. Fourth, the model may be integrated into the long tradition of expectancy-value models in psychology (for a review, see Feather, 2021), with subjective importance reflecting the value component, and perceived norm violation and perceived risks shaping the expectancy that a response will lead to negative consequences. Finally, the model allows for the inclusion of predictors at both the state level (i.e., contextual factors) and the trait level (i.e., dispositional characteristics).

At the state level, the model can help explain how contextual manipulations used in previous research and in Part II of the present research influence self-presentation. This is illustrated in Panel A2 of Figure 2. While each manipulation may affect multiple antecedents of the push and pull motivations, the following effects seem particularly plausible. In the bogus pipeline condition, where participants believe any misreporting will be detected, the perceived risks component of the pull motivation is high, encouraging honest responses (Alexander & Fisher, 2003; Fisher & Brunell, 2014). High levels of perceived anonymity (e.g., Part I, Part II) reduce the perceived risks component of the push motivation, as there is no risk of social

backlash if no one can identify the respondent. The item sum (IS) group (Part II) represents an extreme version of this: with no possibility of being associated with their true response because participants never even report their true value in the first place, the perceived risk of social backlash approaches zero, thereby facilitating honest responding. On the other hand, the exposure threat condition (Alexander & Fisher, 2003; Fisher & Brunell, 2014) is nearly the polar opposite. In this scenario, others are present and likely to detect one's response, which strengthens the perceived risk component of the push motivation, promoting socially desirable responding. This underscores that the exposure threat condition differs from the default scenario in contemporary research. It might therefore be more appropriate to view it as an extreme context that highlights the *potential* impact of self-presentation, rather than as a control condition that reflects *typical* misreporting in scientific studies.

While trait predictors have played a minor role in previous research on sexual self-presentation, the conceptual model provides a foundation for generating hypotheses. Although speculative at this stage, some initial ideas are illustrated in Panel A3 of Figure 2. For example, individuals high in social conservatism (i.e., tendency toward traditional worldviews and resistance to change, Stankov, 2018) may have internalized that society judges people very differently depending on their gender and level of sexual activity. Compared to liberal individuals, conservative persons may, for instance, expect a more extreme devaluation of women with multiple sexual partners. This polarization of evaluations could lead to high perceived norm violations, and thus push motivation, for many levels of sexual activity. In addition, individuals with a strong tendency for moral flexibility (i.e., convincing themselves that dishonest behavior is not immoral) may exhibit more pronounced self-presentation, as this weakens the perceived norm violation component of the pull motivation (Gino & Ariely, 2012; Liu et al., 2022). Finally, high agreeableness might enhance the subjective importance component

of the push motivation (i.e., making a good impression), thereby facilitating self-presentation, whereas high conscientiousness might strengthen the subjective importance component of the pull motivation (i.e., the importance of being honest), thereby counteracting self-presentation.

The Process of Response Selection: A Cybernetic Perspective

The conceptual part of this working model suggests that people may select a value other than their true value if their push motivation outweighs their pull motivation. However, what remains to be specified is how the response selection process unfolds in detail. I illustrate my ideas in Figure 2B. I assume that the response selection process follows a test-operate-test-exit (TOTE) cycle, as outlined in cybernetic models of motivation and behavior (Panel B1, e.g., Carver & Scheier, 1981; Powers, 1973). Panel B2 illustrates the three-step process for Robert, a fictional participant who is asked the sensitive question, "How many people have you had sex with in your life?". Robert has been intimate with one girl (oral stimulation) and one boy (mutual genital stimulation), but he tends to describe these encounters as "second base" rather than "having sex." He perceives that society expects men to be sexually active (ideally with 5-10 partners) and fears social backlash for being considered a male virgin if he answers 0.

The critical second step of response selection can be conceptualized as a TOTE loop, starting with a comparison ("test") of the initial push motivation for the true value against zero. For Robert, the push motivation is positive, which is why he shifts his attention to another possible response value that is closer to the societal ideal ("operate"). For every value that deviates from the truth, a pull motivation is defined, and the comparison is between the respective strength of the push and pull motivation for that value. The loop continues as long as the net motivation (i.e., push motivation minus pull motivation) remains positive. When net motivation turns negative, the loop ends, and the person selects the last value with a positive net motivation ("exit").

In Robert's case, net motivation was positive (green arrows in Figure 2B) for the values 0, 1, and 2. Although some motivation pulled him back, Robert's perceived norm violation for these values was not too pronounced, allowing him to rationalize that his sexual experiences might fit the researcher's definition of "having sex." Since he has only shared private moments with two people, this rationalization would be impossible for the value 3, so the pull motivation increases, and the net motivation becomes negative (red arrow in Figure 2B). In the third and final step, Robert selects the last value for which the net motivation was positive: 2.

The fictional case of Robert highlights two additional points. First, missing or vague definitions (e.g., "who counts as a sex partner?") can benefit self-presentation and should be carefully avoided by researchers. Second, although the fictional case featured a combination of several factors that facilitated self-presentation, Robert reported a number that was still well below what he considered an ideal number of sex partners.

Integrating the Present Research into the Model: Perceived Norms and Gender Differences

Results from the present research can be integrated into this working model to better predict sexual self-presentation. Specifically, the combined results of Parts I to III can explain why self-presentation may be less of an issue than widely assumed and why gendered patterns in perceived societal norms do not necessarily have to find expression in gendered socially desirable responding.

First, for self-presentation to occur, participants must fear that their accurate responses will lead to negative consequences such as social backlash. However, the results of Part II suggest that most participants perceive their participation in online surveys to be completely anonymous, so the perceived risk of social backlash is close to zero. According to our model, this should minimize the push motivation and thus the self-presentation motivation of the average participant.

Second, for studies in which perceived risks are *not* close to zero, self-presentation tendencies do not necessarily have to be gendered. If perceived social norms took the form of a strong sexual double standard, then perceived norm violations would follow a fixed pattern: the average man's true value would always be below the ideal, and the average woman's true value would always be above the ideal. This has led to the prediction of male overreporting and female underreporting of sexual events in scientific studies.

The present research can help update these predictions in two ways. In a first step, Part III offers a new perspective on perceived societal norms as a central determinant of self-presentation. On average, evaluations peaked for moderate rather than extreme levels of sexual activity. This suggests that the true level of sexual motivation of any individual, male or female, may be above or below their perceived societal ideal. As a consequence, a possible push motivation can be in the direction of underreporting or overreporting for both men and women. In a second step, the combined consideration of the findings from Parts I to III can explain why, on average, self-presentation tendencies may be similar in direction and strength for men and women. Part III showed that the ILSA for events indicative of sexual motivation is higher for men than for women. The results from Parts I and II found higher values for men than for women for these indicators of sexual motivation, which are unlikely to be due to bias. Taken together, the discrepancy between actual and perceived ideal levels of sexual motivation may be very similar for the average man and woman—and so may be the discrepancy in their evaluations.

This correspondence between descriptive norms ("how is it?") and perceived societal norms as injunctive norms ("how should it be?") may not appear by chance. Perceptions of what is socially rewarded may affect people's sexual motivation, and perceived differences in sexual motivation may in turn be used to construe the perceived societal norms. A closer examination of

how these two elements are related, including how they shape each other over time, is an interesting avenue for future research.

A Final Flexible Feature, First Empirical Evidence, and Future Research

Finally, while perceptions of how society might evaluate sexual events is likely to be important in many situations in research and daily life, perceived societal norms may also vary as a function of contextual factors. For instance, imagine Robert would not report his lifetime sexual partners in a scientific study but in the context of a birthday party with close friends, a locker-room chat with soccer teammates, or on a first date. While making a good impression might be important to Robert in all of these contexts, the perception of *what* is needed to make a favorable impression is likely different in each context, strongly affecting the self-presentation motivation.

For instance, when two persons who are on a date talk about their sexual desire, the context-specific norm may primarily reflect their perceptions of what may make a particularly good impression on their date. If the average man feels that his true level of sexual desire is too high and the average woman feels that her true level of sexual desire is too low to be taken into consideration as a dating partner, then this may even lead to male underreporting and female overreporting. In a yet unpublished series of two online studies, this is indeed what we found. Participants were assured of their anonymity and asked to honestly answer questions about their sexual desire. We then asked them to imagine that they were on a date with an opposite sex stranger who may be a possible relationship partner and to indicate how they would likely adjust their original (true) responses if they were in this context. We told them that they are free to not change their response or change their response in any direction.

The preliminary results are consistent with the working model of self-presentation. First, many participants did not change their response, consistent with the assumption that multiple factors need to be in place for self-presentation to occur (e.g., perceived norm violation for the

true value). Second, for those who did change their responses, overreporting *and* underreporting occurred among both genders, consistent with the idea that true levels can be below or above ideal levels for men and for women. Third, compared to their responses in the anonymous scenario, which are likely to be accurate, participants' average responses in the dating scenario were lower for men and higher for women—a pattern of male underreporting and female overreporting that is inconsistent with a strong sexual double standard but fully consistent with the working model proposed here. Fourth, open-text responses asking participants to explain their changes provide further qualitative evidence for the relevance and contextualized nature of perceived norm violations: Most explanations referred to trying to make a positive impression, often based on the idea that a perception of congruence in sexual desire would be particularly attractive (e.g., male participant: "to not make the impression that I only want sex and no relationship"; female participant: "Because I don't want to 'scare off' a potential partner, because he might like to be sexually active").

Future research may extent this line of research and quantitatively assess perceived norms across a wide range of situations to better estimate their context-specific nature. Such studies could also more systematically manipulate and assess the antecedents of the two competing motivations to evaluate their impact and to critically test and further develop this working model of flexible self-presentation.

Practical and Societal Implications: The Bigger Picture

How Basic Research Can Help Answering Applied (Sexual) Questions

Basic research can aid applied research by ensuring the trustworthiness of its findings. While this dissertation focused on gender differences—a question of great practical relevance and intense scientific debate—a careful examination of (threats to) validity can benefit a wide range of important research questions: How does state sexual motivation fluctuate over time? Do

critical life events such as breakups, new relationships, moving in together, or marriage affect sexual motivation? When and for whom does trait sexual motivation begin to decline? Are there interventions that can increase sexual motivation in both the short and long term? All of these questions have very practical implications for the lives of millions of people, and the results of scientific studies are likely to be disseminated around the world. However, their investigation again calls for measures that are suited for these research questions, which may here include additional qualities such as sensitivity to change and measurement invariance over time. I firmly believe that basic and applied research can stimulate one another, advancing both science and society. This mutual stimulation is especially vital in the sensitive yet relevant area of sexuality.

Reflection on the Potential Clinical Use of the TSMS

An important contribution of current research is to provide the basis for trustworthy applied research. There are two outcomes of the current research that may promise even more direct benefits to society: the TSMS and the S&D model.

This dissertation approached sexual motivation from a personality and social psychological perspective, aiming to deepen our understanding of how it manifests in a population of healthy adults. However, the importance of sexual motivation and its interaction with non-sexual experiences and behaviors has also been long recognized in clinical psychology. For instance, diagnoses covering clinically relevant distress associated with low levels of sexual motivation and arousal are included in the latest editions of both the International Classification of Diseases and Health-Related Disorders (ICD-11; “6C72: Hypoactive Sexual Desire Dysfunction”; World Health Organization, 2019) and in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; “302.72: Female Sexual Interest/Arousal Disorder” and 302.71: Male Hypoactive Sexual Desire Disorder”; American Psychiatric Association, 2022). Demonstrating the dynamics between sexual and non-sexual experience and behavior, changes in sexual

motivation can also appear in the context of more general families of mental disorders, including unipolar and bipolar affective disorders (with reductions during depressive phases and increases during manic phases) and eating disorders (typically reduction in the context of anorexia nervosa and bulimia nervosa). Therefore, I was delighted to read that a clinical expert published a response to our TSMS article (Part I), highlighting the scale's significant clinical potential (Anderson, 2024). Specifically, Anderson emphasized her dissatisfaction with the mismatch between the clinical importance of sexual motivation and its inadequate measurement in clinical settings and suggested that the TSMS could help address this void. I greatly appreciate the positive clinical reception of the TSMS and the effort to connect basic and applied science, as well as to bridge the gap between different psychological subdisciplines.

In the spirit of a dynamic exchange, I'd like to briefly add a few thoughts. Result of our validation process suggests that the TSMS is a valid measure of (1) *inter*individual differences in (2) *trait* sexual motivation among (3) young, mostly *healthy* adults. However, whether the TSMS is also valid when one or more of these features are modified is an open question and requires ongoing validation. For instance, considerations about the scale's intended use were integrated early in the TSMS construction process, such as empirically deriving response categories from a sample of the target population. This approach might limit the scale's applicability, at least in its current form, to clinical (sub)populations. The strategy employed in Part I could serve as a blueprint for the continued validation of the TSMS.

Social Norms: Using Results from Basic Research to Promote a Healthy (Sex) Life

In this dissertation, I considered sexual norms from a validity perspective, focusing on their role as an antecedent of social desirability bias. However, sexual norms may influence more than just discrepancies between measured and actual levels. They are also likely to shape people's actual sexual experiences, including sexual motivation, and have significant downstream

consequences for sexual and mental health. For instance, there is evidence that endorsement of a sexual double standard is associated with a reduced likelihood of condom use (Crawford & Popp, 2003; Lefkowitz et al., 2014) and with shame and guilt following sexual activity among women (Tolman, 2005). Numerous studies have documented that adolescence and early adulthood are critical periods in which gendered sexual norms are likely to be internalized and lead to negative consequences (e.g., Kiefer & Sanchez, 2007; Sanchez et al., 2012).

Previous research has demonstrated that sex education programs—some of which address social norms and perceptions of norms—can help reduce stigma and promote sexual health (for a review, see Abrams et al., 2023). I believe these programs could further benefit from an accurate representation of gendered sexual norms. The present research makes an important contribution by highlighting important similarities and differences in how people perceive relevant others might evaluate expressions of sexuality. Exploring people's actual evaluations, where previous studies have shown minimal differences between men and women, could be a valuable next step. In terms of the dynamic interplay between research and practice, it would be important for future studies to explore whether sexual norms and social desirability bias may be linked in another way: that people may be motivated to present themselves as less judgmental and more gender-equal than they actually are when asked to contrast male and female sexual behavior.

Limitations and Strengths

Generalizability Across Populations

Within each part of this dissertation, the samples studied showed some level of homogeneity that may limit the generalizability of the findings. All samples can be classified as WEIRD (i.e., Western, educated, industrialized, rich, demographic). Finding gender differences across different countries (Part I: US residents, Part II: UK residents) and replicating the more favorable evaluations of high sexuality for men than for equally active men in German samples

(Part III) speak to some generalizability of our findings, at least across the Western world. Similarly, we replicated gender differences and the complex curvilinear pattern for perceived societal norms (i.e., similarities and differences) across student/community samples and crowdsourcing samples that are relatively heterogeneous in terms of their demographic and socioeconomic background (Goodman et al., 2013). In terms of age, all samples were composed of young to middle-aged participants, and most individuals (approximately 80% across our studies) identified as heterosexual. Preliminary results from an unpublished conceptual replication of Part III suggested that the curvilinear patterns for perceived societal evaluations may generalize to target persons who are older (65 years) or homosexual. Future studies could investigate how the dissociation between sexual desire and activity observed in older adults (i.e., unchanged desire but reduced activity; Yılkan et al., 2024) might influence the factor structure of the TSMS and thus its comparability across different age groups.

Other Biases

In this dissertation, I focus on what I consider to be some of the major threats to the validity of self-reported sexual motivation in general and of gender differences in sexual motivation in particular. Needless to say, it was not possible to examine all potential biases that might limit the validity and generalizability of sexual self-reports in one project. Two other factors that I believe are worth considering are nonresponse bias and recall bias, both of which have been associated with inaccurate self-reported frequencies of sexual events (e.g., Catania et al., 1993; Graham et al., 2003; McCallum & Peterson, 2012). In what follows, I will briefly discuss their potential implications for gender differences in sexual motivation.

Imagine that Andrea, James, Joyce, and David come across a study advertised as “a survey about sexual experiences and behavior”. Andrea and James find this survey a little dubious. Although the ad looks trustworthy, they are concerned about their privacy and are

reluctant to answer sexual questions. They decide not to participate in the study. If people's decision to (not) participate is correlated with their true level of sexual motivation, this nonresponse bias could threaten the validity and generalizability of the results. For nonresponse bias to explain (part of) the higher values found for men than for women, gender differences in sexual motivation would have to be less pronounced among those not participating, so that their exclusion would result in an exaggeration of gender differences. Thus, on average, Andrea's and other non-participating women's values need to be more above or less below the values of participating women than James's and other non-participating men's values are compared to the average participating man. The devaluation asynchrony observed in Part III (i.e., the tendency toward more pronounced devaluation of men below the ILSA and of more pronounced devaluation of women above the ILSA) supports the idea that low-motivation men and high-motivation women may be most hesitant to respond. However, I find it plausible that norm perception and salience are not randomly distributed across participants but are likely to vary across participants in expected ways. For example, a woman may be more likely to develop or maintain a high (true) level of sexual motivation if she has found a way to protect herself from the effects of society. If this were true, it would be unlikely that these would be the individuals most likely to contribute to biases due to non-response or socially desirable responding. Empirically, we found robust gender differences in studies that were explicitly advertised as sexuality studies and those that were not. For now, I lean toward the cautious tentative conclusion that nonresponse bias is likely to play a rather minor role in explaining measured gender differences, but more dedicated research is needed to address this question convincingly.

Joyce and David *did* participate in the online survey, in which they answered the items of the TSMS. The three parts of this thesis suggest that their reports are likely to (1) measure sexual motivation (2) independently of gender and (3) without a strong impact of self-presentation

motivation. However, to report average event frequencies, they must recall, process, and aggregate information from the past. This poses a threat to the reliability of their self-reports, often referred to as recall bias (Catania et al., 1993; Graham et al., 2003). A problem for the validity of gender differences arises when the result of this complex process is different for men and women. Recalling and processing information from the past creates uncertainty, which opens the door to systematic biases such as gendered social desirability bias (Tversky & Kahneman, 1974). However, we have also found consistent gender differences in experience sampling studies that are characterized by short retrospective intervals of only a few hours (e.g., Part I, Study 4). This provides indirect evidence that recall bias is unlikely to account for much of the gender difference in sexual motivation.

Methodological Rigor, Diversity, and Comprehensibility

An important strength of this research, in my view, is its methodological rigor and diversity. My colleagues and I carefully selected methods that we believed would maximize the fit between what is being measured and what should be measured, thereby setting the stage for valid conclusions. In Part I, we used dynamic cutoffs that were specifically tailored to the characteristics of the model and the sample for an appropriate test of model fit (Study 1). We also considered the sampling characteristics of our dependent variables when running generalized linear mixed models (Study 4). In Part II, we used the innovative Item Sum Technique to establish a strong comparison standard for sexual self-reports in online research. In Part III, we used polynomial linear mixed models to test the predictions of the strong SDS and the S&D model against each other.

We have also tried to ensure that the increased validity promised by sophisticated methods does not come at the expense of the readability and comprehension of our manuscripts. To accomplish this, we implemented two key strategies. First, we introduced complex concepts

such as measurement invariance in plain language (e.g., “comparing apples with oranges”) and linked them directly to their function (e.g., “prerequisite for valid gender differences”). Second, whenever appropriate, we complemented advanced methods with traditional approaches that readers might expect and be more familiar with (e.g., Part I: latent group differences within a structural equation modeling framework and manifest group differences using t-tests; Part III: polynomial linear mixed models and ANOVA).

Openness and Transparency

Another strength of this project is that it has been conducted in the spirit of openness and transparency. All data, scripts, and study materials for the studies presented in this dissertation are openly available on the Open Science Framework, enabling readers to understand the analytical procedures and reproduce our results. For seven of the eight samples, we preregistered the research questions, hypotheses, exclusion criteria, and data analytic strategies, and we transparently reported any deviations from these preregistered plans. To limit researchers’ degrees of freedom, the preregistrations are highly detailed. For instance, in Part I, we specified the fit indices and precise (fixed and dynamic) cutoffs for tests of factorial validity and measurement invariance—a level of detail rarely, if ever, provided in previous scale validation projects.

Concluding Remarks

The present research identified three major threats to validity and examined how they undermine valid measurement and valid conclusions: (1) the *defining threat* of not measuring the construct of interest, (2) the *delicate threat* of comparing groups using an instrument that has different meanings for these groups, and (3) the *social threat* of biased responses in the direction of perceived societal norms. Finding higher sexual motivation in men than in women, with effect sizes similar to previous research, under conditions where biased responses are unlikely, argues

against the assumption that gender differences are largely due to bias. The Trait Sexual Motivation Scale (TSMS), which was designed to address the first two threats, measures sexual motivation reliably and in accordance with theory, and may therefore contribute to valid measurement of sexual motivation in future research. The similarities and differences (S&D) model provides a novel perspective on perceived societal norms that differs significantly from previous research. I discussed how these findings can be reconciled with research showing no gender differences, presented a novel model of flexible self-presentation, and reflected on the practical relevance of the present research. In doing so, I aimed to link basic research with relevant scientific and societal implications in the area of sexual motivation, a construct of paramount importance in life.

References

Abrams, R., Nordmyr, J., & Forsman, A. K. (2023). Promoting sexual health in schools: A systematic review of the European evidence. *Frontiers in Public Health, 11*, 1193422. https://doi.org/10.3389/fpubh.2023.1193422

Ågmo, A., & Laan, E. (2022). Sexual incentive motivation, sexual behavior, and general arousal: Do rats and humans tell the same story? *Neuroscience & Biobehavioral Reviews, 135*, Article 104595. https://doi.org/10.1016/j.neubiorev.2022.104595

Alexander, M. G., & Fisher, T. D. (2003). Truth and consequences: Using the bogus pipeline to examine sex differences in self-reported sexuality. *Journal of Sex Research, 40*(1), 27–35. https://doi.org/10.1080/00224490309552164

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425787

Anderson, J. L. (2024). The trait sexual motivation scale (TSMS): Clinical application of a new instrument. *Journal of Personality Assessment*, 1–2. https://doi.org/10.1080/00223891.2024.2305192

Baumeister, R. F., Brewer, L. E., Tice, D. M., & Twenge, J. M. (2007). Thwarting the need to belong: Understanding the interpersonal and inner effects of social exclusion. *Social and Personality Psychology Compass, 1*(1), 506–520. https://doi.org/10.1111/j.1751-9004.2007.00020.x

Baumeister, R. F., Catanese, K. R., & Vohs, K. D. (2001). Is there a gender difference in strength of sex drive? Theoretical views, conceptual distinctions, and a review of relevant evidence. *Personality and Social Psychology Review, 5*(3), 242–273. https://doi.org/10.1207/S15327957PSPR0503_5

Baumeister, R. F., & Leary, M. R. (2007). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. In In R. Zikauskine (Ed.), *Interpersonal Development* (1st ed., pp. 57–89). Routledge. https://doi.org/10.4324/9781351153683

Borsboom, D., Cramer, A. O. J., Kievit, R. A., Scholten, A. Z., & Franic, S. (2009). The end of construct validity. In R. W. Lissitz (Ed.), *The concept of validity: Revisions, new directions, and applications* (1st ed., pp. 135–170). IAP Information Age Publishing. https://psycnet.apa.org/record/2009-23060-007

Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review*, *111*(4), 1061–1071. https://doi.org/10.1037/0033-295X.111.4.1061

Callea, A., & Rossi, G. (2021). Italian validation of the sexual desire inventory (sdi-2): Psychometric properties and factorial structure. *Clinical Neuropsychiatry*, *18*(4), 223–230. https://doi.org/10.36131/cnfioritieditore20210405

Carver, C. S., & Scheier, M. F. (1981). The self-attention-induced feedback loop and social facilitation. *Journal of Experimental Social Psychology*, *17*(6), 545–568. https://doi.org/10.1016/0022-1031(81)90039-1

Catania, J. A., Turner, H., Pierce, R. C., Golden, E., Stocking, C., Binson, D., & Mast, K. (1993). Response bias in surveys of AIDS-related sexual behavior. In D. G. Ostrow & R. C. Kessler (Eds.), *Methodological issues in AIDS behavioral research. AIDS prevention and mental health* (1st ed., pp. 133–162). Springer. https://doi.org/10.1007/0-306-47137-X_7

Conley, T. D., Moors, A. C., Matsick, J. L., Ziegler, A., & Valentine, B. A. (2011). Women, men, and the bedroom: Methodological and conceptual insights that narrow, reframe, and eliminate gender differences in sexuality. *Current Directions in Psychological Science*, *20*(5), 296–300. https://doi.org/10.1177/0963721411418467

Crawford, M., & Popp, D. (2003). Sexual double standards: A review and methodological critique of two decades of research. *Journal of Sex Research*, *40*(1), 13–26. https://doi.org/10.1080/00224490309552163

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. https://doi.org/10.1037/h0040957

Davies, S., Katz, J., & Jackson, J. L. (1999). Sexual desire discrepancies: Effects on sexual and relationship satisfaction in heterosexual dating couples. *Archives of Sexual Behavior*, *28*(6), 553–567. https://doi.org/10.1023/A:1018721417683

Dawson, S. J., & Chivers, M. L. (2014). Gender differences and similarities in sexual desire. *Current Sexual Health Reports*, *6*(4), 211–219. https://doi.org/10.1007/s11930-014-0027-5

Day, L. C., Muise, A., Joel, S., & Impett, E. A. (2015). To do it or not to do it? How communally motivated people navigate sexual interdependence dilemmas. *Personality and Social Psychology Bulletin*, *41*(6), 791–804. https://doi.org/10.1177/0146167215580129

Deci, E. L., & Ryan, R. M. (2012). Self-determination theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (1st ed., pp. 416–436). SAGE Publications Ltd. https://doi.org/10.4135/9781446249215

Eastwick, P. W., Joel, S., Carswell, K. L., Molden, D. C., Finkel, E. J., & Blozis, S. A. (2023). Predicting romantic interest during early relationship development: A preregistered investigation using machine learning. *European Journal of Personality*, *37*(3), 276–312. https://doi.org/10.1177/08902070221085877

Endendijk, J. J., van Baar, A. L., & Deković, M. (2020). He is a stud, she is a slut! A meta-analysis on the continued existence of sexual double standards. *Personality and Social Psychology Review*, *24*(2), 163–190. https://doi.org/10.1177/1088868319891310

Feather, N. T. (2021). *Expectations and actions: Expectancy-value models in psychology* (1st ed.). Routledge.

Fisher, T. D. (2013). Gender roles and pressure to be truthful: The bogus pipeline modifies gender differences in sexual but not non-sexual behavior. *Sex Roles*, *68*(7–8), 401–414. https://doi.org/10.1007/s11199-013-0266-3

Fisher, T. D., & Brunell, A. B. (2014). A bogus pipeline approach to studying gender differences in cheating behavior. *Personality and Individual Differences*, *61–62*, 91–96. https://doi.org/10.1016/j.paid.2014.01.019

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, *8*(4), 370–378. https://doi.org/10.1177/1948550617693063

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*(6), 1011–1027. https://doi.org/10.1037/0022-3514.80.6.1011

Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science*, *13*(2), 83–87. https://doi.org/10.1111/j.0963-7214.2004.00280.x

Frankenbach, J., Weber, M., Loschelder, D. D., Kilger, H., & Friese, M. (2022). Sex drive: Theoretical conceptualization and meta-analytic review of gender differences. *Psychological Bulletin*, *148*(9–10), 621–661. https://doi.org/10.1037/bul0000366

Gailliot, M. T., & Baumeister, R. F. (2007). Self-regulation and sexual restraint: Dispositionally and temporarily poor self-regulatory abilities contribute to failures at restraining sexual behavior. *Personality and Social Psychology Bulletin*, *33*(2), 173–186. https://doi.org/10.1177/0146167206293472

Gino, F., & Ariely, D. (2012). The dark side of creativity: Original thinkers can be more dishonest. *Journal of Personality and Social Psychology*, *102*(3), 445–459. https://doi.org/10.1037/a0026406

Goldey, K. L., Cital, M. N., Rodriguez, S. C., Espinosa, A., & Barton, E. A. (2024). Desire on lockdown? Sexual desire and COVID-19 stress among LGBTQ+ and cisgender, heterosexual college students. *Psychology of Sexual Orientation and Gender Diversity*, *11*(2), 337–352. https://doi.org/10.1037/sgd0000596

Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk samples. *Journal of Behavioral Decision Making*, *26*(3), 213–224. https://doi.org/10.1002/bdm.1753

Graham, C. A., Catania, J. A., Brand, R., Duong, T., & Canchola, J. A. (2003). Recalling sexual behavior: A methodological analysis of memory recall bias via interview using the diary as the gold standard. *Journal of Sex Research*, *40*(4), 325–332. https://doi.org/10.1080/00224490209552198

Greiff, S., & Scherer, R. (2018). Still comparing apples with oranges? Some thoughts on the principles and practices of measurement invariance testing. *European Journal of Psychological Assessment*, *34*(3), 141–144. https://doi.org/10.1027/1015-5759/a000487

Henry, R. G., & Miller, R. B. (2004). Marital problems occurring in midlife: Implications for couples therapists. *The American Journal of Family Therapy*, *32*(5), 405–417. https://doi.org/10.1080/01926180490455204

Hinzmann, J., Khalaidovski, K., Kullmann, J. S., Brummer, J., Braun, S., Kurz, P., Wagner, K. M., & Schultheiss, O. C. (2023). Developing a causally valid picture-story measure of sexual motivation: I Effects of priming. *Motivation Science*. https://doi.org/10.1037/mot0000300.supp

Jonason, P. K. (2008). A mediation hypothesis to account for the sex difference in reported number of sexual partners: An intrasexual competition approach. *International Journal of Sexual Health*, *19*(4), 41–49. https://doi.org/10.1300/J514v19n04_05

Jonason, P. K., & Marks, M. J. (2009). Common vs. uncommon sexual acts: Evidence for the sexual double standard. *Sex Roles*, *60*(5–6), 357–365. https://doi.org/10.1007/s11199-008-9542-z

Kiefer, A. K., & Sanchez, D. T. (2007). Scripting sexual passivity: A gender role perspective. *Personal Relationships*, *14*(2), 269–290. https://doi.org/10.1111/j.1475-6811.2007.00154.x

Kim, J. J., Muise, A., Barranti, M., Mark, K. P., Rosen, N. O., Harasymchuk, C., & Impett, E. (2021). Are couples more satisfied when they match in sexual desire? New insights from response surface analyses. *Social Psychological and Personality Science*, *12*(4), 487–496. https://doi.org/10.1177/1948550620926770

Kuhn, W., Koenig, J., Donoghue, A., Hillecke, T., & Warth, M. (2014). Psychometrische Eigenschaften einer deutschsprachigen Kurzversion des Sexual Desire Inventory (SDI-2). *Zeitschrift für Sexualforschung*, *27*(02), 138–149. https://doi.org/10.1055/s-0034-1366582

Kutylowski, J. (2024). *DeepL Write*. https://www.deepl.com/en/write

Lefkowitz, E. S., Shearer, C. L., Gillen, M. M., & Espinosa-Hernandez, G. (2014). How gendered attitudes relate to women's and men's sexual behaviors and beliefs. *Sexuality & Culture*, *18*(4), 833–846. https://doi.org/10.1007/s12119-014-9225-6

Levine, S. B. (2003). The nature of sexual desire: A clinician's perspective. *Archives of Sexual Behavior*, *32*(3), 279–285. https://doi.org/10.1023/A:1023421819465

Li, G. (2022). Sexuality development in childhood. In D. P. VanderLaan & W. I. Wong (Eds.), *Gender and Sexuality Development* (pp. 323–356). Springer International Publishing. https://doi.org/10.1007/978-3-030-84273-4_12

Lippa, R. A. (2006). Is high sex drive associated with increased sexual attraction to both sexes? It depends on whether you are male or female. *Psychological Science*, *17*(1), 46–52. https://doi.org/10.1111/j.1467-9280.2005.01663.x

Lippa, R. A. (2009). Sex differences in sex drive, sociosexuality, and height across 53 nations: Testing evolutionary and social structural theories. *Archives of Sexual Behavior*, *38*(5), 631–651. https://doi.org/10.1007/s10508-007-9242-8

Liu, Z., Zhang, H., Wei, L., & Ge, X. (2022). Moral Chameleons: The positive association between materialism and self-interest-triggered moral flexibility. *Journal of Research in Personality*, *100*, 104268. https://doi.org/10.1016/j.jrp.2022.104268

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*(3), 635–694. https://doi.org/10.2466/pr0.1957.3.3.635

Mark, K. P. (2014). The impact of daily sexual desire and daily sexual desire discrepancy on the quality of the sexual experience in couples. *The Canadian Journal of Human Sexuality*, *23*(1), 27–33. https://doi.org/10.3138/cjhs.23.1.A2

Marks, M. J., & Fraley, R. C. (2005). The sexual double standard: Fact or fiction? *Sex Roles*, *52*(3), 175–186. https://doi.org/10.1007/s11199-005-1293-5

Marshall, B. L. (2012). Medicalization and the refashioning of age-related limits on sexuality. *Journal of Sex Research*, *49*(4), 337–343. https://doi.org/10.1080/00224499.2011.644597

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, *45*(6), 633–644. https://doi.org/10.1509/jmkr.45.6.633

McCallum, E. B., & Peterson, Z. D. (2012). Investigating the impact of inquiry mode on self-reported sexual behavior: Theoretical considerations and review of the literature. *The Journal of Sex Research*, *49*(2–3), 212–226. https://doi.org/10.1080/00224499.2012.658923

Meston, C. M., Heiman, J. R., Trapnell, P. D., & Paulhus, D. L. (1998). Socially desirable responding and sexuality self-reports. *The Journal of Sex Research*, *35*(2), 148–157. https://doi.org/10.1080/00224499809551928

Mitchell, K. R., Wellings, K. A., & Graham, C. (2014). How do men and women define sexual desire and sexual arousal? *Journal of Sex & Marital Therapy*, *40*(1), 17–32. https://doi.org/10.1080/0092623X.2012.697536

Muise, A., Impett, E. A., Kogan, A., & Desmarais, S. (2013). Keeping the spark alive: Being motivated to meet a partner's sexual needs sustains sexual desire in long-term romantic relationships. *Social Psychological and Personality Science*, *4*(3), 267–273. https://doi.org/10.1177/1948550612457185

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*(134). https://doi.org/10.1098/rsif.2017.0213

Nilsen, F. A., Bang, H., Boe, O., Martinsen, Ø. L., Lang-Ree, O. C., & Røysamb, E. (2020). The multidimensional self-control scale (MSCS): Development and validation. *Psychological Assessment*, *32*(11), 1057–1074. https://doi.org/10.1037/pas0000950

Nye, C. D., & Drasgow, F. (2011). Effect size indices for analyses of measurement equivalence: Understanding the practical importance of differences between groups. *Journal of Applied Psychology*, *96*(5), 966–980. https://doi.org/10.1037/a0022955

Park, Y., & MacDonald, G. (2022). Single and partnered individuals' sexual satisfaction as a function of sexual desire and activities: Results using a sexual satisfaction scale demonstrating measurement invariance across partnership status. *Archives of Sexual Behavior, 51*(1), 547–564. https://doi.org/10.1007/s10508-021-02153-y

Paulhus, D. L. (1988). Balanced inventory of desirable responding (BIDR). *Acceptance and Commitment Therapy. Measures Package*, *41*, 79586–79587.

Penke, L., & Asendorpf, J. B. (2008). Beyond global sociosexual orientations: A more differentiated look at sociosexuality and its effects on courtship and romantic relationships. *Journal of Personality and Social Psychology*, *95*(5), 1113–1135. https://doi.org/10.1037/0022-3514.95.5.1113

Pospeschill, M. (2022). *Testtheorie, Testkonstruktion, Testevaluation* (2nd ed.). UTB. https://doi.org/10.36198/9783838558332

Powers, W. T. (1973). Behavior: The control of perception. *Science*, *179*(4071), 351–356. https://doi.org/10.1126/science.179.4071.351

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Reddit. (2022). *Meta-analysis shows stronger sex drive in men*. https://www.reddit.com/r/science/comments/zc78qe/metaanalysis_shows_a_stronger_sex_drive_in_men/

Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality*, *43*(2), 137–145. https://doi.org/10.1016/j.jrp.2008.12.015

Sanchez, D. T., Fetterolf, J. C., & Rudman, L. A. (2012). Eroticizing inequality in the United States: The consequences and determinants of traditional gender role adherence in intimate relationships. *The Journal of Sex Research*, *49*(2–3), 168–183. https://doi.org/10.1080/00224499.2011.653699

Schultheiss, O. C., Hinzmann, J., Bergmann, S., Matthes, M., Beyer, B., & Janson, K. T. (2023). Developing a causally valid picture-story measure of sexual motivation: II Effects of film clips. *Motivation Science*, *9*(4), 272–287. https://doi.org/10.1037/mot0000301.supp

Simms, L. J. (2008). Classical and modern methods of psychological scale construction. *Social and Personality Psychology Compass*, *2*(1), 414–433. https://doi.org/10.1111/j.1751-9004.2007.00044.x

Skorska, M. N., Yule, M. A., Bogaert, A. F., & Brotto, L. A. (2023). Patterns of genital and subjective sexual arousal in cisgender asexual men. *Journal of Sex Research*, *60*(2), 253–270. https://doi.org/10.1080/00224499.2022.2071411

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory–2: The BFI-2-S and BFI-2-XS. *Journal of Research in Personality*, *68*, 69–81. https://doi.org/10.1016/j.jrp.2017.02.004

Spector, I. P., Carey, M. P., & Steinberg, L. (1996). The sexual desire inventory: Development, factor structure, and evidence of reliability. *Journal of Sex & Marital Therapy*, *22*(3), 175–190. https://doi.org/10.1080/00926239608414655

Stankov, L. (2018). Psychological processes common to social conservatism and terrorism. *Personality and Individual Differences*, *120*, 75–80. https://doi.org/10.1016/j.paid.2017.08.029

Stark, R., Kagerer, S., Walter, B., Vaitl, D., Klucken, T., & Wehrum-Osinsky, S. (2015). Trait sexual motivation questionnaire: Concept and validation. *The Journal of Sexual Medicine, 12*(4), 1080–1091. https://doi.org/10.1111/jsm.12843

Strang, J. F., Wallace, G. L., Michaelson, J. J., Fischbach, A. L., Thomas, T. R., Jack, A., Shen, J., Chen, D., Freeman, A., Knauss, M., Corbett, B. A., Kenworthy, L., Tishelman, A. C., Willing, L., McQuaid, G. A., Nelson, E. E., Toomey, R. B., McGuire, J. K., Fish, J. N., … Yang, J. S. (2023). The gender self-report: A multidimensional gender characterization tool for gender-diverse and cisgender youth and adults. *American Psychologist, 78*(7), 886–900. https://doi.org/10.1037/amp0001117

Toates, F. (2009). An integrative theoretical framework for understanding sexual motivation, arousal, and behavior. *Journal of Sex Research, 46*(2–3), 168–193. https://doi.org/10.1080/00224490902747768

Tolman, D. L. (2005). *Dilemmas of desire: Teenage girls talk about sexuality* (1st ed.). Harvard University Press. https://doi.org/10.4159/9780674044364

Touraille, P., & Ågmo, A. (2024). Sex differences in sexual motivation in humans and other mammals: The role of conscious and unconscious processes. *Behavioral Sciences*, *14*(4), 277. https://doi.org/10.3390/bs14040277

Trappmann, M., Krumpal, I., Kirchner, A., & Jann, B. (2014). Item sum: A new technique for asking quantitative sensitive questions. *Journal of Survey Statistics and Methodology, 2*(1), 58–77. https://doi.org/10.1093/jssam/smt019

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124

Vallejo-Medina, P., Rojas-Paoli, I., & Álvarez-Muelas, A. (2020). Validation of the sexual desire inventory in Colombia. *Journal of Sex & Marital Therapy*, *46*(4), 385–398. https://doi.org/10.1080/0092623X.2020.1739181

Weber, M., Reis, D., & Friese, M. (2024). Development and validation of the trait sexual motivation scale (TSMS). *Journal of Personality Assessment*, *106*(2), 267–282. https://doi.org/10.1080/00223891.2023.2206896

World Health Organization. (2019). *International statistical classification of diseases and related health problems* (11th ed.). https://icd.who.int/

Yılkan, I., Apak, E., & Artan, T. (2024). Opinions of the elderly on sexuality in old age: "I want but I can't." *Archives of Gerontology and Geriatrics Plus*, *1*(1), 100004. https://doi.org/10.1016/j.aggp.2024.100004

Zuk, M., & Simmons, L. W. (2018). *Sexual selection: A very short introduction* (1st ed.). Oxford University Press. https://doi.org/10.1093/actrade/9780198778752.001.0001