Current Psychology
https://doi.org/10.1007/s12144-022-04227-z

Are women truly "more emotional" than men? Sex differences in an indirect model-based measure of emotional feelings

Ella Givon[1] · Rotem Berkovich[1] · Elad Oz-Cohen[1] · Kim Rubinstein[1] · Ella Singer-Landau[1] · Gal Udelsman-Danieli[1] · Nachshon Meiran[1]

Accepted: 28 December 2022

Abstract

Common beliefs regard women as being more emotional than men. However, assessing differences in emotional feelings holds methodological challenges because of being based on explicit reports. Such research often lacks an explicit measurement model, and reports are potentially biased by stereotypical knowledge and because of existing sex differences in the ease of emotion-label retrieval. This pre-registered analysis employed an evidence accumulation model that has previously been validated for describing binary (un)pleasantness reports made in response to normed emotion-eliciting pictures. This measurement model links overt binary (un)pleasantness reports with the latent variables processing efficiency and a bias to report a certain emotional feeling. Employing online rather than retrospective reports that do not involve intensity rating, together with an explicit measurement model overcome the aforementioned methodological challenges. Across nine different experiments (N = 355) women generated negative emotions more efficiently than men. There was no sex difference in the bias to report negative emotions and in positive emotions. Post hoc account of the results emphasizes the greater relevance of negative emotions for women, given their evolutionary role as primary caregivers who should show enhanced sensitivity for dangers to their offspring ("fitness threat"), given their heightened likelihood of being themselves exposed to physical violence and given their traditional social roles that still remain relevant in many societies.

Keywords Emotional experience · Evidence-accumulation modeling · Sex-differences · Reaction-time

Conventional wisdom holds that men and women differ from one another, emotionally (e.g., Durik et al., 2006). Particularly, women are perceived as being more emotional as compared to men. Such beliefs are deeply rooted in present day culture, media, and society, with origins dating back to much earlier periods. Despite this stereotype being so widespread, whether it is justified remains hotly debated. To test the validity of this conventional wisdom, an extensive amount of empirical work has been conducted. Sex differences have been assessed in various domains of emotion, covering expressions, autonomic activity, brain imaging, action tendencies and subjective experiences (see more below).

A recent analysis of facial expressions employing various subtle measures revealed that women are not universally more emotionally expressive then men, and that sex differences in expressiveness of negative valence depend on emotion type (McDuff et al., 2017). In a different review (Fischer & LaFrance, 2015), the authors argued that women are generally more expressive than men (particularly in crying and smiling), but that the magnitude of these sex differences is determined by three factors: gender-specific norms, social role and situational constraints, and emotional intensity. Research further found that the sexes differ in the pattern of autonomic arousal (e.g., Kring & Gordon, 1998) and other studies suggested that such sex differences are restricted to specific types of emotion (e.g., Deng et al., 2016). Yet, other studies failed to discriminate between sexes based on physiological measures (Kelly et al., 2006). From a neurological perspective, a relatively recent meta-analysis (Filkowski et al., 2017) revealed sex differences in the activation of emotion-related brain areas. In another meta-analysis of neuroimaging studies, Stevens and Hamann (2012) focused on valence and found that the pattern of sex differences was different for positive and negative emotions. Research on

✉ Ella Givon
ellasharon123@gmail.com

[1] Department of Psychology and Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, 84105 Beer-Sheva, Israel

Published online: 16 January 2023

hormones and genes further pointed to interactions between hormones levels and gene expression as an underlying mechanism for sex differences in emotional behavior (Kret & De Gelder, 2012). In the behavioral domain, women show disposition to respond defensively to all kinds of aversive stimuli, while men reacted appetitively when viewing erotica (Bradley et al., 2001).

Nonetheless, all these findings do not directly answer the question regarding differential emotional *subjective experience* between the sexes. Specifically, LeDoux and Hofmann (2018) argued that subjective emotional experience (feeling), is the core of emotion. They further argued that biological and behavioral findings (such as those reviewed above) are indirect and imperfect indicators of these inner experiences. Additional evidence for why studying objective manifestations of emotion is limited comes from a recent study on sex differences in emotional concordance (Rattel et al., 2020). Specifically, this research showed that women display higher coherence between diverse components of emotion such as physiology and explicit report of feeling. This implies that indirect physiological and behavioral measures may be reasonable indicators of feelings for women. However, given the lower coherence, physiology and behavioral manifestations are poor indicators of feeling for men. Following LeDoux and Hofmann (2018), we believe that sex differences in the biological and behavioral correlates of emotion fall short of addressing the essence of this topic, which is whether sex-differences exist in the psychological emotional experience.

Regarding feelings, studies found that women differ from men mostly in negative emotions (Gard & Kring, 2007; Simon & Nath, 2004), report experiencing more powerless emotions (Fischer et al., 2004), rate lower (more negative) valance across various content categories (Hillman et al., 2004; Maffei & Angrilli, 2019), and differ from men in how they experience their emotions in the body (Volynets et al., 2020). Furthermore, there are sex differences in depression, and these differences in depression have been documented across different countries and cultures (Hopcroft & Bradley, 2007). Relevant suggestions for an underlying reason include sex differences in emotional reactivity and in prevalence of negative life events (Hyde et al., 2008).

Although direct reports are arguably the method of choice for assessing feelings, simply asking people to report how they feel has its challenges (Schwarz, 2012). For example, Fugate et al. (2009) suggested that self-reports of emotions are prone to retrospective biases and gender stereotypes, thus influencing findings regarding sex differences in emotions. This can be exemplified when participants are asked to report the intensity of a felt emotion over the last week or even today (PANAS; Watson et al., 1988). Given the imprecise memory of feelings, one must employ some fill-in, which involves semantic memory, including stereotypes. Accordingly, Barrett et al. (1998) showed that sex differences in emotional experience depend on social context, and particularly appear strongly when measurement involves global retrospective self-reports. Furthermore, Fugate et al. (2009) raise the possibility that women may report more about their feelings simply because of being relatively more fluent with emotion language. An additional complication is that reports may reflect "semantic valence" i.e., whether a stimulus would be commonly regarded as pleasant or unpleasant (Itkes et al., 2017) rather than reflecting the current subjective experience.

There are several different self-report methodologies that may help overcome some of these challenges. One relatively new and popular approach is to count emotion words in social media. The obvious advantage is that these are spontaneous, online expressions that are available for huge samples. However, a recent study casts doubt on the validity of this method (Kross et al., 2019). Another approach is to report about how one feels the emotion in the body (Volynets et al., 2020), but this highly intriguing approach had so far been used to assess typical emotion (e.g., what you feel when you feel love) rather than online feeling (what you feel *now*), thus potentially involving the influence of stereotypes.

A broader concern relates to the general issue of psychological measurement, which is the potential confusion between observed measures and theoretical attributes, or "latent variables" (Borsboom, 2006). Accordingly, observed scores such as emotion reports are often mistakenly used as substitutes for theoretical constructs (e.g., feelings). The point is that while ratings of emotional feelings are clearly related to actual feelings, they are not identical with them. This is an especially challenging problem in the case of sex differences, since reports of emotional feelings (rather than the emotional feelings proper) are arguably biased by stereotypes as explained beforehand. In order to make inferences regarding feelings (rather than regarding feeling *ratings*), one needs an explicit and validated model which describes the relationship between the observed variable (reports) and the latent variable (feelings).

Our review above indicates that reports of emotional feelings may (1) be biased by stereotypes especially when they are retrospective and global, (2) be influenced by differential fluency in emotional language, and (3) reflect semantic valence rather than truly felt experience. Furthermore, to properly measure feelings rather than rely on raw self-reports, (4) one needs an explicit, validated measurement model.

In the current study, we tried to overcome these measurement challenges. We asked participants to report their (locally, in the moment, felt) emotion while being exposed to emotion-eliciting pictures, rather than have asked them to provide retrospective / global reports. By dropping the requirement to recall past emotion experiences, we have possibly minimized the influence of stereotypes that arguably help filling-in blanks in episodic recall. Additionally, we obtained a graded metric without explicitly asking for a graded response. Specifically, *reports involved a binary choice between the same response options (pleasant vs. unpleasant)*. We then used an explicit processing model, that describes how feeling reports are being generated, as a measurement model. This model thus enabled us to deduce the intensity of the felt emotion indirectly from binary pleasant vs. unpleasant reports without explicitly asking participants to report feeling intensity. Thus, our measure of intensity minimally involved (if at all) stereotypes regarding sex-differences in the intensity of felt emotion which may exist when directly grading the intensity.

Our method has several additional advantages. One such advantage is that the response labels were constant, and were provided in advance. This method does not challenge the ability to come up with verbal labels for emotions, an ability which presumably differentiates between the sexes (Fugate et al., 2009). Moreover, we have also very explicitly instructed participants regarding the difference between semantic valence and actual experience, and have asked them to report only about their authentic and momentary felt experience, even if they believe that their experience deviates from what most people would feel. Finally, given that some emotions may be embarrassing to report, and given that such embarrassment may also differ across sexes, we limited the stimuli to those which are unlikely to invoke such feelings (i.e., we excluded stimuli associated with morbid curiosity and porn). Perhaps the most substantial advantage of the current work is using an explicit measurement model. We describe the model and its implications in the following sections.

Feeling generation as evidence accumulation

Overview of the model

Evidence accumulation models (EAMs; see e.g., Donkin & Brown, 2018) consist of different models that describe speeded decision making (in this case, the decision to report "pleasant" vs. "unpleasant"). The models describe the underlying processes mathematically, and thus make it possible to "back engineer" and use RTs and response identity information (pleasant, unpleasant) to uncover psychological processes that underlie decision making (see more below), and most importantly – quantify emotion intensity.

EA models are typically used to study perceptual decisions and the different EA models all share a fundamental principle of sequential sampling of evidence over time. The evidence that is being sampled in the sampling process is noisy and evidence favoring each response alternative is accumulated until it reaches threshold. In verbal terms, the proposed process goes as follows: When a person is exposed to an emotion-eliciting stimulus, s/he starts gathering emotional evidence from own cognitive appraisals, bodily sensations, action drives, etc. (all emotion components, see Scherer, 1987), and when enough emotional evidence favoring a given response alternative has been accumulated and has reached threshold – that person can report his/her feeling. Two of the most commonly used EA models are the Drift Diffusion Model (DDM; Ratcliff & McKoon, 2008) and the Linear Ballistic Accumulator model (LBA; Brown & Heathcote, 2008). It has been previously shown that while these models rely on different theoretical frameworks, they ultimately lead to quite similar conclusions (Donkin et al., 2011; Rae et al., 2014). Given their similarity, we (Givon et al., 2020) chose the LBA as a model of feeling reports, mainly because of its theoretical and practical simplicity.

LBA parameters of feeling reports

The generic LBA has 5 parameters, two of which represent the core decision making mechanism: Drift-Rate and Threshold (see Fig. 1 for a graphic illustration of these parameters). Drift-Rate is the most important parameter in the present context, because it consists our measure of emotion intensity. Technically, it represents the mean (across trials) rate of evidence accumulation, and it accounts for the quality of information derived from the stimulus. In the context of feeling-reports, drift-rate represents the joint influence of the emotional intensity of the stimulus, the changes in the emotion system when exposed to the stimulus, and participant's ability to detect these changes. Ultimately, a higher Drift-Rate (given equally potent stimuli) corresponds to a more efficient accumulation of emotional evidence (and to a more strongly felt emotion). The Drift-Rate is graphically presented as the slope of the function relating evidence to time in the accumulator in the graph describing evidence accumulation, such that a high Drift-Rate corresponds to a steep slope (see Fig. 1b). The Threshold is the total amount of emotional evidence that is required to reach a decision, or when applied to feeling reports, how much evidence is needed in order to feel. This parameter mainly reflects speed-accuracy tradeoff. In other words, a low threshold enables making quick decisions that are based on little evidence and as such are error-prone, whereas a high threshold enables making high quality decisions that are based on more evidence, at the expense of taking more time. Accordingly, cautious people (in terms of readiness to consciously experience emotion) require a larger amount of emotional evidence and would take longer to decide how they feel, while people with lower threshold require less emotional evidence and would decide sooner how they feel. Graphically, the

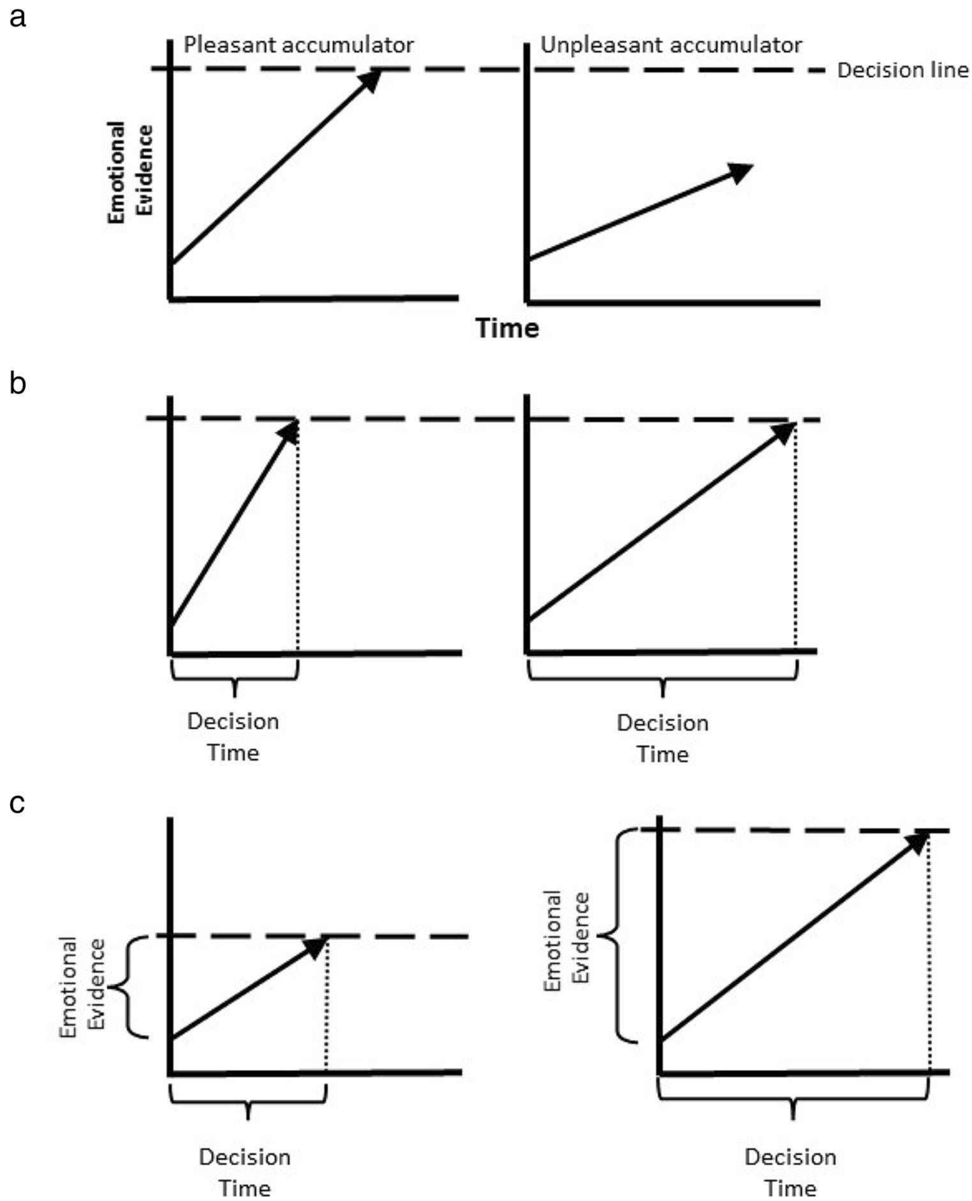

Fig. 1 (**a**) A schematic illustration of competition between accumulators according to the LBA. X-axis represents time and Y-axis represents emotional evidence. The first accumulator to cross the decision line is the one to determine the decision what one feels: pleasant or unpleasant. (**b**) Influence of Drift-Rate. The left panel presents a higher Drift-Rate, graphically displayed as steeper slope of the left accumulator. For this accumulator, evidence is accumulated more efficiently, and decision time is faster with no decrease in the amount of evidence on which the decision is based. (**c**) Influence of Threshold. The left panel presents a low Threshold (the height of the decision line), meaning that less evidence is required to decide. Decision time is faster, but the decision relies on less evidence thus resulting in a higher error-rate

threshold is presented as the height of the decision line, and a higher threshold corresponds to a higher line (see Fig. 1c). Importantly, the Drift-Rate and the Threshold parameters account for two distinct mechanisms. When applied to sex-differences, a higher Drift-Rate in one of the sexes would indicate greater sensitivity to relevant emotional events/stimuli. This greater sensitivity means greater changes in inner emotion-related happenings (e.g., changes in emotion-related action tendency, autonomic arousal, cognition) that support reportable/conscious emotional experiences. Alternatively, sex-differences in the Threshold imply that one of the sexes needs less evidence concerning these inner happenings in order to consciously feel and report.

The LBA, as other EA models, uses accuracy and RTs to yield its parameters, which are latent variables (i.e., resembling factors in Structural Equations Modeling). While in perceptual tasks accuracy is obtained through comparing participant's response to an objective value (the correct answer), here we use social norms of emotion from the NAPS database (Marchewka et al., 2014) as a substitute for an objective value (see more on this issue below). The LBA parameters are estimated by fitting the LBA model to observed results (measures of accuracy and RTs). This fitting is possible, since the LBA makes precise predictions regarding the shape of the RT-distributions of correct (normative) and incorrect (aberrant) responses as well regarding the relative rate of normative/aberrant responses. Loosely speaking, model fitting is accomplished by gradually changing the LBA parameter-values until the LBA predictions (which change when the parameter-values change) closely resemble the observed results, including the relative rate of aberrant/normative responses as well as the shape of their RT distributions. The LBA parameter-values that yield the closest match are regarded as the final estimates of the latent variables. This process conceptually resembles how regression weights are being computed by changing the weights so that the prediction most strongly correlates with the actual outcome.

Previous validations of the LBA in feeling reports

In perceptual tasks, the LBA has been extensively validated (e.g., Donkin et al., 2011; Rae et al., 2014). We (Givon et al., 2020) proposed to use the LBA as a process model for describing the emergence of reportable feeling and we have established it in various studies (Givon et al., 2020; see also Singer-Landau & Meiran, 2021; Berkovich & Meiran, 2022 for additional support). Model validation, reported in the aforementioned papers, involved all the necessary steps. To summarize, we (1) showed that the model fits the data excellently (e.g., Berkovich & Meiran showed RMSEA (Steiger & Lind, 1980) values of 0.038 and 0.013, which is about 1 order of magnitude less (= better) than the recommended (0.08, Schubert et al., 2017), and similar to that found in perceptual decisions (RMSEA = 0.041 and 0.043, respectively), which is the domain in which the model was originally developed. We additionally (2) compared between different variants of the model, to show "selective influence"—i.e., that a manipulation influences only the parameters which are hypothesized to be affected by it. Specifically, we showed that the normative rated intensity of the stimuli had an impact on the Drift-Rate but did not change the Threshold (Berkovich & Meiran, 2022; Givon et al., 2020), and that cognitive reappraisal (an emotion down-regulation strategy) has an analogous effect (Singer-Landau & Meiran, 2021). Berkovich and Meiran (2022) have (3) additionally supported the ratio-scale properties of the Drift-Rate by showing that drift-rate follows the classic Weber's Law which describes how encoding uncertainty is affected by intensity. We (4) validated a critical auxiliary assumption that we employed in our modeling, in which we treat counter-normative emotional reports as errors. We specifically showed that the counter-normative emotional reports closely resemble perceptual decision errors, in terms of the reaction-time effects and distributions (Givon et al., 2022). Some of these steps are replicated in the present work (see Results). Last, to ensure that participants report genuine versus expected feelings, we (Givon et al., 2020) ran experiments that compared between two groups that differed only on the instructions regarding how to respond to the emotion stimuli (different instructions were used to dissociate affective and semantic valence). To this end, in the self-focused group, participants were requested to respond only according to their own feelings and answer the question: "Does this photo make you feel pleasant?". In the stimulus-focused group, participants were requested to ignore their own feelings, focus on the stimulus, and answer the question: "Is this photo supposed to create a pleasant feeling?". We found a higher rate of normative responses in stimulus-focused, and meaningful group-differences in the LBA processing parameters. These findings validate the task (participants report experienced as opposed to expected feeling) and our model (since it captures these differences).

An important aspect in our modeling is that we classified responses as "normative" and "aberrant" and treated them in the model as if they represented "correct" and "incorrect" decisions. This treatment is justified by a recent work from our lab (Givon et al., 2022), showing that "normative" and "aberrant" emotion reports show very similar processing characteristics to "correct" and "incorrect" responses made in a perceptual decision task in which an objective answer exists. Specifically, the resemblance has been shown in several landmarks of errors including the shape of the RT-distribution, the responsiveness to speed vs. accuracy emphasis instructions, the slowing that takes place after aberrant responses (A.K.A, "post-error-slowing") and the (lack of) error-related EEG marker. Importantly, we do not claim that emotions are not subjective, nor do we claim that emotions have a truth value. Instead, we tackled this complicated issue empirically. Based on our results (Givon et al., 2022), we suggest that at the cognitive-performance / brain level, counter-normative emotion reports closely resemble errors in perceptual decision tasks and can thus be modelled using the same processing model (LBA).

The current study

So far, we have used the LBA model solely in order to make inferences at the group level (Givon et al., 2020; Singer-Landau & Meiran, 2021), in which the LBA parameters accounted for differences between experimental conditions. Here, we wish to examine whether the LBA parameters (specifically the Drift-Rate and the Threshold) can stand as reliable measures of individual differences and can eventually be used to reflect sex differences. For that purpose, data were aggregated across 9 different experiments, in each of them participants were required to respond whether an emotional picture made them feel pleasant or unpleasant. The experiments were each conducted to address very different research questions and thus differed from one another in the type of stimuli, instructions, and in other aspects. Although this is an accidental aspect of our study, it strengthens its external validity because it shows that the conclusion holds across different samples, stimuli and instructions. We used the 'ggdmc' R package for Hierarchical Bayesian LBA modeling (Lin & Strickland, 2020) to generate a unique set of LBA parameters for each participant in 2 types of emotion-valence: negative and positive. We hypothesized that sex differences in feeling-reports, if exist, would appear in negative emotions, either in the Drift-Rate, the Threshold, or both. Specifically, our hypotheses concerning sex differences (H1) were as follows: (H1.1) Women will present a higher Drift-Rate in negative emotions, as compared to men; (H1.2) Women will present a lower Threshold in negative emotions, as compared to men; (H1.3) Women will present a higher Drift-Rate and lower Threshold, as compared to men. These hypotheses were based on past literature that showed sex differences mostly in the experience of negative but not positive emotion (e.g., Bradley et al., 2001; Fischer et al., 2004).

Method

Pre-registration

Core analyses as well as main hypotheses were pre-registered, see: https://osf.io/j9aun/. Eight out of the nine experiments were pre-registered, see the following list:

Experiment 2: https://osf.io/apfts/
Experiment 3: https://osf.io/gvy3t/
Experiment 4: https://osf.io/d9a2s/
Experiment 5: https://osf.io/8g932/
Experiment 6: https://osf.io/guhtb
Experiment 7: https://osf.io/3q68t
Experiment 8: https://osf.io/h7kcx
Experiment 9: https://osf.io/wbeqg

Participants

All participants were recruited through the Ben-Gurion university psychology participant pool. Participants received course credit or payment for participating. All participants were at least 18 years old and not older than 40 years old. They were native speakers of Hebrew (to ensure that the meaning of "pleasant" and "unpleasant" in the instructions remains uniform across participants) and all reported having normal or corrected-to-normal vision. In the current study, we included only participants who were instructed to respond according to their own feelings.

Stimuli and procedure

In nine different experiments, participants (total N = 357) were required to respond whether an emotion-evoking picture made them feel pleasant or unpleasant (the emotion task). Given that each experiment was designed to answer a different question, the experiments differed from one another (for a summary of all differences between experiments, see Table 1). The common denominator of all the experiments was that they comprised trials each made of two tasks: beginning with an emotion report task and followed by a generally emotional-neutral perceptual-decision task. This perceptual decision task mainly served to disguise the real focus of the study (in which emotion was the dependent variable) by depicting the study as focusing on how emotions (depicted as an independent variable) influence perceptual decision. The perceptual decision task also served to minimize emotion carryover effects from one trial to the next and (in some experiments) the task served as a benchmark in which emotion reports were compared to.

The emotion task is the focus of the current study. A version of it was used in each of the 9 experiments. Participants viewed emotion-eliciting photos with established norms (NAPS; Marchewka et al., 2014). Stimuli were presented on computer screens and each emotion-photo was proceeded with a fixation frame matched to the stimulus (vertically or horizontally). Unbeknown to the participants, the photos were taken from two different categories according to their valence, negative (below 5) and positive (above 5). Valence values were obtained from the NAPS established norms (Marchewka et al., 2014). Valence scale ranged from 1 (mostly negative) through 5 (neutral) to 9 (mostly positive). Under each photo appeared a question regarding the feeling (i.e., whether the stimulus elicits a pleasant or an unpleasant feeling?). Accuracy (or adherence to norm) and RTs were collected, and errors were defined as responding "yes" to a normatively (according to the NAPS) negative photo or responding "no" to a normatively (according to the NAPS) positive photo. RT was defined as the time from picture presentation until the key press.

Importantly, we ensured that in all experiments, instructions for the emotion task were highly detailed regarding how to respond, with an emphasis to respond only according to one's own feeling and not what participants believe to be the expected emotion. The experimenter provided examples for experiencing an emotion that deviates from the expected emotion of the stimulus (i.e., liking venomous snakes), and emphasized that there is no "right" answer, and requested participants to respond only according to what they truly feel.

Results

Analysis

Preprocessing and Bayesian Hierarchical modeling were conducted using the R software (R Core Team, 2021). Bayesian Analysis of Variance (BANOVA) and Person correlations were conducted using the JASP software (JASP Team, 2022). Based on Givon et al. (2022), who showed processing similarity between perceptual errors and aberrant emotion reports, we defined "accuracy" in the emotion task as adherence to social norms. Simply put, an "error" in the

Table 1 Differences between experiments

Exp	N (F/M)	Blocks (#trials per block)	Valence range, pleasant stimuli	Valence range, unpleasant stimuli	Phrasing the emotion question	Lab/ online	Perceptual decision task	Presentation time of the emotion stimuli	Manipulation	NAPS content categories	Comments
1	29 (20/9)	5 (40)	6.5–7.5	3.5–4.5	Does this photo make you feel pleasant?	Lab	Faces task (deciding whether a face presented is of a male or female)	Until participant's response		An equal representation of Animals, Objects, Faces, People, and Landscapes	Experiment 1 from Givon et al. (2020)
2	40 (22/18)	4 (50) only the 2 baseline blocks were used here	6.5–7.5	3.5–4.5	Does this photo make you feel pleasant?	Lab	Faces task	Until participant's response	Speed-accuracy manipulation that appeared after the baseline phase	An equal representation Animals, Objects, Faces, People, and Landscapes	Experiment 3 from Givon et al. (2022)
3	60 (34/26)	10 (20)	6.5–7.5	3.5–4.5	Choosing between pleasant and unpleasant	Lab	Faces task	5000 ms		An equal representation of Animals, Objects, Faces, People, and Landscapes	EEG was recorded during the experiment. Experiment 4 from Givon et al. (2022)
4	34 (28/12)	4 (60) only the 2 blocks of the "real" condition were used here	5.5–6.5	2.0–4.0	Does this photo make you feel unpleasant?	Lab	Faces task	Until participant's response	Reality challenge (fictitious vs. real)	Representation of: Faces (33.33%), Objects (33.33%), People (20%), Landscapes (13.33%)	Experiment 1 from Singer-Landau and Meiran (2021)
5	34 (21/13)	4 (60) only the 2 blocks of the "real" condition were used here	5.5–6.5	2.0–4.0	Does this photo make you feel unpleasant?	Online	Faces task	Until participant's response	Mixed reality challenge (real + fictitious vs. real)	Representation of: Faces (33.33%), Objects (33.33%), People (20%), Landscapes (13.33%)	Experiment 2 from Singer-Landau and Meiran (2021)

Table 1 (continued)

Exp	N (F/M)	Blocks (#trials per block)	Valence range, pleasant stimuli	Valence range, unpleasant stimuli	Phrasing the emotion question	Lab/ online	Perceptual decision task	Presentation time of the emotion stimuli	Manipulation	NAPS content categories	Comments
6	34 (20/14)	4 (60)	5.5–6.5	2.0–4.0	What emotion does this photo evoke in you?	Online	Waldo searching task	Until participant's response	Approach tendency (close vs. far from screen keys)	Representation of: Faces (33.33%), Objects (33.33%), People (20%), Landscapes (13.33%)	
7	34 (29/5)	6 (23–26)	5.85–7.92	3.0–4.36	Does this photo make you feel pleasant?	Online	Circle task (Deciding whether a circle presented is larger than a size of a permanent circle)	4500 ms		An equal representation of Animals, Objects, Faces, People, and Landscapes	Main experiment from Berkovich and Meiran (2022)
8	34 (24/10)	6 (23–26)	5.85–7.92	3.0–4.36	Does this photo make you feel pleasant?	Online	Circle task	4500 ms		An equal representation of Animals, Objects, Faces, People, and Landscapes	Replication study from Berkovich and Meiran (2022)
9	58 (40/18)	4 (2 contained 35 trials and 2 contained 45 trials), trials of "intense negative" (20) were excluded	6.5–7.6	2.5–3.5	Does this photo make you feel pleasant?	Online	Faces task	4500 ms	Emotion regulation manipulation for blocks involve intense negative trials	An equal representation of Animals, Objects, Faces, People, and Landscapes	

emotion task would be responding 'pleasant' to a picture with an unpleasant valence and vice versa. Note that we do not make the claim that aberrant emotion reports are errors, only that they are processed like errors, and thus merit being regarded as errors when applying a processing model (LBA).

Pre-processing pipeline

We used the same preprocessing pipeline (4 steps) as that used in previous published analyses from our lab (Singer-Landau & Meiran, 2021) and we performed it systematically on the raw data from all 9 experiments. Each step received as input the output of the preceding step:

Step 1: Excluding deviant participants—those whose responses indicated less than 50% accuracy. (2 participants whose accuracy rates did not exceed 50% were excludes from the analyses, one male and one female).
Step 2: Excluding manipulated conditions and extreme valence stimuli (above 8 and under 2) for the purpose of standardization across experiments.
Step 3: Excluding trials with RTs shorter than 150 ms.
Step 4: Excluding trials with RT > 2.5 SDs above the mean. For each combination of emotion and participant, we computed the mean and SD, and excluded RTs > Z = 2.5.

Bayesian Hierarchical LBA modeling

Using the 'ggdmc' R package (Lin & Strickland, 2020), we generated a unique set of LBA parameters per emotion for each participant. Hierarchical Bayesian modeling produces posterior probability distributions for each parameter based on priors and the data. We used priors that are informed by previous modeling of similar data (Singer-Landau & Meiran, 2021), and ensured that the priors for types of emotions (positive and negative) were equal in all models. Thus, any difference found between posterior parameters could not have been related to differences between priors. Fitting was done using 1000 samples per chain in an initial "burn-in" phase, followed by an actual sampling phase in which we used 5000 samples per chain (while "thinning" i.e., keeping each 12th sample). Finally, we computed Potential Scale Reduction Factor (PSRF) for each participant to assess convergence between the chains. An acceptable level of convergence is when all the individual PSRFs fall below 1.1 (Brooks & Gelman, 1998). For all participants in all analyses, PSRF values were all lower than 1.003.

Model fit

This analysis was not pre-registered. An important issue concerns whether the model fits the data sufficiently well, to legitimize using its estimates. We used a recent extension of the RMSEA index (Steiger & Lind, 1980) to evidence accumulation models (Schubert et al., 2017). According to Schubert et al. (2017), RMSEA should be below 0.08 to indicate good fit. Our results show a RMSEA = 0.00529, (CI = 0.00507,0.00553). Our 95% confidence interval was calculated using Bootstrap (#samples = 500), with resampling conducted over participants. These results indicate the excellent fit of the model because RMSEA values were considerably smaller (= better) than the recommended upper threshold-value, a fact that further supports model validity.

Reliability

All reliability analyses were performed on data from Experiments 1–3 (total N = 129). Using Bayesian Hierarchical LBA modeling, we extracted for each participant 4 values of the Threshold parameter. The four values are the combinations of valence (2, positive or negative) and odd–even status of the trial (2, odd or even). We additionally extracted 8 values of the Drift-Rate parameter (combinations of odd–even-numbered trials (2), valence (2), and correct vs incorrect (2)). All these LBA parameters showed good internal reliability (r > 0.85 between odd-and even, and Spearman-Brown boosted reliabilities of at least 0.92; See Table 2 for a summary of reliability analyses). For a graphic illustration of the core parameters' reliabilities (Threshold and Drift-Rate correct in negative emotions), see Fig. 2.

Sex differences

Following the establishment of the Threshold and the Drift-Rate as reliable measures of individual differences, we examined whether they account for sex differences in emotion. We performed Bayesian hierarchical LBA modeling on the pre-processed data from all 9 experiments. To simplify matters, this model receives as an input data with accuracy rates and RTs for each condition (negative and positive) and generates estimated LBA-parameters for each participant. The more interesting parameters, including the Drift-Rate parameters and Threshold are estimated separately for "pleasant" and for "unpleasant". Eventually, there are six values for each participant: Threshold in negative emotions, Threshold in positive emotions, Drift-Rate correct (normative) in negative emotions, Drift-Rate correct in positive emotions, Drift-Rate error (aberrant) in negative emotions and Drift-Rate error (aberrant) in positive emotions. For each of these values, we performed B/ANOVA in which the LBA parameter was the dependent variable, while the independent variables were sex (male vs. female) and experiment (8 levels, each experiment stand as a group except for Experiments 1 and 2 that were united into one group because they were identical; See Table 3 for a

Table 2 Summary of parameters' reliability

Parameter	Pearson correlation between odd and even trials	95% Credible Interval	Spearman-Brown boosted reliability
Threshold negative	0.871	[0.817, 0.906]	0.931
Threshold positive	0.871	[0.817, 0.906]	0.931
Drift-Rate correct negative	0.874	[0.822, 0.908]	0.932
Drift-Rate correct positive	0.905	[0.865, 0.931]	0.950
Drift-Rate error negative	0.860	[0.802, 0.897]	0.924
Drift-Rate error positive	0.866	[0.811, 0.902]	0.928

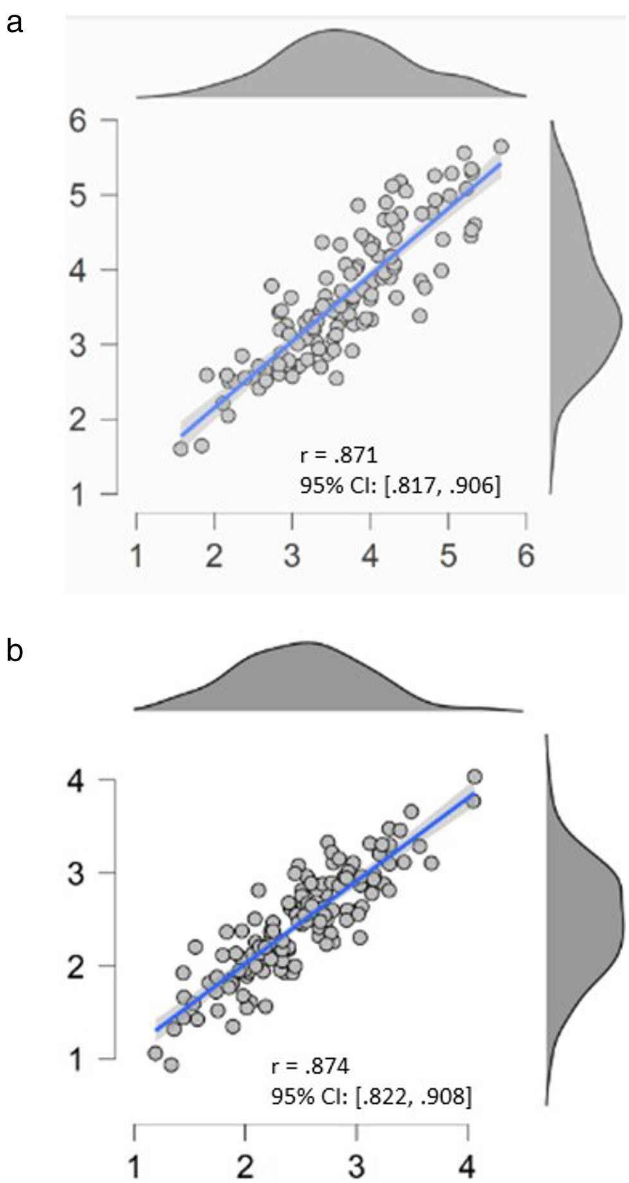

Fig. 2 (**a**) Pearson correlation between odd and even numbers for Threshold in negative emotions. (**b**) Pearson correlation between odd and even numbers for Drift-Rate correct in negative emotions. CI represent credible intervals

summary of B/ANOVA). Creating a set of analyses with these two independent variables (sex and experiment), enable us to look whether potential differences between sexes are replicated across different experiments. We were thus interested in both the main effects of sex (telling whether the sexes differ in each parameter) as well as in the interaction between sex and experiment (telling whether the sex difference is consistent across experiments). With reference to the hypotheses, our focus of interest was Threshold in negative emotions and Drift-Rate correct in negative emotions. Regarding H1.2, the analysis indicated lack of main effect for sex in Threshold-Negative (decisive support for lack of effect, i.e., the null hypothesis, H0), $BF_{01} > 100$. Regarding H1.1, we found decisive evidence for a main effect of sex in Drift-Rate-Correct-Negative, ($F_{(1,339)} = 16.8$, $p < 0.001$, $BF_{10} = 73.26$). Specifically, across all experiments, women showed higher Drift-Rate-Correct-Negative as compared to men. The analysis indicated lack of interaction between sex and experiment ($BF_{01} = 4.54$, supporting the null hypothesis assuming zero interaction, H0), meaning that the sex difference was consistent across experiments. For a graphic illustration of sex differences in LBA core parameters, see Fig. 3. All other parameters that were examined failed to show sex differences (see Table 3), except for some evidence for sex differences in Drift-Rate-Correct in positive emotions ($F_{(1,339)} = 6.258$, $p = 0.013$, $BF_{10} = 3.392$). We examined the interaction between sex and experiment in this analysis to check if the sex differences in Drift-Rate-Correct in positive emotions was consistent across experiments. This analysis revealed decisive support for an interaction effect between sex and experiment ($F_{(7,339)} = 5.354$, $p < 0.001$, $BF_{10} = 562.6$). It seems that there was a robust effect for sex in Experiment 6, but this effect did not replicate across the experiments. In fact, when data from Experiment 6 were excluded from the analysis, there was no effect for sex in Drift-Rate-Correct in positive emotions ($F_{(1,307)} = 0.089$, $p = 0.766$, $BF_{01} > 10{,}000$, a result providing decisive support for H0).

Table 3 Summary of B/ANOVA of sex differences in LBA parameters

LBA Parameter Dependent Variable	Fsex	*P*-value	BF10	Comments
Threshold negative	3.497	0.062	0.15	Acceptance of H0
Threshold positive	4.629	0.032	0.173	Acceptance of H0
Drift-Rate correct negative	**16.802**	**< 0.001**	**73.262**	Acceptance of H1
Drift-Rate correct positive	6.258	0.013	3.392	Effect for Sex*Experiment interaction, no effect for sex when Experiment 6 was excluded from analysis
Drift-Rate error negative	1.854	0.174	0.372	Undecided
Drift-Rate error positive	0.11	0.74	0.126	Acceptance of H0

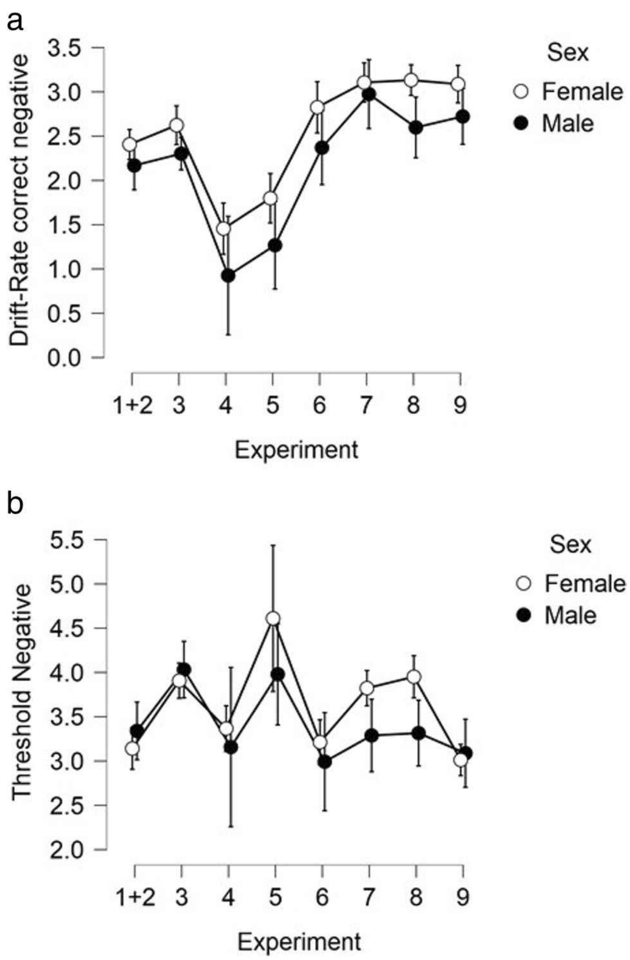

Fig. 3 (**a**) Sex differences in Drift-Rate correct in negative emotions for all experiments. (**b**) Sex differences in Threshold in negative emotions for all experiment. Error bars represent 90% credible intervals

Additional non-pre-registered analyses

Following the finding of sex differences in Drift-Rate in negative emotions, we wanted to rule out alternative explanations regarding general differences in speed or adherence to norms. For this purpose, we performed t-tests between sexes for RTs and accuracy rates (adherence to norms) specifically in negative emotions. We could not find any significant evidence for sex differences in negative emotions in RTs ($BF_{10} = 1.001$) or in accuracy rates ($BF_{10} = 2.09$).

Discussion

Do men and women differ in their subjective experience of emotion? The extant research on this topic suggests a positive answer to this question. However, this research has typically employed raw self-reports which have their serious challenges. In the current study, we have tried to overcome all the challenges of which we were aware. We have measured truly felt in-the-moment emotion experiences, and thus have minimized the influence of stereotypes that is most serious with global and retrospective reports. We have further supplied the response labels in advance, thus overcoming issues associated with potential sex-differences in the accessibility of emotional labels. Moreover, although we have studied emotion intensity, participants made binary choices of pleasant vs. unpleasant, thus minimizing whatever influence of gender stereotypes on the reported intensity of the felt emotion. Perhaps most importantly, we employed an explicit measurement model (Borsboom, 2006) which links feeling, a latent variable, with observable report.

Across nine different experiments, women showed a higher rate of evidence accumulation (Drift-Rate), as compared to men, but only in negative emotions. In other words, when facing a normatively negative stimulus, women accumulate evidence favouring reporting negative emotion more efficiently than men. To simplify, when men and women are asked to indicate whether photos previously judged to convey negative emotions are "pleasant" or "unpleasant", women either have shorter reaction times, make fewer normatively aberrant responses, or both as compared with men. Note however that our core measure was not the observed report or its corresponding response-time. Rather, we have focused on the underlying latent process (rate of evidence accumulation) that presumably expresses in reports and response-time. Importantly, our results show that this difference cannot be explained by sex differences in general speed or in adherence to stimulus norms. Sex differences in other LBA parameters as well as in positive emotions, were not found.

Finding an effect in the Drift-Rate but not in the Threshold, allows us to make inferences regarding the underlying process of observed sex differences in emotion-reports, as these two parameters account for distinct mechanisms. The findings suggest that while women display greater sensitivity to emotional evidence in response to a negative stimulus, they do not need less evidence than men in order to feel negative emotions. Thus, our results indicate that the stereotype of women being more emotional is partially supported, but only regarding negatively valenced information and only with respect to efficiency (women are more efficient than men) and not in terms of bias (because women do not need less evidence to feel a negative emotion). This finding seems to align with Rattel et al.'s (2020) showing that in women, there is greater concordance between reports and other emotion indicators. Specifically, given that the evidence which accumulates to support experiencing emotion arguably comes from other emotion components (such as physiology, see Givon et al., 2020), one would predict greater concordance to reflect a higher Drift-Rate (more evidence favoring feeling).

The current work additionally compared negative and positive emotions. We found a null effect of sex in positive emotions that was coupled with the clear sex effect in negative emotions. Importantly, the current findings concerning valence asymmetry are not an outlier and align well with recent work (Jones et al., 2020) that showed greater negativity bias in women. Specifically, negativity bias is the tendency to attend to, learn from, and use negative information far more than positive information, and specifically experience and recall more negative emotions (Vaish et al., 2008). Our findings additionally align with similar findings regarding sex differences in negative emotions (e.g., Fischer et al., 2004; Gard & Kring, 2007; Hillman et al., 2004) as well as the higher incidence of depression among women (e.g., Hopcroft & Bradley, 2007; McGuire & Troisi, 1998).

The aforementioned valence asymmetry can help narrow down the range of possible explanations. Specifically, the theoretical debate regarding sex differences focuses mostly on two central approaches: the evolutionary perspective (e.g., Ellis, 2011; Hehman & Salmon, 2020; Schmitt, 2015) and the social role theory (e.g., Eckes & Trautner, 2012; Grossman & Wood, 1993). A plausible relevant evolutionary explanation (traditionally tested in studies of emotional facial recognition) would be the primary caretaker hypothesis proposed by Babchuk et al. (1985). According to this hypothesis, the evolutionary role that females play as primary caregivers would result in evolved adaptations that increase the survival chances of their offspring. Hampson et al. (2006), suggested two partly contradicting interpretations for the primary caretaker hypothesis. The first, the "attachment promotion" hypothesis, suggests that mothers who are responsive to all of their babies' non-verbal signals (i.e., crying, smiling) are likely to have secure attachment with their babies (Ainsworth, 1979), that in turn will display optimal immune system and better social outcomes. This hypothesis predicts a female superiority across the entire emotional spectrum, not being restricted to negative emotions. A different interpretation of the primary caretaker hypothesis is the "fitness threat" hypothesis, which suggests that the advantage of women is restricted to negative emotions, because they serve as an indication for potential threat to the survival of offspring (Hampson et al., 2006).

From the perspective of the social role theory (Eckes & Trautner, 2012; Grossman & Wood, 1993), gender roles facilitate gender typical behaviours. Indeed, in many cultures, women are expected to be more compassionate and other-oriented, which in turns facilitate their activities within the family and their work in typically feminine occupations such as teacher, social worker etc. We tend to think that this hypothesis also predicts valence asymmetry because adopting a compassionate position usually implies sharing negative, not positive emotions.

An additional (not mutually exclusive) explanation for valence asymmetry considers the power differential between the sexes. Kemper (1978) argues that people with higher status and more power (men) tend to experience more positive emotions whereas people with lesser power and status (women) tend to experience negative emotions such as sadness and anxiety. Another issue concerning the power differential is the physical size: women are physically smaller and weaker than men, on average, and are thus more prone to be attacked, physically. For example, according to a US survey (Truman & Morgan, 2014), women are about three times more likely to be victims of domestic violence than men. This observation further stresses the greater need of women to be sensitive to danger cues and other potential sources of misfortune. Indeed, a recent evolutionary theory, "the fearful ape hypothesis", suggests that heightened fearfulness is adaptive, and is higher in humans than in chimps, for example (Grossmann, 2022).

To sum up, of all the various explanations that we have outlined, only one seems to have difficulty with the current findings concerning valence asymmetry. The "attachment promotion" hypothesis, would predict female advantage (higher Drift-Rate) in both positive and negative emotions. Admittedly, the current study may not be optimally suited to deal with the valence asymmetry issue because we have only studied (un)pleasantness and did not study specific pleasant and unpleasant emotions.

Our study has several additional limitations. First, our participants are not a representative sample, since they are all Israeli and almost all of them were students when tested, i.e., relatively high functioning young adults. Second, while the effect in Drift-Rate was replicated across different experiments, the conclusions are limited considering the type of stimuli (pictures, with porn and morbid curiosity-related pictures excluded) and to the type of emotions (pleasant and unpleasant). Future research should thus explore these sex differences in other cultures and in wider (including clinical) populations, and should be extended to emotions other than pleasantness and unpleasantness and to non-visual stimuli.

In conclusion, using online (rather than retrospective) binary emotion reports together with the LBA model, the current study helps overcoming methodological shortcomings associated with studying sex differences in felt emotions. This study additionally shed light on the mechanism underlies sex differences in negative emotions. Regarding the core latent variables of the LBA model, our findings indicate that men and women share an equal Threshold, while women showed higher Drift-Rate than men. It seems that women are not simply more "emotional" than men, but rather that women are better able than men to detect relevant negative emotional information. Arguably, the greater responsiveness of women to negative information may have been abused in the past to depict women's emotional responses as irrational. Our results indicate just the opposite of this stereotype. They show that women are as rational as men in terms of being equal to men in the amount of evidence required to report unpleasantness (Threshold). The results instead show that greater responsiveness of women to negative information is not a matter of irrationality but a matter of greater efficiency to correctly detect information that is relevant to negative emotional experiences.

Funding This work was supported by Grant 1547/20 from the Israel Science Foundation to Nachshon Meiran.

Data availability The datasets generated and analysed during the current study are available in the Open Science Framework (see links for each experiment in Method).

Declarations

Conflict of interest The authors declared no conflicts of interest with respect to the authorship or the publication of this article.

References

Ainsworth, M. D. S. (1979). Attachment as related to mother-infant interaction. In J. S. Rosenblatt, R. A. Hinde, C. Beer, & M.-C. Busnel (Eds.), *Advances in the study of behavior* (Vol. 9, pp. 1–51). Academic Press. https://doi.org/10.1016/S0065-3454(08)60032-7

Babchuk, W. A., Hames, R. B., & Thompson, R. A. (1985). Sex differences in the recognition of infant facial expressions of emotion: The primary caretaker hypothesis. *Ethology and Sociobiology, 6*(2), 89–101.

Barrett, L. F., Robin, L., Pietromonaco, P. R., & Eyssell, K. M. (1998). Are women the "more emotional" sex? Evidence from emotional experiences in social context. *Cognition & Emotion, 12*(4), 555–578. https://doi.org/10.1080/026999398379565

Berkovich, R., & Meiran, N. (2022). Pleasant emotional feelings follow one of the most basic psychophysical laws (weber's law) as most sensations do. *Emotion*.

Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika, 71*(3), 425–440.

Bradley, M. M., Codispoti, M., Sabatinelli, D., & Lang, P. J. (2001). Emotion and motivation II: Sex differences in picture processing. *Emotion, 1*(3), 300–319.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics, 7*(4), 434–455. https://doi.org/10.1080/10618600.1998.10474787

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57*(3), 153–178. https://doi.org/10.1016/j.cogpsych.2007.12.002

Deng, Y., Chang, L., Yang, M., Huo, M., & Zhou, R. (2016). Gender differences in emotional response: Inconsistency between experience and expressivity. *PLoS ONE, 11*(6), e0158666. https://doi.org/10.1371/journal.pone.0158666

Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E. J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin & Review, 18*(1), 61–69.

Donkin, C., & Brown, S. D. (2018). Response times and decision-making. *Stevens' handbook of experimental psychology and cognitive neuroscience*, 349–377.

Durik, A. M., Hyde, J. S., Marks, A. C., Roy, A. L., Anaya, D., & Schultz, G. (2006). Ethnicity and gender stereotypes of emotion. *Sex Roles, 54*(7), 429–445. https://doi.org/10.1007/s11199-006-9020-4

Eckes, T., & Trautner, H. M. (2012). *The developmental social psychology of gender*. Psychology Press.

Ellis, L. (2011). Evolutionary neuroandrogenic theory and universal gender differences in cognition and behavior. *Sex Roles, 64*(9), 707–722.

Filkowski, M. M., Olsen, R. M., Duda, B., Wanger, T. J., & Sabatinelli, D. (2017). Sex differences in emotional perception: Meta analysis of divergent activation. *NeuroImage, 147*, 925–933. https://doi.org/10.1016/j.neuroimage.2016.12.016

Fischer, A., & LaFrance, M. (2015). What drives the smile and the tear: Why women are more emotionally expressive than men. *Emotion Review, 7*(1), 22–29.

Fischer, A. H., Rodriguez Mosquera, P. M., van Vianen, A. E. M., & Manstead, A. S. R. (2004). Gender and culture differences in emotion. *Emotion, 4*(1), 87–94. https://doi.org/10.1037/1528-3542.4.1.87

Fugate, J. M. B., Gouzoules, H., & Barrett, L. F. (2009). Separating production from perception: Perceiver-based explanations for sex differences in emotion. *Behavioral and Brain Sciences, 32*(5), 394–395.

Gard, M. G., & Kring, A. M. (2007). Sex differences in the time course of emotion. *Emotion, 7*(2), 429–437. https://doi.org/10.1037/1528-3542.7.2.429

Givon, E., Itzhak-Raz, A., Karmon-Presser, A., Danieli, G., & Meiran, N. (2020). How does the emotional experience evolve? Feeling generation as evidence accumulation. *Emotion, 20*(2), 271–285. https://doi.org/10.1037/emo0000537

Givon, E., Udelsman-Danieli, G., Almagor, O., Fekete, T., Shriki, O., & Meiran, N. (2022). Can feelings "feel" wrong? Similarities between counter-normative emotion reports and perceptual errors. *Psychological Science*, 09567976211063915.

Grossman, M., & Wood, W. (1993). Sex differences in intensity of emotional experience: A social role interpretation. *Journal of Personality and Social Psychology, 65*(5), 1010.

Grossmann, T. (2022). The human fear paradox: Affective origins of cooperative care. *Behavioral and Brain Sciences*, 1–43.

Hampson, E., van Anders, S. M., & Mullin, L. I. (2006). A female advantage in the recognition of emotional facial expressions: Test of an evolutionary hypothesis. *Evolution and Human Behavior, 27*(6), 401–416. https://doi.org/10.1016/j.evolhumbehav.2006.05.002

Hehman, J. A., & Salmon, C. A. (2020). Beyond sex differences: Predictors of negative emotions following casual sex. *Evolutionary Psychological Science, 6*(2), 97–108.

Hillman, C. H., Rosengren, K. S., & Smith, D. P. (2004). Emotion and motivated behavior: Postural adjustments to affective picture viewing. *Biological Psychology, 66*(1), 51–62. https://doi.org/10.1016/j.biopsycho.2003.07.005

Hopcroft, R. L., & Bradley, D. B. (2007). The sex difference in depression across 29 countries. *Social Forces, 85*(4), 1483–1507.

Hyde, J. S., Mezulis, A. H., & Abramson, L. Y. (2008). The ABCs of depression: Integrating affective, biological, and cognitive models to explain the emergence of the gender difference in depression. *Psychological Review, 115*(2), 291.

Itkes, O., Kimchi, R., Haj-Ali, H., Shapiro, A., & Kron, A. (2017). Dissociating affective and semantic valence. *Journal of Experimental Psychology: General, 146*(7), 924–942. https://doi.org/10.1037/xge0000291

JASP Team. (2022). JASP (Version 0.16.2) [Computer software]. https://jasp-stats.org/

Jones, L. L., Wurm, L. H., Norville, G. A., & Mullins, K. L. (2020). Sex differences in emoji use, familiarity, and valence. *Computers in Human Behavior, 108*, 106305. https://doi.org/10.1016/j.chb.2020.106305

Kelly, M. M., Forsyth, J. P., & Karekla, M. (2006). Sex differences in response to a panicogenic challenge procedure: An experimental evaluation of panic vulnerability in a non-clinical sample. *Behaviour Research and Therapy, 44*(10), 1421–1430. https://doi.org/10.1016/j.brat.2005.10.012

Kemper, T. D. (1978). Toward a sociology of emotions: Some problems and some solutions. *The American Sociologist*, 30–41.

Kret, M. E., & De Gelder, B. (2012). A review on sex differences in processing emotional signals. *Neuropsychologia, 50*(7), 1211–1221. https://doi.org/10.1016/j.neuropsychologia.2011.12.022

Kring, A. M., & Gordon, A. H. (1998). Sex differences in emotion: Expression, experience, and physiology. *Jornal of Personality and Social Psychology, 74*(3), 686.

Kross, E., Verduyn, P., Boyer, M., Drake, B., Gainsburg, I., Vickers, B., Ybarra, O., & Jonides, J. (2019). Does counting emotion words on online social networks provide a window into people's subjective experience of emotion? A case study on Facebook. *Emotion, 19*(1), 97.

LeDoux, J. E., & Hofmann, S. G. (2018). The subjective experience of emotion: A fearful view. *Current Opinion in Behavioral Sciences, 19*, 67–72. https://doi.org/10.1016/j.cobeha.2017.09.011

Lin, Y.-S., & Strickland, L. (2020). Evidence accumulation models with R: A practical guide to hierarchical Bayesian methods. *The Quantitative Methods for Psychology, 16*(2), 133–153.

Maffei, A., & Angrilli, A. (2019). E-MOVIE-Experimental MOVies for Induction of Emotions in neuroscience: An innovative film database with normative data and sex differences. *PLoS ONE, 14*(10), e0223124.

Marchewka, A., Żurawski, Ł., Jednoróg, K., & Grabowska, A. (2014). The Nencki Affective Picture System (NAPS): Introduction to a novel, standardized, wide-range, high-quality, realistic picture database. *Behavior Research Methods, 46*(2), 596–610. https://doi.org/10.3758/s13428-013-0379-1

McDuff, D., Kodra, E., Kaliouby, R. E., & LaFrance, M. (2017). A large-scale analysis of sex differences in facial expressions. *PLoS ONE, 12*(4), e0173942.

McGuire, M. T., & Troisi, A. (1998). Prevalence differences in depression among males and females: Are there evolutionary explanations? *British Journal of Medical Psychology, 71*(4), 479–491.

R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(5), 1226.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*(4), 873–922.

Rattel, J. A., Mauss, I. B., Liedlgruber, M., & Wilhelm, F. H. (2020). Sex differences in emotional concordance. *Biological Psychology, 151*, 107845.

Scherer, K. R. (1987). Toward a dynamic theory of emotion: The component process model of affective states. *Geneva Studies in Emotion and Communication, 1*, 1–98.

Schmitt, D. P. (2015). The evolution of culturally-variable sex differences: Men and women are not always different, but when they are… it appears not to result from patriarchy or sex role socialization. In *The evolution of sexuality* (pp. 221–256). Springer, Cham.

Schubert, A.-L., Hagemann, D., Voss, A., & Bergmann, K. (2017). Evaluating the model fit of diffusion models with the root mean square error of approximation. *Journal of Mathematical Psychology, 77*, 29–45. https://doi.org/10.1016/j.jmp.2016.08.004

Schwarz, N. (2012). Why researchers should think "real-time": A cognitive rationale. *Handbook of research methods for studying daily life*, 22.

Simon, R. W., & Nath, L. E. (2004). Gender and emotion in the United States: Do men and women differ in self-reports of feelings and expressive behavior? *American Journal of Sociology, 109*(5), 1137–1176.

Singer-Landau, E., & Meiran, N. (2021). Cognitive appraisal contributes to feeling generation through emotional evidence accumulation rate: Evidence from instructed fictional reappraisal. *Emotion, 21*(7), 1366–1378. https://doi.org/10.1037/emo0001006

Steiger, J. H., & Lind, J. (1980). Statistically based tests for the number of common factors. *In the annual meeting of the Psychometric Society. Iowa City, IA. 1980*. https://ci.nii.ac.jp/naid/10012870999/

Stevens, J. S., & Hamann, S. (2012). Sex differences in brain activation to emotional stimuli: A meta-analysis of neuroimaging studies. *Neuropsychologia, 50*(7), 1578–1593. https://doi.org/10.1016/j.neuropsychologia.2012.03.011

Truman, J. L., & Morgan, R. E. (2014). *Nonfatal domestic violence*. US Department of Justice, Bureau of Justice Statistics.

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social-emotional development. *Psychological Bulletin, 134*(3), 383–403.

Volynets, S., Glerean, E., Hietanen, J. K., Hari, R., & Nummenmaa, L. (2020). Bodily maps of emotions are culturally universal. *Emotion, 20*(7), 1127.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063