Universiteit Leiden

Master Creative Intelligence & Technology

Beyond the Mood Swings: Tracking Anxiety and Social Understanding Across the Menstrual Cycle

Name: Nikki Rademaker
Student ID: 2641887

Date: [03/12/2025]

1st supervisor: Dr. Tessa Verhoef
2nd supervisor: Dr. Liisa Hantsoo

MSc Creative Intelligence & Technology (former Media Technology)

Leiden Institute of Advanced Computer Science
Leiden University
Einsteinweg 55
2333 CC Leiden
The Netherlands

Beyond the Mood Swings: Tracking Anxiety and Social Understanding Across the Menstrual Cycle

Nikki Rademaker
n.s.rademaker@umail.leidenuniv.nl

MSc Creative Intelligence & Technology
Leiden Institute of Advanced Computer Science (LIACS)
Leiden University

Primary supervisor:
Dr. T. Verhoef, Leiden University

Secondary supervisor:
Dr. L. Hantsoo, Johns Hopkins University School of Medicine

December 3, 2025

Abstract. Research on how ovarian hormonal changes across the menstrual cycle influence anxiety and social cognition remains limited. This study examined whether naturally cycling individuals experience varying levels of state anxiety and affective Theory of Mind (ToM) performance across menstrual phases. For two full menstrual cycles, participants completed daily surveys measuring state anxiety using the State-Trait Anxiety Inventory. They also completed a total of four affective ToM tests based on the Adult Faux Pas Recognition Test. State anxiety was assessed daily across three phases (menstrual, follicular, luteal), and affective ToM performance was assessed during two phases (mid-follicular and mid-to-late luteal). A repeated measures ANOVA showed a significant difference in anxiety between menstrual cycle phases. The mean anxiety score was highest during menstruation and significantly greater than during the follicular phase. The lowest anxiety was found in the follicular phase excluding menstruation. No significant difference was observed between the luteal phase and the other phases. Paired samples t-tests showed no significant difference in affective ToM performance between the mid-follicular and mid-to-late luteal phases. In conclusion, the findings indicate that the menstrual cycle influences anxiety levels but not affective ToM performance. However, the small sample sizes ($n = 20$ for anxiety, $n = 17$ for ToM) and phase predictions based on self-reported cycle data limit the generalizability of these results. The findings contribute to closing the data gap on how the menstrual cycle affects everyday fluctuations in anxiety and take a step toward exploring potential phase effects on social cognition.

Keywords: Menstrual Cycle · State Anxiety · Affective Theory of Mind.

1 Introduction

In 2024, the global population was estimated to be 8.2 billion people [1]. Of this population, about 1.8 billion menstruate each month [2][3]. This means that roughly 22% of the world's population currently undergoes the menstrual cycle. The menstrual cycle does not come without impact. It was shown that among 42,000 women who completed a survey on menstrual symptoms, 85% of them reported abdominal pain during menstruation. In addition, 77% experienced psychological complaints before and during their period, and 38% indicated that they were not able to carry out all of their daily activities [4]. Altogether, this shows that the menstrual cycle influences many people worldwide. The impact is not only physical but also psychological. Despite affecting nearly a quarter of the global population and the reported symptoms of pain and mood changes, research on the menstrual cycle has been systematically overlooked. As a result, data gaps remain in our understanding of the overall menstrual cycle and its effects, largely due to its under-researched nature and complexity. [5][6].

The menstrual cycle is a recurring process regulated by hormonal fluctuations that influence physical and psychological functioning. It prepares the female body for ovulation and a potential pregnancy. However, if pregnancy does not occur, hormone levels drop, and the cycle repeats itself. A person who menstruates goes through four phases each cycle: menstrual, follicular, ovulation, and luteal [7]. These phases are characterized by fluctuations in hormones such as estrogen and progesterone.
Although data gaps in research on the menstrual cycle exist overall, significant gaps also remain in understanding how menstrual cycle-related hormonal changes affect cognition and mental health in females. In this context, hormonal changes refer to fluctuations in ovarian hormones. Estrogen and progesterone are ovarian hormones that have been found to influence mood, anxiety, and brain function [8][9]. This means that these hormonal fluctuations could contribute to variations in feelings of nervousness or tension throughout the cycle. While studies have examined the ovarian hormone fluctuations during the menstrual cycle, research exploring the potential effects of the cycle on general anxiety levels remains limited. Most studies focus on clinical disorders or do not look at daily anxiety fluctuations. Bridging this data gap is important. Understanding how the menstrual cycle affects everyday anxiety can improve knowledge of mental well-being across the population. These insights can help individuals recognize and manage their anxiety more effectively. They may allow people to plan around their symptoms or understand that their experiences are not irrational. Moreover, raising awareness of these effects within society is essential so that the influence of the menstrual cycle is more widely acknowledged and considered.

To address the gap in research, this thesis aims to explore how the cycle phases (menstrual, follicular, and luteal) affect individuals' anxiety throughout the menstrual cycle while looking at daily general anxiety levels. In this context, general anxiety refers to everyday feelings of anxiety. This includes feelings of tension, apprehension, nervousness, unease, and worry. The paper focuses on anxiety as an emotion rather than clinical anxiety disorders. Based on this, the main research question this thesis investigates is:
"How do general anxiety levels vary across different phases of the menstrual cycle in individuals who menstruate and do not use hormonal birth control?"
This leads to the following hypothesis: *In naturally cycling people, general anxiety levels vary across menstrual cycle phases. Anxiety is expected to be higher during the menstrual phase and luteal phase when estrogen and progesterone are assumed to be low or declining. Furthermore, anxiety levels will likely be lower before and around ovulation when estrogen is likely to be high.*

Fluctuations in mood and anxiety also raise the question of whether menstrual cycle phases influence the ability to recognize and respond to others' emotions. Being able to understand what someone else is feeling is related to a concept called affective Theory of Mind (ToM). Anxiety has been linked to difficulties in ToM performance, especially when it comes to under-

standing others' emotions [10]. At the same time, ovarian hormonal shifts across the menstrual cycle appear to affect social sensitivity and emotion recognition [11]. Consequently, they could also influence affective ToM. Affective ToM is relevant for social functioning and daily interactions [12]. It plays a crucial role in fostering appropriate social behavior, which is essential for maintaining relationships and overall mental well-being. Still, research on whether the menstrual cycle influences affective ToM abilities remains very limited. For this reason, this paper looks at both anxiety levels and affective ToM performance across different menstrual cycle phases. The sub-research question that will be answered is:
"How does affective Theory of Mind performance of naturally cycling individuals differ between the mid-follicular and the mid-to-late luteal phase of the menstrual cycle?"
Accordingly, the hypothesis is: *Affective Theory of Mind performance differs during the mid-follicular phase and the mid-to-late luteal phase. Moreover, performance will likely be better during the mid-follicular phase than during the mid-to-late luteal phase in naturally cycling people, due to suspected lower anxiety levels.*

To answer the research questions and test the hypotheses, participants complete daily surveys across two full menstrual cycles to track anxiety levels during different phases. Anxiety is measured with the State-Trait Anxiety Inventory, which produces a numerical score. Additional information is also collected. This includes medication intake, diagnosed anxiety disorders (e.g., Generalized Anxiety Disorder, Social Anxiety Disorder, Panic Disorder, Post-Traumatic Stress Disorder), other diagnoses that participants believed could influence their anxiety, and diagnosed menstrual disorders. A repeated measures ANOVA is used to test whether anxiety scores differ significantly between menstrual cycle phases. This is followed by pairwise post-hoc comparisons to determine between which phases the differences are significant. Participants are also asked to complete four tests to assess affective ToM abilities during two different phases of the menstrual cycle. These tests are conducted on cycle days 7-10 (mid-follicular phase) and 21-24 (mid-to-late luteal phase) for two cycles. To compare ToM performance between the two phases, scores are analyzed using a paired samples t-test.

The remainder of this thesis discusses the structure and findings of the research. Section 2 explains the main concepts needed to understand the study, including the menstrual cycle, hormonal regulation, anxiety, and affective ToM. After this, Section 3 reviews earlier research related to this study. Sections 4 and 5 give an overview of the methodology, describing the participants, experimental setup, data collection, and data analysis. This is followed by the results of the data analysis in Section 6. Subsequently, Section 7 discusses these results and their implications. This includes limitations and directions for future research. To conclude, Section 8 summarizes the key findings and answers the research questions.

2 Background

This research brings together three areas: the menstrual cycle and its ovarian hormonal fluctuations, changes in state anxiety, and the concept of ToM in social cognition. The following section outlines these concepts and their relevance to the study.

2.1 The Menstrual Cycle

As mentioned earlier, the menstrual cycle is a recurring physiological process that prepares the female body for the possibility of pregnancy. It begins at menarche (the first menstrual period), which typically occurs around age twelve. The cycle continues to repeat itself throughout the reproductive years, except during pregnancy, until menopause [13][14]. Menopause usually begins in a female's late forties or early fifties [7][15]. For most individuals, cycle lengths range between 21 and 35 days, with 28 days as the average reference [16][17]. Some medical sources consider cycles of up to 38 days to still fall within the normal range [18][19]. While many individuals have cycles within these ranges, others experience irregular menstrual cycles

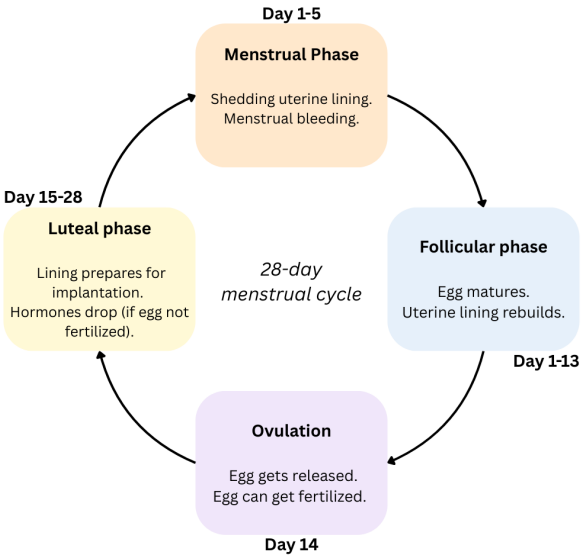

Fig. 1: Overview of the four phases of a 28-day menstrual cycle. The cycle begins with the menstrual and follicular phase, followed by ovulation and the luteal phase. The menstrual and follicular phases start on the same day, but the follicular phase continues for a longer duration.

(atypical patterns in cycle length or timing) or menstrual disorders. A menstrual disorder is a clinical condition involving abnormal menstrual symptoms, severe pain, or disruptions in normal bleeding patterns [20]. Examples of such disorders are premenstrual dysphoric disorder (PMDD), dysmenorrhea, and endometriosis. However, the menstrual cycle generally follows a typical pattern in most individuals. The menstrual cycle has different phases, as can be seen in Figure 1. The cycle phases will be further explained using a 28-day cycle as the baseline. The phases involve changes in the ovaries and the uterine lining. This is mainly guided by fluctuations in estrogen, progesterone, follicle-stimulating hormone (FSH), and luteinizing hormone (LH) [7]. It is commonly divided into four different phases:

1. The **menstrual** phase (day 1-5): The start of a new menstrual cycle happens on the first day of menstruation. This is marked by menstrual bleeding. The menstrual phase occurs when the egg from the previous cycle has not been fertilized. This leads to a drop in estrogen and progesterone levels, causing the uterine lining to shed [7][16]. The menstrual phase typically lasts around five days, although a duration up to seven days is also considered normal [14][21].
2. The **follicular** phase (day 1-13): The follicular phase starts on the first day of the period and continues until ovulation. It can vary in length for different people. In this phase, the anterior pituitary gland, located near the base of the brain, releases FSH. This stimulates the growth of several follicles. A follicle is a tiny sac of fluid in the ovary that contains an immature egg [21]. Usually, only one follicle continues to mature [16]. During menstruation, the uterine lining is shed. However, as the follicle develops, estrogen is produced. Estrogen helps the uterine lining thicken again, preparing the body for potential pregnancy [21].
3. The **ovulation** phase (day 14): Ovulation occurs around the midpoint of the cycle, approximately 14 days before menstruation. For a 28-day cycle, ovulation typically occurs around day 14 [13][22]. High levels of estrogen trigger an increase in the release of LH [16]. This causes the dominant follicle to release a mature egg from the ovary. The egg then travels into the fallopian tube, where fertilization can take place. Ovulation typically lasts for about 24 hours. During ovulation, the egg can be fertilized by sperm. The days leading up to ovulation also fall within the fertile window since sperm can survive for several days [21].

4. The **luteal** phase (day 15-28): Around day 15, after ovulation, the luteal phase starts. The luteal phase typically lasts between 11 and 17 days. In most cycles, it is about 12 to 14 days [23]. During this phase, the follicle transitions into the corpus luteum, and the corpus luteum produces progesterone and estrogen [17]. The hormones help prepare the uterus by thickening its lining in case fertilization of the egg has occurred. However, if fertilization has not occurred, the corpus luteum breaks down, and hormone levels will decline [16]. Then, the uterine lining will be shed during menstruation as the new cycle begins.

The menstrual cycle can also be described as the ovarian cycle (includes the follicular, ovulation, and luteal phases), and the uterine cycle (consists of the menstrual, proliferative, and secretory phases) [24]. In this paper, it is explained as one cycle with four phases for clarity, as this perspective provides all the information relevant to this thesis.

As explained above, ovarian hormone levels fluctuate throughout the cycle phases. Estrogen and progesterone are ovarian hormones that play crucial roles in regulating the menstrual cycle. However, these two hormones have also been linked to mood regulation [8][9]. In short, as can be seen in Figure 2, during the menstrual phase, estrogen and progesterone are at their lowest levels. In the follicular phase, estrogen rises as the uterine lining thickens. Around ovulation, estrogen levels peak and then later drop. During the luteal phase, the uterus prepares for possible implantation of a fertilized egg. This is when progesterone rises quickly and estrogen rises slightly. If fertilization does not occur, both hormone levels drop and the cycle restarts [25][26].
According to a literature review by Del Río et al. (2018), estrogen and progesterone are associated with emotional regulation and mental health across life stages. Estrogen supports brain functions such as memory, learning, and emotional states in females. It influences the activity of serotonin and dopamine, which are neurotransmitters important for mood regulation [8]. Low estrogen levels are associated with lower serotonin activity and weaker dopamine signaling. This could negatively affect mood and cause emotions of sadness and anxiety [8][27]. Progesterone and its metabolite allopregnanolone enhance signaling through GABA by interacting with GABA receptors. GABA is a major inhibitory neurotransmitter in the brain. It works by blocking specific signals in the central nervous system. This can help in reducing anxiety. Low levels of allopregnanolone have also been linked to depressive symptoms [8]. This suggests that the hormones regulating the menstrual cycle also affect neurotransmitter systems involved in emotion and mood. Consequently, anxiety levels might fluctuate with hormonal changes per menstrual phase rather than remaining stable across the cycle.

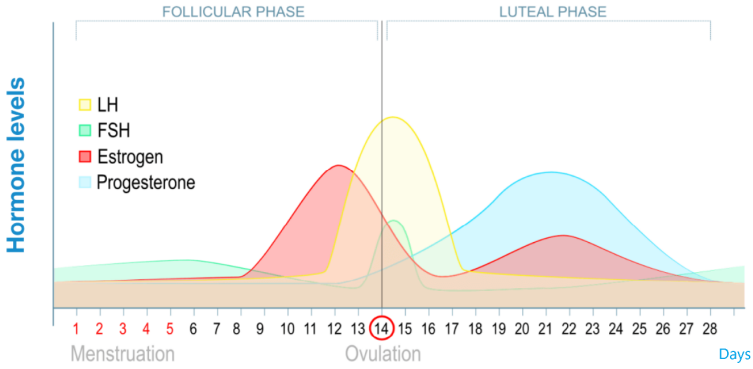

Fig. 2: Overview of the hormone levels of luteinizing hormone (LH), follicle-stimulating hormone (FSH), estrogen, and progesterone across the menstrual cycle [28]. LH and FSH levels peak during ovulation. Estrogen levels are highest after menstruation during the late follicular phase. Progesterone rises during the luteal phase. Estrogen and progesterone levels are low before menstruation.

2.2 State Anxiety

Anxiety can be described as feelings of tension, nervousness, fear, and worry. It is often accompanied by physical responses such as an increased heart rate or restlessness. Experiencing anxiety in stressful situations is normal. However, when it is persistent or excessive, it can interfere with daily functioning [29]. In this research, anxiety is meant as a general state of anxiety that is not necessarily tied to a disorder.

The State-Trait Anxiety Inventory (STAI) is a validated and commonly used tool for assessing anxiety. The STAI was developed by Spielberger, Gorsuch, and Lushene in 1964 [30]. The survey distinguishes between two forms of anxiety:

- **State anxiety** (STAI-S): a temporary feeling of anxiety or nervousness. It reflects how a person feels in the moment or in a specific situation.
- **Trait anxiety** (STAI-T): a more lasting form of anxiety. It reflects how often a person tends to feel anxious in general.

The full STAI contains forty statements, twenty for each anxiety type. Answers are given on a four-point scale, with some items being reverse-coded. The answers given for assessing trait anxiety tend to remain stable. On the other hand, for state anxiety, answers fluctuate depending on circumstances. This research only uses the STAI-S to measure state anxiety, as this enables tracking daily fluctuations in anxiety across the menstrual cycle. Participants complete the state anxiety questions each day, reflecting on how anxious they felt throughout the day. More information on this can be found in Section 4.2.2.

2.3 Theory of Mind

ToM refers to the ability to infer and recognize mental states of oneself and others. Examples of such mental states are emotions, beliefs, intentions, desires, and thoughts [31][32]. It also means being able to understand that these mental states might be different from one's own [33]. ToM is commonly divided into two components:

- **Cognitive ToM**: the ability to understand what others know, think, or intend. It involves making inferences about beliefs and motivations [33].
- **Affective ToM**: the ability to understand what others feel, and to infer their emotional state. It builds on cognitive ToM and also depends on empathy [32][33]. Brain studies indicate that it involves a region of the brain that integrates emotional and cognitive information during social interactions [32].

In this research, the focus is on affective ToM as it is essential for understanding the emotional side of social interactions. It helps people recognize how others feel and respond in appropriate ways [32]. As mentioned earlier, anxiety levels may vary across menstrual cycle phases, and anxiety has been suggested to negatively impact affective ToM [10]. Therefore, this study examines both anxiety and affective ToM. The Adult Faux Pas Recognition Test (FPRT) will be used, developed by Stone and Baron-Cohen [34]. Its tasks are designed to assess social sensitivity and ToM abilities. The Adult FPRT consists of twenty short written stories where a character may unintentionally say something inappropriate or hurtful (a faux pas). After each story, participants answer a series of questions to assess whether they detected a faux pas and understood the characters' feelings. This will be further explained in Section 4.2.3.

3 Related Work

Research on the menstrual cycle has long been underrepresented. Knowledge of its physiological mechanisms remains limited, and there is no consistent approach to studying cycle-related topics [6]. While research on this topic has become more common, important knowledge gaps

persist. How ovarian hormonal fluctuations across the cycle influence psychological processes, such as mood and cognition, remains unclear [35]. Before addressing these gaps in the present study, it is necessary to analyze the available research. Therefore, this section reviews relevant studies in three areas. It starts by discussing studies on anxiety across the menstrual cycle. This is followed by research about the relationship between anxiety and ToM abilities, and studies on ovarian hormonal influences on emotion recognition and empathy. Finally, it outlines methodological approaches used in earlier work to measure anxiety and ToM. Parts of these approaches were adapted for use in this thesis.

3.1 Anxiety Across the Menstrual Cycle

Research has consistently shown that hormonal shifts in progesterone and estrogen across menstrual phases impact mood and anxiety. Reynolds et al. (2018) conducted a longitudinal study where they measured progesterone and general anxiety at three different points in the menstrual cycle. The participants had no diagnosed conditions affecting the female reproductive system and reported regular menstrual cycles. Anxiety-related conditions were not screened for [36]. They found that women with higher average progesterone levels across their cycles reported more anxiety. However, within-person fluctuations in progesterone were not significantly associated with anxiety [36]. In a review, Wieczorek et al. (2023) discussed evidence that fluctuations in estrogen and progesterone influence mental health through serotonin, dopamine, and GABA pathways in both healthy and clinical women. Some studies measured hormones directly, but others only looked at reported symptoms across menstrual cycle phases. They reported that mood and anxiety symptoms often worsen during the luteal and menstrual phases, which is when estrogen and progesterone levels (eventually) drop. Furthermore, based on the studies they reviewed, mood tends to improve around ovulation, when estrogen levels peak [9]. In other words, Reynolds et al. (2018) found differences in general anxiety between women with varying progesterone levels but not within individuals, whereas Wieczorek et al. (2023) summarized evidence of phase-related changes within individuals [9][36]. These mixed findings show that the link between ovarian hormones and anxiety is not straightforward. Adding another perspective, Hantsoo et al. (2022) analyzed premenstrual symptoms in more than 238,000 users of the Flo mobile application through a cross-sectional (one-time) survey [37][38]. Participants reported on a range of symptoms. Mood swings and anxiety were reported by about 64% of participants every menstrual cycle. These symptoms remained consistent across age groups [38]. This large-scale self-report study suggests that, regardless of hormone patterns, many women experience anxiety symptoms regularly in connection with their cycles.

Most studies have mainly or specifically examined clinical populations. Nillni et al. (2011) reviewed findings showing that women with panic disorder and PMDD often experience more anxiety and panic symptoms before menstruation. They proposed a model linking ovarian hormone changes, anxiety sensitivity, and stress. This model suggests that the decline in these hormones may lower the brain's calming mechanisms. This makes women with panic disorders or high anxiety sensitivity more likely to misinterpret bodily sensations, which possibly causes them to feel more anxious, especially when external stressors are present [39]. Furthermore, Green and Graham (2022) reviewed studies on anxiety disorder, post-traumatic stress disorder (PTSD), and obsessive-compulsive disorder (OCD) [40]. They discuss that symptoms are often worse during the premenstrual and early follicular phases when ovarian hormone levels are low, which is consistent with the phase-related changes described by Wieczorek et al. (2023) [9][40]. The review indicates that ovarian hormonal changes across the cycle may increase anxiety risk for some women, but the effects vary across individuals. They also note that most studies used small samples, sometimes with as few as 10 participants, and defined menstrual phases inconsistently. This makes the findings less reliable and more difficult to compare across studies [40].

Taken together, these studies suggest that hormonal changes across the menstrual cycle can affect anxiety and mood, though the evidence varies by population and method. Some work points to phase-related differences, while others fail to find within-person effects or report them only in a subset of participants. Much of the existing research has focused on clinical populations such as women with PMDD, panic disorders, or anxiety disorders. On the other hand, little research has examined how anxiety levels vary across menstrual phases in the general female population without necessarily having diagnosed conditions. Moreover, many studies rely on fixed measurement points, such as recording anxiety once per cycle phase in Reynolds et al. (2018), or on cross-sectional reports as in Hantsoo et al. (2022), rather than frequent measurements or daily self-reports [36][38]. Others, such as Green and Graham (2022), highlight inconsistent definitions of menstrual phases across studies [40]. Additionally, Green and Graham (2022) mainly reviewed research focusing on specific phases of the menstrual cycle rather than all phases of the whole cycle [40]. The same can be said for other discussed research. For example, Hantsoo et al. (2022) examined premenstrual symptoms, and Nillni et al. (2011) also concentrated on the premenstrual (late follicular) phase [38][39]. This shows that much of the literature mainly addresses specific phases or premenstrual symptoms, while fewer studies investigate anxiety across all menstrual phases. These approaches limit insight into whether and how anxiety fluctuates dynamically across the entire cycle. Addressing this gap through daily self-reports across complete cycles can provide a more detailed understanding of these fluctuations beyond strictly clinical samples.

3.2 Anxiety Affecting Theory of Mind Abilities

While the previous subsection reviewed how anxiety levels vary across the menstrual cycle, another question is how anxiety itself may influence social cognition. Social cognition involves the mental processes that allow individuals to recognize, interpret, and understand others and social contexts. Affective ToM is a fundamental component of social cognition [41]. Several studies have examined the relationship between anxiety and ToM abilities.

Briscoe et al. (2024) conducted a meta-analysis of multiple studies to research the relationship between anxiety and ToM. The participants in these studies were between 4 and 19 years old. They concluded that higher anxiety in children is linked to reduced ToM performance. The effect seems to be strongest for affective ToM. The studies were cross-sectional, meaning the data was collected at a single point in time, and they used trait measures of anxiety. As a result, they might not have been experiencing anxiety at the time of testing their ToM performance. [10]. Despite these limitations, this study does suggest that anxiety possibly impairs the ability to infer how someone else is feeling (affective ToM). Similarly, Baez et al. (2023) reviewed studies focusing specifically on adults with anxiety disorders. They found that social anxiety disorder was associated with impairments in emotion recognition and ToM in general. Findings for generalized anxiety disorder were inconclusive due to the small number of studies [42]. Foulds et al. (2025) did not look at clinical groups. They distinguished between state and trait anxiety, and between feelings of general and social anxiety in young adults. Using false-belief tasks, they found that state and trait anxiety (whether general or social) did not impair ToM performance [43].

Overall, findings on anxiety and ToM are mixed. Some evidence shows reduced ToM performance, especially in affective ToM, while other work finds no effect. Since this thesis examines both anxiety and affective ToM across the menstrual cycle, it is important to consider that higher anxiety levels may be associated with lower affective ToM performance.

3.3 The Effect of Ovarian Hormones on Emotion Recognition and Empathy

The previous subsection showed that anxiety might influence ToM abilities. Hormonal changes across the menstrual cycle not only affect mood and anxiety but have also been linked to social processes such as emotion recognition and empathy. These processes are closely related

to affective ToM. Since no studies seem to have directly examined affective ToM across the menstrual cycle, reviewing research on emotion recognition and empathy can provide important context for understanding how the menstrual cycle may influence affective ToM.

Most research has focused on emotion recognition as a first step in understanding how ovarian hormonal changes may shape social cognition. Early research by Derntl et al. (2008) examined the link between menstrual cycle phase and emotion recognition. They found that women in the follicular phase, when progesterone levels are low, recognized facial emotions more accurately than those in the luteal phase. They also showed that higher measured progesterone levels were linked to lower accuracy, which suggests that progesterone may have a negative effect on emotion recognition [44]. The systematic review by Osório et al. (2018) presents mixed findings, where in some cases, no significant difference was found in emotion recognition across menstrual phases. However, most of the reviewed studies have reported that the follicular phase (high estrogen and low progesterone levels) is associated with better accuracy in recognizing emotions from facial expressions [45]. Consistent with the findings of Derntl et al (2008), higher progesterone levels were often associated with lower performance [44][45]. It is important to note that while certain studies measured progesterone directly, others relied on estimates derived from menstrual cycle phases, which limits comparability across studies [45]. On the other hand, Maner and Miller (2014) found that during the early luteal phase, higher progesterone increased accuracy in reading facial expressions. They also suggest that higher progesterone makes women more attentive to social cues. The former finding relied on estimated progesterone levels from cycle day, whereas the latter was based on directly measuring progesterone levels [11]. A more recent large-scale study, however, did not find evidence for cycle-related changes in emotion recognition of faces, voices, or both combined. Rafiee et al. (2023) conclude that women's emotion recognition ability remains stable across late follicular and mid-luteal phases [46]. Jang et al. (2024) also did not find overall differences in emotion recognition when comparing the cycle phases. Instead, their analyses of ovarian hormone levels revealed a more complex pattern. Women with generally higher estrogen levels across the study performed better during the follicular phase and mid-cycle. However, temporary increases in estrogen relative to their own baseline levels seemed to worsen emotion recognition [47].

The differences in the literature show that findings are inconsistent. Several studies, such as Derntl et al. (2008), and those discussed by Osório et al. (2018) suggest better emotion recognition during the follicular phase [44][45]. In contrast, Maner and Miller (2014) found improved performance in the luteal phase [11]. Other studies, such as Rafiee et al. (2023) and Jang et al. (2024), did not detect significant phase differences in emotion recognition accuracy [46][47]. Differences in sample size, methodologies, and phase definitions likely contribute to this variation. Nonetheless, the results point to the possibility that ovarian hormone shifts may influence emotion recognition and the understanding of social cues. A remaining question is whether such effects extend beyond emotion recognition to broader concepts in social cognition. Empathy, for instance, goes a step further than recognizing basic emotions since it involves inferring, understanding, and sharing others' emotions. This is relevant because cognitive empathy conceptually overlaps with affective ToM [48], making empathy studies an important reference point.
Derntl et al. (2013) extended their previous work by testing emotion recognition, as well as emotional perspective-taking, and affective responsiveness [48]. They measured estrogen (specifically estradiol) and progesterone in order to relate hormone levels to task performance. Based on the model of Decety and Jackson, these three components are considered to cover the concept of empathy [48][49]. Derntl's work describes empathy as having cognitive and affective aspects. Emotional perspective-taking is considered a measure of cognitive empathy. They explain that cognitive empathy refers to the ability to understand others' feelings. It is strongly connected to ToM, and especially affective ToM is highly comparable to cognitive empathy. Affective responsiveness reflects affective empathy, which is the ability to feel others'

inferred emotions. In their study, women in the follicular phase were better at recognizing emotions than those in their mid-luteal phase, which matched their earlier findings in Derntl et al. (2008) [44][48]. In addition, women in the luteal phase responded more quickly, but not more accurately, to negative situations. For affective responsiveness, this was taken as evidence of greater reactivity to negative situations. No significant differences were found in emotional perspective-taking or in self-reported empathy. They conclude that emotion recognition and affective responsiveness both seem to be influenced by progesterone levels, but in different ways. Lower progesterone was linked to better emotion recognition accuracy, whereas higher progesterone was associated with greater affective responsiveness [48]. Overall, these findings suggest that menstrual cycle phases influence emotion recognition and affective empathy, while cognitive empathy appears unaffected. A study by Kimmig et al. (2021) notes that earlier evidence suggests that hormonal status may influence affective but not necessarily cognitive empathy in women. In this context, hormonal status refers to a woman's menstrual cycle phase or the use of oral contraceptives [50]. Their study examined the separate effects of emotional closeness and hormonal status, as well as their combined impact on female empathy. They compared three hormonal status groups, which were verified using measured ovarian hormone levels: women taking oral contraceptives as a form of hormonal birth control (low natural estrogen and progesterone), naturally cycling women during the early follicular phase (low estradiol and progesterone), and naturally cycling women around ovulation (high estradiol and low progesterone). The results showed that overall empathy was higher towards friends than towards people they disliked. Affective empathy was more strongly shaped by closeness than cognitive empathy. Hormonal status only influenced affective empathy towards disliked people. Oral contraceptive users showed reduced affective empathy compared to women who were not using oral contraceptives and were in the early follicular phase. No differences emerged for the naturally cycling individuals around ovulation. This study concludes that there was no evidence that hormonal status influenced cognitive empathy, while affective empathy seemed more sensitive to both interpersonal closeness and hormonal group [50].

The discussed studies present mixed results on whether and how hormones involved in the menstrual cycle influence social cognition. Emotion recognition appears more accurate in the follicular phase in several studies, though others report no phase differences or even an advantage during the luteal phase. Empathy builds on emotion recognition by involving the understanding and sharing of others' emotions. Findings of Derntl et al. (2013) and Kimmig et al. (2021) suggest that affective empathy is influenced by hormonal variation [48][50]. Cognitive empathy, and with it, affective ToM, seems to remain stable across the menstrual cycle. According to Mier et al. (2010), emotion recognition and affective ToM have overlapping brain activation. This means that they are not fully separate processes, but suggests that affective ToM builds on emotion recognition [51]. Therefore, examining whether ovarian hormonal changes influence affective ToM could be an interesting next step, especially since it extends beyond emotion recognition and is essential for everyday social interactions. Affective ToM abilities can be measured straightforwardly using performance tasks that require identifying others' mental states. For these reasons, this thesis examines whether the menstrual cycle influences affective ToM abilities specifically.

3.4 Methodological Inspirations

This study uses STAI-S to assess daily general anxiety (Section 2.2). The STAI has previously been applied in studies on anxiety across the menstrual cycle [30]. For instance, Mentese and Kutlu (2023) aimed to investigate how different phases of the menstrual cycle (early follicular, late follicular, and luteal) affect cognitive and motor skills in women. STAI was completed once per phase to measure anxiety symptoms [52]. Flores-Ramos et al. (2017) used the STAI to measure state and trait anxiety in peri- and post-menopausal women. The results showed that peri-menopausal women reported significantly higher state anxiety than post-menopausal

women [53]. Building on these approaches, this thesis uses the STAI-S daily across two full menstrual cycles. This design captures day-to-day fluctuations in anxiety, covering all phases of the cycle. However, a common problem in daily self-report studies is the occurrence of missing data points. Last Observation Carried Forward (LOCF) is an often-used strategy for accounting for this. Missing entries get substituted by the most recent available value. For example, Born et al. (2009) used it in long-term mood charting of bipolar patients [54]. Nordholt et al. (2024) applied LOCF in daily psychotherapy monitoring of patients with major depressive disorder [55]. This approach will also be used in this thesis to ensure data continuity and minimize gaps.

The Adult FPRT, used for measuring affective ToM in this study (Section 2.3), is a well-established and publicly available test for research purposes [34]. It involves reading short stories to assess recognition of faux pas and understanding of the characters' feelings. Dodell-Feder et al. (2013) introduced the Short Story Task (SST), demonstrating that a narrative-based measure can be used to assess affective and cognitive ToM performance [56]. This motivates the application of the Adult FPRT, since it is also a narrative-based ToM task that uses short stories. The Adult FPRT is suitable for examining differences in social reasoning. It has been used in several studies to identify ToM variations in individuals with social cognitive impairments. Gregory et al. (2002) used it to show ToM deficits in frontotemporal dementia [57]. More recently, Đorđević et al. (2025) validated a Serbian version of the Adult FPRT. It demonstrated reliable measurement qualities and confirmed its capacity to distinguish between healthy participants and individuals with schizophrenia or bipolar disorder [58]. These findings further support its value as a measure of ToM across diverse populations or conditions.

4 Experimental Setup and Data Collection

The methodology of this study builds on the approaches discussed in Section 3.4. The current section outlines the experimental setup and the data collection process. This includes participant recruitment and consent, data collection through daily surveys, the distribution of ToM tests, and the procedures used for data management.

4.1 Participant Recruitment and Ethics

Participants were recruited through posters and flyers distributed around university buildings and public spaces, as well as through personal networks and social media platforms. A website was created to provide additional information about the study and to allow interested individuals to register via email [59]. Eligible participants were individuals who menstruated, were at least 18 years old, and were fluent in English. Participants had not used hormonal contraception for at least two months prior to and during participation. After registration, participants received an information sheet describing the study and outlining the tasks they were expected to complete. They then signed an informed consent form sent to the email address they used to register.

The participants' ages ranged from 18 to over 35 years old, with the majority falling between the ages of 22 and 25. In total, 24 participants completed the daily anxiety surveys, and 17 completed all four ToM tests. Participation was voluntary, and participants could withdraw from the study at any time without providing a reason or facing any consequences. All data was collected anonymously and identified only by participant ID codes. The link between participant names and ID was kept confidential. It was only used to distribute ToM tests at the correct points in the menstrual cycle and to remove data from participants who chose to withdraw or had to be excluded from the study. Once the data collection and the one-month period during which participants could request data deletion ended, the link between names and IDs was permanently deleted. Furthermore, participants signed a debriefing

form in which they reconfirmed consent for the use of their data. This thesis received ethical approval from the Creative Intelligence & Technology ethics board at Leiden University.

4.2 Data Collection

Data was collected longitudinally over approximately 2.5 months, from July 2 to September 23, 2025. Each participant with a regular cycle completed two full menstrual cycles, beginning on whichever cycle day they were on at the time of joining the study. This means that data collection continued until the same cycle day was reached twice after the starting point, ensuring coverage of two complete cycles. For instance, if a participant started on day 10 of their first recorded cycle, data collection continued through the remainder of that cycle, the entirety of the next cycle, and until day 10 of the third cycle. Individuals with irregular cycles were only asked to participate for one full menstrual cycle. In this study, irregular cycles were defined as cycles shorter than 21 days or longer than 38 days. Furthermore, cycles were considered irregular when an individual's cycle length varied by more than nine days between cycles [19].

Qualtrics was used for the data collection. This enabled secure data storage and automated survey distribution. This included all surveys, tests, informed consent forms, and debriefing forms. Participant progress and cycles were tracked and updated daily in the participant-tracking Excel file to ensure precise cycle day tracking, accurate scheduling of ToM tests, and monitoring of survey completion. Another separate Excel file was created to record missed survey entries.

4.2.1 Preliminary Survey

Before the start of the daily surveys, each participant completed a preliminary survey. This survey gathered general information relevant to the study, including whether participants considered themselves generally anxious ("definitely yes", "maybe a little", or "not at all"), and whether they had been diagnosed with a menstrual disorder, anxiety disorder, or any other condition that could affect anxiety. Participants also reported the start and end dates of their previous menstruation. The reported period start date was used to determine the participant's initial cycle day at the beginning of the data collection.

4.2.2 Daily Surveys

After the preliminary survey was submitted, participants were added to the Qualtrics daily survey list. Qualtrics automatically distributed a daily survey to all active participants. Participants had to rate twenty short STAI-S statements describing how they felt at that moment on a scale from 1 ("not at all") to 4 ("very much so"). The list of statements is provided in Appendix A.1. Based on these statements, the STAI-S score was automatically calculated in Qualtrics for every participant's survey entry. The total score was derived by summing the responses to all twenty items, after reverse-scoring the ten positively worded items. The resulting total score ranged from 20 to 80. Higher scores indicated greater state anxiety.
Additional information was collected to gain further insights by asking about menstrual status (currently menstruating; if yes, start date), sleep quality (1-5; 1 = very poor, 5 = very good), and medication use that affects anxiety levels (yes/no). Furthermore, participants were asked whether a situational stressor occurred during the day, such as an exam, presentation, or family issue ("yes, definitely", "possibly a little bit", or "no, not at all"). Daily stressors were assessed using this survey question. For the analysis, the three response options were categorized as no stressor, slight stressor, and high stressor.

Each daily survey was sent out at 8:00 p.m. (CEST) and remained available for 48 hours to ensure responses reflected recent experiences. As a result, survey windows overlapped

across consecutive days. Automated reminders were sent at 8:00 a.m. and 4:00 p.m. (CEST) on the following day to participants who had not yet responded. On the second day, a more detailed manual reminder was scheduled to be sent at 2:00 p.m. (CEST), explicitly noting that six hours remained to complete the survey for the specifically missed date. An additional automated reminder was sent at 6:00 p.m. (CEST) on the second day, notifying participants that they had two hours left to complete the missed survey.

4.2.3 Theory of Mind Tests

Participants with regular menstrual cycles were asked to complete four affective ToM tests based on the Adult FPRT. The test measures participants' ability to recognize social missteps and understand the emotions and intentions of others. The original Adult FPRT consists of twenty short stories. Appendix B.1 contains the instructions given to participants, as well as an example of a faux pas story with its corresponding questions. For this study, the test was divided into two subsets of ten stories presented in random order, each containing five faux pas stories and five control stories. These two subsets were then duplicated and slightly modified to create two additional versions with the same balance of faux pas and control stories. However, the story order was randomized again, and the stories were slightly altered. This was done, for instance, by changing names, genders, or settings to reduce repetition effects while maintaining comparability across test results within a participant. Each participant completed one subset during days 7-10 of their cycle and one during days 21-24. This was done over two consecutive cycles, meaning participants completed four ToM tests in total. All participants started with Subset 1 and ended with Subset 4. Depending on their starting cycle day, the subsets could be completed in different phase orders. The story composition and numbering of all subsets can be found in Appendix B.2, and the phase orders are listed in Table B1 (Appendix B.3).
Days 7-10 were chosen to represent the mid-follicular phase, when estrogen levels are supposed to rise. Days 21-24 typically fall within the mid-to-late luteal phase, when both progesterone and estrogen should begin to decline. These phases were selected because they differ in hormone levels. Despite mixed findings, previous research has shown that ovarian hormonal fluctuations may affect emotion recognition and empathy (processes related to affective ToM) [11][45][48][50]. The hormones involved have also been linked to fluctuations in anxiety and mood regulation [8][9]. Since several studies have suggested that anxiety influences ToM [10][42], comparing these two phases allows examination of ToM performance under different ovarian hormonal and possibly emotional conditions.

Cycle days were checked daily in the participant-tracking Excel file to determine which participant ID should receive a ToM test. If a participant was expected to enter their testing window (days 7-10 or days 21-24) the next day, the correct test was scheduled in Qualtrics to be sent before 10:00 a.m. (CEST). Following this, reminder emails were set up to be sent daily until the participant completed the test or until their testing window ended. Participants who did not complete their ToM test within the testing window did not receive further ToM tests and were excluded from this part of the research. Furthermore, participants with irregular cycles were excluded from the ToM analysis, as their testing days could not be reliably determined in relation to their menstrual phase.

4.2.4 Daily Tracking and Data Management

Every day throughout the data collection period, participant progress, period start dates, and ToM test completion were manually monitored and updated in a structured Excel file. The participant-tracking Excel file contained automated formulas to calculate each participant's current cycle day, cycle number, and which ToM test needed to be sent. The newly reported period start dates from the daily surveys were logged in the Excel file each day to ensure accurate cycle tracking. This enabled monitoring participants' cycle days, identifying when they entered their ToM testing window, and scheduling the appropriate ToM test for the following

morning. It also enabled tracking of when participants had completed the required number of menstrual cycles. This was important during the final month of the data collection. Once a participant had completed two full menstrual cycles, they needed to be removed from the Qualtrics daily survey mailing list and sent a debriefing form. Considering participants had different cycle lengths and started the experiment on different cycle days, the total tracking duration varied across individuals. This is why it was necessary to check daily which participants had completed the experiment, ensuring that only unfinished participants continued to receive surveys.

Since missing data entries are likely in a study requiring daily participation, specific exclusion rules were applied. Participants who did not complete the daily survey for more than three consecutive days were excluded from the study. In addition, participants who completed fewer than 75% of all daily surveys were also excluded from the analysis. These thresholds were chosen to ensure that conclusions about phase-related effects remained reliable while still allowing for occasional missing data entries. A separate Excel file was used to keep daily records of which participants missed a survey. Participants who were close to meeting the exclusion criteria were manually sent a personal reminder email as a final warning. Those who met the exclusion criteria were excluded immediately and removed from the mailing list.

5 Data Preprocessing and Analysis

Following the approaches outlined in Section 3.4 and the data collection process in Section 4, this section explains how the data was processed and analyzed. It first describes the steps taken to clean and structure the datasets, and concludes with an overview of the data analysis used to answer the research questions and test the hypotheses.

5.1 Data preprocessing

The raw data files containing all survey responses required preprocessing before they could be interpreted and used for statistical analysis. The following subsection describes the steps taken to clean, structure, and prepare the data. These steps involved handling missing survey responses, categorizing participants' daily data by menstrual cycle phase, and computing average anxiety (STAI-S) scores for each phase per participant. Finally, the ToM test results were scored, and the participants' mean scores per phase were calculated.

5.1.1 Preparing Daily Survey Data

The daily survey data was exported from Qualtrics and preprocessed in Python using the pandas library. This dataset contained daily entries for each participant, including their responses to the STAI-S, the STAI-S anxiety score, whether they were currently menstruating, and information on contextual factors such as stress, sleep quality, and medication use. Each row represented one day of participation per individual and was linked to a unique participant ID. The dataset, therefore, captured daily reports across one or two menstrual cycles per participant, depending on whether they had a regular or irregular cycle. The data was divided into two datasets: one containing participants with regular cycles and one with irregular cycles. Both datasets underwent the same preprocessing steps.

The first step was removing the data of participants who dropped out or had to be excluded according to the previously mentioned exclusion rules (Section 4.2.4). Throughout the data collection process, missing survey days were manually identified and logged in a separate Excel file that matched the structure of the exported full dataset. Each missing entry was added to this full dataset. These entries were marked with a flag indicating that they were inserted placeholders with the correct survey date and participant ID, but without any recorded data. This way, even if a participant missed a day, all survey days were represented

in the dataset. This ensured that each participant's data remained continuous.

To combine participants' daily survey data with their menstrual cycle information, a second Excel file was created containing one row per participant. This file included each participant's experiment start date, the cycle day on which they began the study, and their average cycle length. Before merging the two files, all dates were standardized by separating them into numerical day and month values to avoid inconsistencies in date formats. The merge was then performed by matching the participant ID and the experiment start date from the cycle information file to the first recorded survey date for that participant in the full dataset.

After merging the files, the surveys were treated as if they had been completed in chronological order. However, participants may have occasionally submitted surveys from the past 48 hours in a different order. It was not possible to know the exact order in which these surveys were meant to be completed. Therefore, the assumed intended order was kept. Since each row represented a new survey day, the survey dates were reconstructed by taking each participant's experiment start date as the reference point and counting each subsequent entry as the next day.

Once the assumed survey dates were assigned to each survey entry, the data was linked to each participant's menstrual cycle. This alignment was based on the cycle day on which they began the experiment to account for the fact that participants did not all start the daily surveys on the same cycle day. Additionally, it was based on participants' newly reported period dates throughout the study. In each daily survey, participants indicated whether they were currently menstruating and, if so, entered the starting date of that period. These reported menstrual start dates were matched to the corresponding survey dates to identify the first day of menstruation for each new cycle. Each of these entries was then reset to cycle day 1, and the cycle number was incremented by 1. Cycle days continued to be counted for each survey until a new period start was reported.

Missing values for survey entries that were added as placeholders were imputed using the LOCF method. This method replaces a missing entry with the most recent available data from the same participant. The survey dates and cycle days continued to count as usual, while all other information was carried forward. Across the dataset of participants with regular cycles, 24 out of 1,149 daily survey entries were missing. The lowest percentage of completed daily surveys among participants was 89%, and only six participants had one or more missing survey entries.

The next preprocessing step was labeling every survey entry according to its menstrual cycle phase. The menstrual phase was identified based on the reported start and duration of menstruation. Ovulation was estimated to occur 14 days before the next reported period. If no next period was recorded within the data collection period, ovulation was estimated to occur halfway through the participant's average cycle length (calculated from their recorded cycles). The luteal phase covered the days after ovulation until the next period, while the follicular phase included the days from the first day of menstruation to ovulation.

At the end of the preprocessing, any extra data entries were removed so that each participant contributed the same amount of data. This resulted in exactly two menstrual cycles for those with regular cycles, or one full cycle for those with irregular cycles. Furthermore, unnecessary columns were removed from the dataset, keeping only the relevant data for analysis.

To prepare the data for analysis, the average anxiety score was calculated for each participant within every menstrual cycle phase. Separate mean scores were computed for the menstrual, follicular including menstrual, follicular excluding menstrual, and luteal phases. For each participant, information from the preliminary survey regarding potential diagnoses and whether

they consider themselves an anxious person was added as separate columns. Means are sensitive to extreme values. However, they were used here because each phase included multiple daily observations per participant, which reduced the influence of single outliers. Simply using the sum of STAI-S scores per phase would be biased because menstrual cycle phases vary in length both within and between participants. Longer phases would naturally produce higher sums regardless of anxiety levels. Therefore, mean scores were used to allow fairer comparisons between phases. Furthermore, the mean was preferred over the median because the goal was to capture overall anxiety levels per phase. This is better represented by the mean when multiple observations are available.

5.1.2 Preparing Theory of Mind Test Data

ToM test answers were exported from Qualtrics and manually scored in an Excel file, with separate sheets per test. All four affective ToM tests were based on the Adult FPRT. Therefore, each ToM test was scored according to the Adult FPRT scoring system [34]. Before scoring, an answer sheet was created for all tests, reflecting what each question was intended to measure. For each participant, answers were marked as correct (1) or incorrect (0) for each question per story. Every story contained eight questions (Q1-Q8). The last two questions (Q7-Q8) assessed story comprehension. If one of the comprehension questions was answered incorrectly, all other answers for that story (Q1-Q6) were not allowed to be scored [34].
The remaining six questions (Q1-Q6) were used to calculate five sub-scores, each expressed as a ratio ranging from 0 to 1:

1. **Faux Pas Detection** (Q1, Q2): Participants need to recognize whether a faux pas occurred. Was something socially inappropriate said, and who said it?
2. **Understanding Inappropriateness** (Q3): Participants need to understand the social or emotional impact of what was said. Why was the remark inappropriate?
3. **Intentions** (Q4): Participants need to infer the character's motivation. Why did this person make that remark?
4. **Belief** (Q5): Participants need to be aware that the character had no bad intentions. Was the faux pas unintentional?
5. **Empathy** (Q6): Participants need to identify the appropriate emotional reaction. How did the person affected by the faux pas feel?

Participants who answered "no" to the faux pas detection question (Q1) were only asked questions 1, 7, and 8. If their "no" response to Q1 was correct (meaning no faux pas was present), the skipped questions were counted as correct. If their "no" response was incorrect (meaning a faux pas was present), the skipped questions were counted as incorrect.

For the Faux Pas Detection sub-score, the number of correct answers to Q1 and Q2 across all stories for one ToM test was divided by the total number of correctly answered comprehension questions (Q7-Q8) across all stories. For the other four sub-scores (Q3-Q6), each ratio was calculated by dividing the number of correct answers to that question across all stories of a single test by the number of stories in which *both* comprehension questions were answered correctly [34]. The total ToM score per test was then computed as the average of the five sub-scores. All sub-scores and total scores for each participant across the four ToM tests were calculated in Excel.

Another Excel file was created manually, linking all ToM scores and ratios to the menstrual cycle phase during which each test was completed. This was based on the phase overview shown in Table B1 (Appendix B.3). For each participant, the scores from all tests taken during the same phase were grouped together. The mean total ToM scores, Empathy ratios, and Understanding Inappropriateness ratios per participant were then calculated separately for the follicular and luteal phases. The resulting dataset contained the average of these scores and ratios for each participant per phase. This dataset was used for paired samples t-tests to determine statistical significance.

5.2 Data Analysis

This subsection explains how the collected data was analyzed to answer the research questions. The data analysis was divided into two parts. The first focused on anxiety levels across menstrual cycle phases. The second compared affective ToM performance between the mid-follicular and mid-to-late luteal phases. For both parts of the study, an overview of the statistical tests used to evaluate phase-related differences is provided. All statistical analyses were conducted using JASP (version 0.95.3.0) [60], and all visualizations were produced in Python using pandas for data handling and matplotlib for plotting.

5.2.1 Analysis of Anxiety Levels

During the categorization of participants' survey entries into menstrual cycle phases, all phase assignments were based on predictions. As ovulation could only be estimated and only represents a single day, it was not included as a separate phase in the analysis. Therefore, the analysis focused on the menstrual, follicular, and luteal phases. The follicular phase was analyzed both including and excluding menstruation. This distinction was made to examine whether anxiety levels during the follicular phase differ when the menstrual days are included compared to when they are excluded. Furthermore, the statistical analysis included only participants with regular menstrual cycles. Irregular or unusually long cycles can be anovulatory (no ovulation occurs) and are therefore not hormonally representative of a typical menstrual cycle.

To test whether mean anxiety levels differed across these phases, a repeated measures ANOVA test was conducted using participants' average STAI-S scores per phase. Before performing the repeated-measures ANOVA, several assumptions must be verified [61]:

- **Independence**: This assumes that each participant's measurements are unrelated to the other participants' measurements.
- **Normality**: This assumes that participants' mean STAI-S scores are approximately normally distributed within each menstrual cycle phase. Normality was verified using the Shapiro–Wilk test (for every phase $p > 0.05$). Q–Q plots showed that the data points closely followed the diagonal line, which further supports this assumption.
- **Sphericity**: This assumes that the variability in how participants' STAI-S scores differ between any two menstrual cycle phases is approximately the same. Mauchly's test indicated that this assumption was violated ($p < 0.05$, estimated Greenhouse–Geisser $\epsilon = 0.531$). ϵ is a value between 0 and 1 that describes the severity of the sphericity violation. Since there was a moderate violation ($\epsilon < 0.75$), the Greenhouse–Geisser correction was applied to obtain a more accurate p-value for the repeated measures ANOVA.

Statistical significance for the repeated measures ANOVA was set at $\alpha = 0.05$. Pairwise post-hoc comparisons were planned in case the repeated measures ANOVA revealed a significant effect. This was to determine which phases differed significantly from one another. A Holm correction was applied to control for the increased risk of a false positive (incorrectly rejecting the null hypothesis) occurring when multiple tests were performed.

5.2.2 Analysis of Affective Theory of Mind Performance

To examine whether affective ToM performance differs between the mid-follicular and mid-to-late luteal phases of the menstrual cycle, participants' mean scores per phase were analyzed. The Adult FPRT yielded five different scores and a total score as discussed in Section 5.1.2. For each participant, three outcome measures were considered:

- **Total ToM score**: This shows overall performance in recognizing and understanding social situations. It reflects general ToM, which includes both thinking (cognitive) and emotion (affective) aspects.

- **Empathy ratio**: This measures the accuracy with which participants identified how the affected person felt. It measures affective ToM by focusing on recognizing how someone feels.
- **Understanding Inappropriateness ratio**: This shows how well participants understood why a remark was inappropriate, which can include recognizing its emotional impact on others. It shows whether someone understands why a remark could hurt or embarrass another person. This relates to affective ToM.

Before testing the significance of phase-related differences in ToM performance, the distribution of the total ToM score differences between the mid-follicular and mid-to-late luteal phases was assessed using the Shapiro–Wilk test in JASP. It confirmed that the data did not deviate from normality ($p > 0.05$). The same normality assessment was conducted for the Empathy and Understanding Inappropriateness ratios. In both cases, the Shapiro–Wilk test also indicated no deviation from normality. Therefore, paired samples t-tests were conducted. For this test, statistical significance was set at ($\alpha = 0.05$), where a p-value below this threshold indicates a significant phase-related difference. The paired samples t-test assumes the data is normally distributed without any extreme outliers, that the two measurements are paired (the same participants tested in both phases), and that these pairs are independent of one another (participants' results do not influence each other) [62].

6 Results

This section presents the results of the study on daily self-reported state anxiety levels and affective ToM performance across menstrual cycle phases. The analyses are structured into two parts. The first part focuses on daily state anxiety levels. This addresses phase-related differences and the influence of external factors such as situational stressors and sleep quality. The second part reports ToM performance across two menstrual phases and examines whether any observed differences are related to test order rather than hormonal phase.

6.1 Anxiety Levels Results

A total of 20 participants ($n = 20$) with regular menstrual cycles were included in the analysis of state anxiety levels. Of the 20 participants, four reported a diagnosed anxiety disorder, and four reported another condition that may have affected their anxiety levels. No participant reported having a menstrual disorder. In this study, a menstrual disorder was defined according to the preliminary survey question, which asked participants whether they had a diagnosed menstrual disorder. Examples of such disorders include PMDD and endometriosis.

Daily state anxiety was assessed using the STAI-S, which produces scores between 20 (low anxiety) and 80 (high anxiety). The mean STAI-S scores per participant across menstrual cycle phases are shown in Table A1 in Appendix A.2. Most participants followed the same overall trend of highest anxiety during menstruation and lowest anxiety during the follicular phase excluding menstruation. However, some varied from this pattern. Table 1 summarizes the descriptive statistics for each menstrual cycle phase. It presents the mean anxiety scores for each menstrual cycle phase. The standard deviation (SD) indicates the extent to which individual scores vary around the mean within each phase. The standard error of the mean (SE) represents the accuracy of the sample mean in estimating the population mean. As can be seen in Table 1, mean STAI-S scores appeared to fluctuate across menstrual phases. The highest anxiety was observed during the menstrual phase ($M_{\text{STAI-S}} = 45.32$) and the lowest during the follicular phase excluding menstruation ($M_{\text{STAI-S}} = 38.05$). The follicular phase including menstruation ($M_{\text{STAI-S}} = 40.91$) and the luteal phase ($M_{\text{STAI-S}} = 43.02$), showed values between the menstrual phase and the follicular phase excluding menstruation.

The repeated measures ANOVA test (with Greenhouse–Geisser correction) revealed a significant main effect of menstrual cycle phase on STAI-S scores, $F(1.59, 30.27) = 7.02$, $p = 0.005$.

Menstrual Phase	Mean score	SD	SE
Menstrual	45.32	12.083	2.702
Follicular (incl. menstruation)	40.91	8.645	1.933
Follicular (excl. menstruation)	38.05	8.471	1.894
Luteal	43.02	10.610	2.373

Table 1: Descriptive statistics for mean state anxiety (STAI-S) scores across menstrual cycle phases for $n = 20$. The table displays the mean, standard deviation (SD), and standard error of the mean (SE) for participants' average STAI-S scores in each phase.

This indicates that mean anxiety levels differed significantly across menstrual cycle phases. The F-value represents the test result from the ANOVA, and the p-value shows whether the effect is statistically significant. Holm-adjusted pairwise post-hoc comparisons revealed that anxiety levels during the menstrual phase were significantly higher than during both the follicular phase including menstruation ($p = 0.030$, $d = 0.44$) and the follicular phase excluding menstruation ($p = 0.033$, $d = 0.72$). The follicular phase including menstruation also showed slightly higher anxiety than the follicular phase excluding menstruation ($p = 0.046$, $d = 0.28$). No significant differences were found between the luteal phase and any other phase (menstrual-luteal: $p = 0.241$, $d = 0.23$; follicular incl.-luteal: $p = 0.241$, $d = -0.21$; follicular excl.-luteal: $p = 0.057$, $d = -0.49$). The d represents the effect size, which reflects how large the difference is. When compared to the menstrual phase, anxiety during the follicular phase showed a small-to-medium effect when menstrual days were included and a medium-to-large effect when they were excluded (based on Cohen's d interpretation) [63].

Figure 3 shows the phase-related differences. The error bars represent within-subject 95% confidence intervals calculated using the Cousineau–Morey normalization. This normalization accounts for individual baseline differences, ensuring that the error bars reflect within-person variability instead of variability between participants.

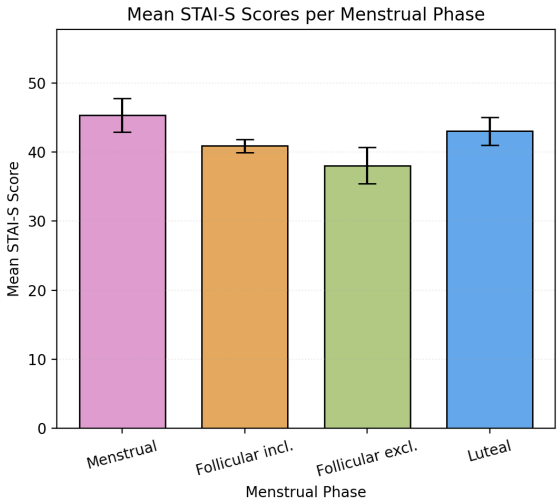

Fig. 3: Mean state anxiety (STAI-S) scores across four menstrual phases for all participants with a regular cycle ($n = 20$): menstrual (pink, $M_{\text{STAI-S}} = 45.32$), follicular including menstrual days (orange, $M_{\text{STAI-S}} = 40.91$), follicular excluding menstrual days (green, $M_{\text{STAI-S}} = 38.05$), and luteal (blue, $M_{\text{STAI-S}} = 43.02$). Error bars represent within-subject 95% confidence intervals (Cousineau–Morey). Higher STAI-S scores indicate higher anxiety levels. The figure shows that anxiety was highest during the menstrual phase and lowest during the follicular phase excluding menstruation.

Additional analyses explored whether the observed phase-related differences could have been influenced by other factors. Figure 4 shows that participants with a diagnosed anxiety disorder

or another condition affecting anxiety ($n = 8$) reported slightly lower or comparable anxiety levels across all phases compared to participants without such a diagnosis ($n = 12$). Figure 5 displays mean STAI-S scores grouped by participants' self-rated general anxiety in the preliminary survey. Those identifying as anxious ($n = 8$) and slightly anxious ($n = 11$) showed comparable anxiety levels. The one participant identifying as not anxious reported relatively high anxiety scores overall. The overall trend appeared similar across both figures (Figures 4 and 5), with the highest anxiety levels during menstruation and the lowest during the follicular phase excluding menstruation, regardless of diagnosis or self-rated general anxiety.

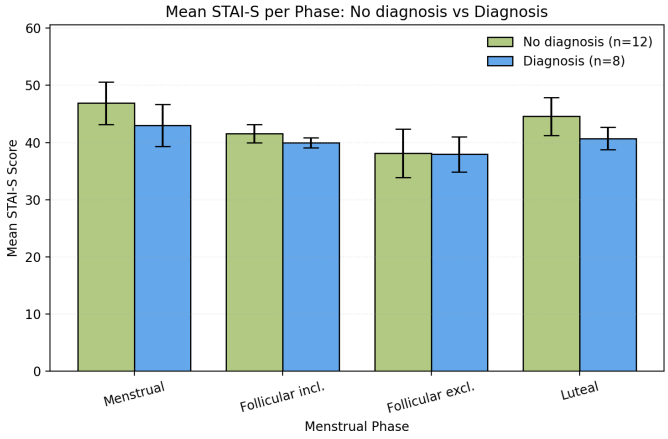

Fig. 4: Mean state anxiety (STAI-S) scores across the four menstrual phases for participants with regular cycles without (green) and with (blue) a diagnosis influencing anxiety ($n = 12$ and $n = 8$, respectively). Error bars represent within-subject 95% confidence intervals (Cousineau–Morey). Higher STAI-S scores reflect higher anxiety levels. In both groups, anxiety scores were generally highest during the menstrual phase and lowest during the follicular phase excluding menstruation. Participants with a diagnosis tended to show slightly lower or comparable anxiety levels to those without a diagnosis across phases.

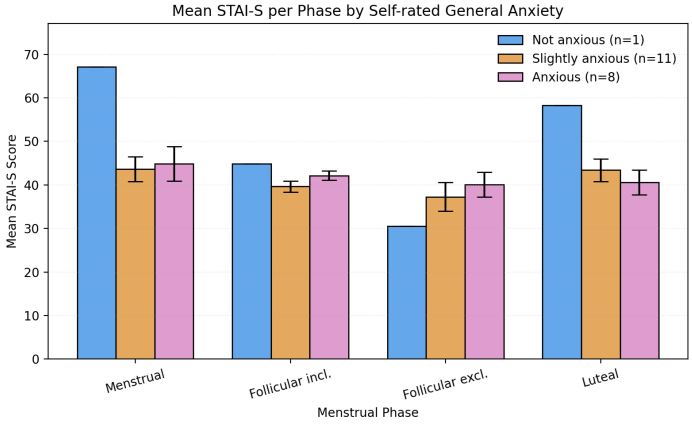

Fig. 5: Mean state anxiety (STAI-S) scores across the four menstrual phases for participants with regular cycles. The bar chart is grouped by participants' self-rated general anxiety: not anxious (blue, $n = 1$), slightly anxious (orange, $n = 11$), and anxious (pink, $n = 8$). Error bars represent within-subject 95% confidence intervals (Cousineau–Morey). Higher STAI-S scores indicate higher levels of state anxiety. State anxiety scores were highest during the menstrual phase and lowest during the follicular phase excluding menstrual days across all groups. The "not anxious" group consists of a single participant and should be interpreted with caution. The single participant reported considerably higher anxiety scores during most phases compared to the group means, despite identifying as non-anxious.

The potential influence of situational factors was assessed by examining the relationship between daily STAI-S scores and daily stressors (Figure 6) and between daily STAI-S scores and sleep quality (Figure 7).

Figure 6 shows that daily state anxiety (STAI-S) scores differed across levels of reported daily stressors. Median STAI-S scores increase with higher stressor ratings. The interquartile ranges (IQRs) show increasing overall score ranges from the no stressor to the high stressor categories. Participants tended to report higher STAI-S scores on days when they experienced a high stressor. As illustrated in Figure 7, daily state anxiety scores varied across reported levels of sleep quality. Lower sleep quality was associated with higher daily anxiety scores since the IQR shifted downward with higher sleep quality ratings. Median STAI-S scores were highest when participants rated their sleep as very poor and lowest when they rated their sleep as very good. For both Figures 6 and 7, the long whiskers and IQRs show that daily anxiety scores varied considerably within each stressor and sleep quality category.

Medication intake data was collected but excluded from visualization because only three participants reported taking medication affecting anxiety. One participant took medication daily, while the other two reported only one or two instances. This resulted in insufficient variation for meaningful analysis.

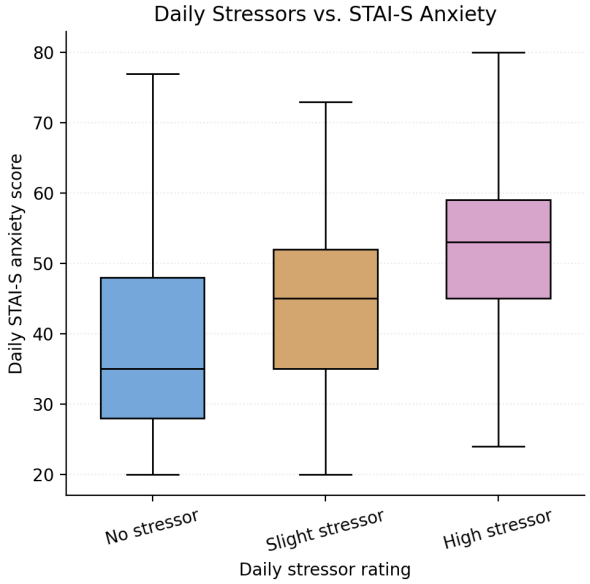

Fig. 6: Daily state anxiety (STAI-S) scores plotted against participants' reported daily situational stressors. During the daily surveys, participants ($n = 20$) indicated whether they had experienced an external stressor that day using three categories: no stressor (blue), slight stressor (orange), and high stressor (pink). Each box represents the distribution of daily STAI-S scores for that category. The line inside marks the median, and the box height shows the interquartile range (middle 50% of values). Whiskers represent the range of typical scores. Higher STAI-S scores indicate higher state anxiety. The highest anxiety scores were observed on days when participants experienced a high stressor.

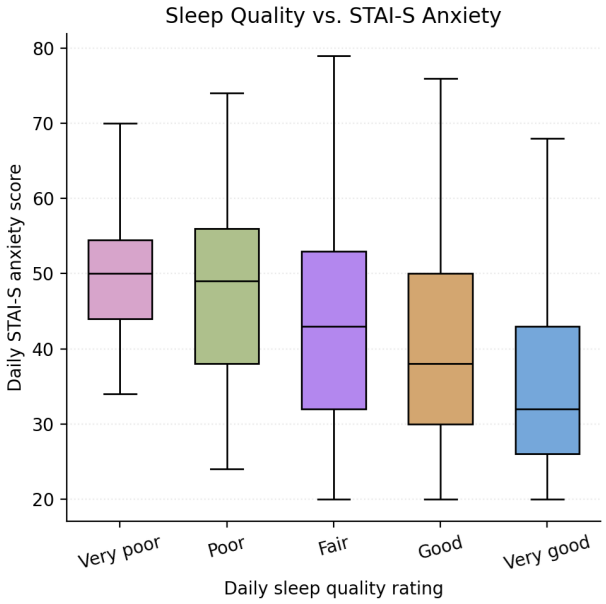

Fig. 7: Daily state anxiety (STAI-S) scores in relation to participants' reported sleep quality. Each day, participants ($n = 20$) rated their sleep on a scale from 1 (very poor) to 5 (very good). Each box shows the distribution of daily STAI-S scores for that sleep rating, with the line inside indicating the median and the box height representing the interquartile range (middle 50% of scores). Whiskers extend to the typical range of values. Higher STAI-S scores correspond to greater state anxiety. Lower sleep quality was associated with higher daily anxiety scores.

Participants with irregular cycles ($n = 4$) were analyzed separately. Their mean STAI-S scores per phase are presented in Table A2 in Appendix A.2. The data showed no clear pattern across menstrual phases. Two participants (P007, P012) had nearly constant anxiety levels across all phases. P015 reported overall higher anxiety in the luteal phase, and P021 showed slightly lower anxiety levels in the luteal phase.

6.2 Affective Theory of Mind Performance Results

Seventeen participants ($n = 17$) with regular cycles completed four ToM tests. All subscores ranged from 0 to 1, with higher values indicating better performance. Table B2 in Appendix B.4 displays the total ToM score per test for each participant. The mean total ToM score of all participants was 0.857 ($SD = 0.109$) during the mid-follicular phase and 0.859 ($SD = 0.087$) during the mid-to-late luteal phase. A paired samples t-test compared these mean ToM scores between the two phases. The test showed no significant difference between phases, $t(16) = -0.10$, $p = 0.918$, $d = -0.03$.
The mean Empathy ratio was 0.904 ($SD = 0.109$) during the mid-follicular phase and 0.894 ($SD = 0.076$) during the mid-to-late luteal phase. A paired samples t-test showed no significant difference between phases, $t(16) = 0.45$, $p = 0.662$, $d = 0.11$. The mean Understanding Inappropriateness ratio was 0.808 ($SD = 0.145$) during the mid-follicular phase and 0.843 ($SD = 0.112$) during the mid-to-late luteal phase. This difference was also not significant, $t(16) = -1.07$, $p = 0.303$, $d = -0.26$.

The mean ToM performance across the two menstrual phases for all seventeen participants is shown in Figure 8. The bars show the total ToM score, Empathy ratio, and Understanding Inappropriateness ratio with within-subject 95% confidence intervals (Cousineau–Morey normalization). According to these intervals, there was limited variability within participants for

each measure. Across both phases, the mean total ToM score and the mean ratios remained similar, with no significant phase-related differences.

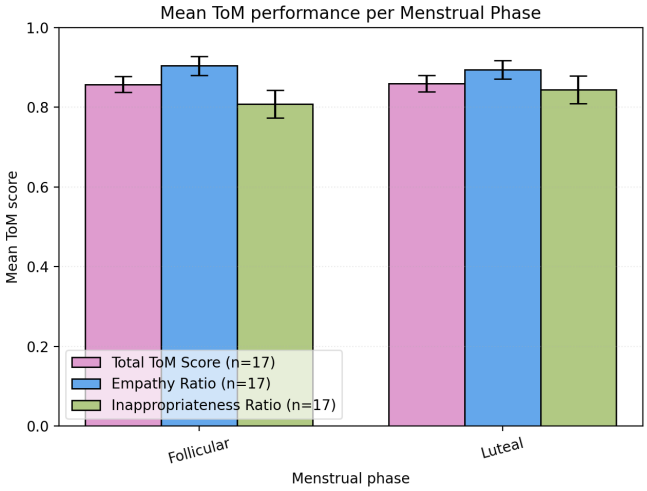

Fig. 8: Mean Theory of Mind (ToM) performance across the mid-follicular phase and mid-to-late luteal phase ($n = 17$). Three measures are shown: total ToM score (pink), Empathy ratio (blue), and Understanding Inappropriateness ratio (green). Error bars represent within-subject 95% confidence intervals (Cousineau–Morey). The higher the score, the better the performance. The total ToM score reflects overall ToM performance ($M_{\text{follicular}} = 0.857$, $M_{\text{luteal}} = 0.859$), The Empathy ratio ($M_{\text{follicular}} = 0.904$, $M_{\text{luteal}} = 0.894$), and the Understanding Inappropriateness ratio ($M_{\text{follicular}} = 0.808$, $M_{\text{luteal}} = 0.843$) involve affective ToM specifically. Mean scores and ratios appeared comparable across both phases.

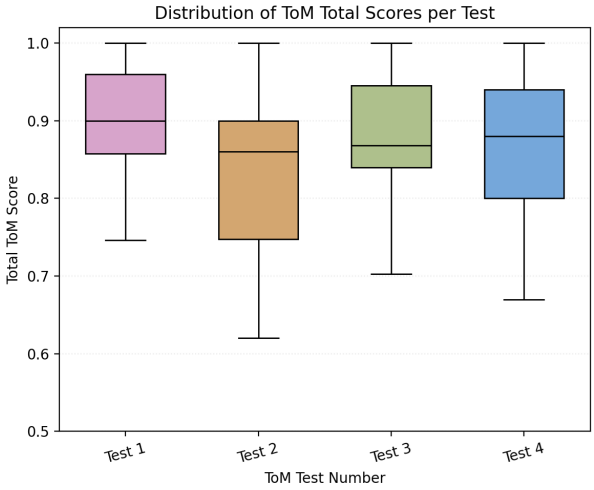

Fig. 9: Distribution of total Theory of Mind (ToM) scores across the four ToM tests ($n = 17$). Each box represents the score range for one test: Test 1 (pink), Test 2 (orange), Test 3 (green), and Test 4 (blue). The line inside each box marks the median, and the box height represents the interquartile range (middle 50% of values). Whiskers indicate the range of typical scores, excluding outliers. The plot visualizes variation in ToM scores by test order to check for potential learning or recognition effects across sessions. The overall test scores show slight variations with comparable medians.

Most participants mentioned recognizing some of the stories from the earlier tests, as Tests 3 and 4 contained modified versions of the stories from Tests 1 and 2. To examine whether

total ToM score differences or similarities could be related to test order rather than menstrual phase, Figure 9 visualizes the distribution of total ToM scores across the four test sessions. The median scores were generally comparable across tests, though Test 2 appeared to have slightly lower overall ToM scores compared to the other tests. The overlapping IQRs and whiskers show that score distributions were largely comparable across the four tests.

7 Discussion

Despite its impact on both physical and psychological well-being, the menstrual cycle remains under-researched [5][6]. Earlier studies suggest that hormonal shifts in estrogen and progesterone throughout the menstrual cycle may influence mood and anxiety [9][40]. However, most have used isolated rather than continuous measurements or have primarily examined clinical samples [38][40]. Less is known about how these hormonal and potential anxiety fluctuations relate to social cognition. Previous work has linked ToM to anxiety [10][42], and other research has connected processes related to affective ToM to hormonal variation across the menstrual cycle [11][48][50]. Still, findings across these studies are mixed, and no work appears to have directly examined affective ToM across menstrual cycle phases. Therefore, this study investigated whether the menstrual cycle affects daily state anxiety and whether affective ToM performance differs between the mid-follicular and mid-to-late luteal phases in naturally cycling participants.
This discussion section interprets the findings of both parts of the study and addresses its limitations and implications.

7.1 Reflection on Anxiety Across the Menstrual Cycle

The results show that state anxiety fluctuated across menstrual phases, with the highest levels reported during menstruation and the lowest during the follicular phases. The difference in effect sizes between the menstrual and follicular phases suggests that anxiety decreased more clearly once menstruation ended. When menstrual days were excluded, the contrast between the menstrual and follicular phases became stronger. This indicates that the follicular phase post-menstruation may represent a more stable period with low anxiety. In short, the findings suggest that anxiety fluctuates across menstrual cycle phases. This pattern likely reflects hormonal variations in estrogen and progesterone that influence emotional regulation as discussed in Section 2.1 [8]. During menstruation, both estrogen and progesterone levels are at their lowest, which potentially contributes to the heightened anxiety levels. As estrogen increases in the follicular phase, anxiety levels appear to decline. This could be explained by the influence of estrogen on mood regulation, as lower levels are often associated with more anxiety (and vice versa) [8].
The interpretation aligns with research linking low ovarian hormone levels to increased anxiety. Wieczorek et al. (2023) and Green and Graham (2022) observed that anxiety symptoms tend to worsen during both the menstrual and luteal phases, when estrogen and progesterone levels are lower [9][40]. This is mostly consistent with the pattern found here, as participants in the current study also reported higher anxiety during menstruation and slightly higher levels during the luteal phase. However, unlike Hantsoo et al. (2022), who reported premenstrual (late luteal phase) anxiety [38], the present findings did not show significant luteal-phase differences. This may be because ovarian hormone levels vary considerably across the luteal phase. Progesterone and estrogen rise after ovulation but drop before menstruation. These hormonal differences make it difficult to detect a consistent effect on anxiety throughout the entire phase. As a result, anxiety may not increase immediately after ovulation but rather when both hormones begin to decline or have declined. In this sample, the pattern suggests that anxiety peaks mainly when hormone levels are lowest, such as during menstruation, rather than earlier in the luteal phase. However, because no hormone levels were directly measured, this interpretation remains speculative and cannot be confirmed based on the collected data. Overall, the findings partly support the hypothesis that anxiety would be higher

during menstruation and the luteal phase, and lower when nearing ovulation.

When examining participant types in more detail, those with a diagnosed (anxiety) disorder reported slightly lower average anxiety levels than those without a diagnosis. This pattern was unexpected and is likely related to individual circumstances rather than a genuine group effect. One possible explanation is that the participant taking daily anxiety medication, who also had a diagnosed disorder, reported overall lower state anxiety scores. Since the sample was small, this single case could have influenced the group mean. In contrast, some participants without a diagnosis reported relatively higher anxiety levels, possibly due to unrecognized symptoms or temporary stress. Regardless of diagnostic background, both groups showed the same general trend in anxiety fluctuations. This suggests that diagnosis alone does not determine how the menstrual cycle affects anxiety, since external factors can also influence it.
Participants also rated themselves as generally anxious, slightly anxious, or not anxious. All three groups showed similar changes in anxiety across phases, suggesting that phase-related fluctuations appeared in each group. Interestingly, the one participant who described themselves as not anxious had some of the highest overall scores, which may reflect differences in how participants define or perceive anxiety in themselves.

Not all participants followed the same pattern of anxiety across the cycle. Some showed stable levels, while others had stronger fluctuations. This variation suggests that the menstrual cycle influences anxiety differently across individuals. Daily factors such as stress or sleep quality could play a role in these differences. Participants reported higher anxiety on days with stronger stressors and lower anxiety on days without an external stressor. Better sleep quality was associated with lower anxiety scores, whereas poorer sleep quality showed higher anxiety. This may help explain why some participants showed stronger fluctuations in anxiety while others remained more stable. It suggests that daily circumstances could influence how changes in anxiety across the menstrual cycle are experienced.

For participants with irregular cycles, no consistent pattern in anxiety was found across phases. These mixed results are difficult to interpret due to the small sample size and uncertainty in phase classification. Potential unusual ovarian hormone fluctuations could explain the inconsistent pattern, as some cycles may have been anovulatory. When ovulation does not occur or when cycle lengths vary significantly, hormone levels do not follow the typical pattern. This could make phase-related effects on anxiety less distinct.

7.2 Reflection on Affective Theory of Mind Performance

Affective ToM performance remained stable when comparing the mid-follicular phase to the mid-to-late luteal phase. The mean total ToM scores during the two phases were almost identical. The Empathy and Understanding Inappropriateness ratios, which reflect affective ToM more specifically, also showed similar means across phases. The stability may suggest that affective ToM is not strongly influenced by the menstrual cycle. However, it could also mean that the effects of the cycle on ToM are too subtle to detect. The findings contradict the hypothesis that higher performance would occur during the mid-follicular phase due to lower anxiety levels.
Previous studies have found links between anxiety and reduced (affective) ToM performance [10][42]. Although anxiety levels varied across phases in this study, it did not seem to influence affective ToM performance. This could be explained by the lack of a significant difference in anxiety between the follicular and luteal phases. However, it is important to note that this study was not designed to determine whether anxiety itself affects ToM. Anxiety was only considered as a potential factor that might influence ToM performance during different menstrual phases.
Earlier research on the effect of ovarian hormones on emotion recognition and empathy has shown mixed results. Some studies found better emotion recognition during certain menstrual

cycle phases [11][44][45], while others reported no differences [46][47]. The current findings align with the latter, suggesting stable performance across phases. The differences compared to studies that did find phase effects may be due to task type. Emotion recognition tasks focus on identifying facial expressions, while ToM tasks involve understanding emotional intentions and social context. Ovarian hormonal changes might influence simple emotional perception more than the reasoning needed to interpret others' thoughts and feelings. Looking at studies that focus on these more complex processes, Derntl et al. (2013) and Kimmig et al. (2021) found that cognitive empathy (which closely resembles affective ToM) remained stable across phases [48][50]. This aligns with the current findings, as mean total ToM scores, mean Empathy ratios, and mean Understanding Inappropriateness ratios remained similar across phases.

To assess whether methodological factors could have influenced these results, the order of the ToM tests was also taken into consideration. The total ToM scores differed only slightly between tests, with no consistent pattern across sessions. This suggests that learning or story familiarity did not affect performance. Test 2 seemed slightly more difficult, with scores spread over a wider range than in the other tests. However, the median remained at a similar level across sessions. The results also show that ToM abilities varied between participants but not systematically across tests. This supports that the stable ToM performance was likely not due to the test order.

7.3 Study Limitations

While the study provides insight into how menstrual cycle phases may relate to anxiety and affective ToM, several limitations should be acknowledged when interpreting the results. The main limitation is the relatively small sample sizes in both parts of the study ($n = 20$ for anxiety (regular cycle) and $n = 17$ for ToM), which limits how well the results can be applied to a larger population. With few participants, it is more challenging to detect small effects due to limited statistical power, and outliers can have a stronger impact on the results.

7.3.1 Anxiety Study Limitations

In addition to the limited sample size, group sizes were uneven in comparisons, such as those between participants with and without a diagnosis affecting anxiety levels. This reduces the reliability of those results. It is still possible that people with a disorder experience stronger or weaker anxiety fluctuations across the cycle, even though this study did not show clear differences between these groups. Beyond these group differences, several methodological limitations might have influenced data accuracy and menstrual cycle phase classification.

All the menstrual cycle phases were predicted based on the reported period dates. These estimations were never confirmed by measuring hormone levels. Furthermore, not all participants started the study on day 1 of a cycle. In some cases, the data covered two full cycles, each beginning and ending mid-cycle. As a result, phase classification was not always precise for every participant. When no next period was recorded to estimate the previous ovulation, the average cycle length was used instead, which may have reduced accuracy. Some phases also contained more data points because they lasted longer. This means that certain phases were represented more heavily in the dataset. If one participant had a short menstrual phase (e.g., 2 days) and another had a longer one (e.g., 7 days), a single outlier would have a stronger impact on the participant's average STAI-S score for the shorter phase. This could also have influenced the overall phase means. These limitations make the phase categorization less accurate and may have caused some individual phase means to influence the results more than others.

Because phase estimation relied on reported menstruation dates, the hormonal patterns underlying the phases could not be verified. This may also explain why certain expected effects

did not appear in the results. No significant luteal phase effects on anxiety were found, whereas several other studies did report such effects [9][38][40]. A possible explanation could be that in this study, the luteal phase was analyzed as one category, even though ovarian hormone levels change within this phase. This may have masked effects that occur when the hormone levels drop right before menstruation. Unlike studies such as Reynolds et al. (2018), which directly measured progesterone, the daily anxiety data relied on predicted phases. Reynolds et al. found no within-person differences in anxiety when progesterone varied [36], which could also explain the absence of luteal effects. It may be the case that progesterone might not strongly influence anxiety within individuals. Low progesterone was hypothesized to lead to higher state anxiety. According to the results, anxiety was lowest during the follicular phase. During this phase, progesterone levels are low. This further supports the interpretation. However, without hormone measurements, it remains unclear whether the observed phase differences are related to estrogen or progesterone. Consequently, all findings reflect general phase-related patterns rather than specific hormonal effects.

Not only were the menstrual periods self-reported, but the daily anxiety scores and contextual factors such as stress and sleep were also self-reported. These results rely on participants' perception and recall. Daily reporting reduces memory bias, but subjective interpretation still plays a role. In addition, several practical issues in survey timing may have affected the precision of the daily data. Tracking survey completion required close monitoring. A separate Excel sheet was used daily to record which participants had completed or missed each survey. Surveys remained open for 48 hours. In some cases, due to this, participants submitted several surveys on the same day, sometimes skipping the current day or filling in earlier ones instead. This may have caused small timeline inaccuracies during preprocessing. The recorded submission dates did not always match the assumed survey dates and could differ by up to two days. Since all entries were treated as if they were completed in chronological order, a few single entries might have been assigned to the wrong phase. Although this was checked by updating the Excel file each day to minimize errors, some surveys may still have been misclassified, as entries in Qualtrics did not indicate which day the responses referred to.

7.3.2 Theory of Mind Study Limitations

Participants completed four ToM tests, with the last two being very similar to the first two. As discussed, order effects such as learning or story familiarity seemed minimal. Yet, seeing similar story structures several times may have made the task more predictable, even without knowing the correct answers from earlier tests. This might have reduced its ability to detect small phase-related differences. In addition, the Adult FPRT may have been too easy for examining affective ToM in a sample not focused on clinical populations. Many participants reached scores near the maximum, suggesting a ceiling effect. This means that the task did not always leave enough room for variation in performance, as most participants performed at or close to the top of the scale. As a result, the test may have been less sensitive to small differences between menstrual phases. A more challenging task could have been more suitable for capturing subtle variation in affective ToM performance.

Cycle day assignment is another limitation that may have affected the results. The intended mid-follicular (days 7-10) and mid-to-late luteal (days 21-24) phases were based on a baseline cycle length of 28 days, as participants' individual cycle lengths were unknown at the start of the experiment. These days may not have represented the exact phases or specific parts of the phases for every participant. For instance, some participants may have been in the early rather than mid-to-late luteal phase when completing a ToM test. For most participants, these days corresponded to the intended phases, but variations in cycle length have caused some mismatches. This reduced the accuracy of phase comparisons. Moreover, only two menstrual phases were examined. Therefore, the findings cannot provide a complete picture of affective ToM performance across the entire cycle. This restricted design was chosen to minimize the

demands on participants during the two- to three-month data collection period.

The statistical approach also introduced certain limitations. The paired samples t-tests compared phase means within participants. This inherently controlled for individual baseline differences. However, it did not consider that the size of phase-related changes could differ between participants, possibly making small within-person effects less apparent.

7.4 Future Work and Implications

For future research, the main limitations of this study should be addressed. A larger sample size would increase the reliability of the findings. Measuring ovarian hormone levels at multiple points throughout the cycle would allow for more accurate phase classification and a clearer understanding of how ovarian hormonal fluctuations relate to anxiety. This study treated the luteal phase as a single category. Separating it into early and late luteal phases would require more precise hormonal information rather than relying solely on cycle days. With such data, future work could examine whether anxiety differs between early and late luteal phases, rather than only assessing the luteal phase as a whole. The existing anxiety dataset could be further explored by examining whether the effects of daily situational stressors and sleep quality on anxiety differ across menstrual phases. In addition, it could be investigated whether individuals with generally higher anxiety scores experienced more daily stressors overall, or whether certain phases were associated with more reported daily stressors. Future work could also include more diverse samples by investigating age differences and contraceptive use. As for the approach to measuring affective ToM performance, using more challenging or varied tasks might reduce ceiling effects. This could help confirm whether ToM is unaffected or reveal subtle changes that the current test could not capture. For instance, the Yoni Task could be more informative. It combines verbal and visual cues to assess both cognitive and affective ToM [64]. The task requires faster and more straightforward reasoning about others' emotions and thoughts than the Adult FPRT. This may provide a clearer view of affective ToM performance. Additionally, all menstrual cycle phases could be investigated instead of only two.

Despite some limitations, this study contributes to the limited understanding of how the menstrual cycle relates to emotional and social functioning. Differences in daily anxiety across menstrual phases show that the menstrual cycle can affect everyday mental well-being. This knowledge can help normalize emotional changes experienced and reduce stigma around menstrual symptoms. It frames these experiences as biologically grounded rather than as overreactions. It can also help people who menstruate understand that their anxious feelings may be heightened by hormonal changes. Recognizing this can lead to greater self-understanding and self-compassion. Gaining insights into affective ToM performance across the menstrual cycle can also enhance self-awareness. It may clarify changes in social understanding and empathy across the cycle, which can influence communication and relationships. Recognizing these patterns may help explain variations in social behavior and emotional connection.
Understanding the relationship between menstrual cycle phases and anxiety also has implications for research and healthcare. Establishing consistent evidence for such patterns would strengthen the case for considering menstrual cycle phases in studies on mood, stress, and emotional health in individuals who menstruate. This could help explain variations in anxiety and lead to more accurate mental health assessments.

8 Conclusion

The main goal of this study was to investigate if and how general anxiety levels vary across menstrual cycle phases in naturally cycling people. Daily state anxiety scores were collected across two full menstrual cycles to answer the question *"How do general anxiety levels vary across different phases of the menstrual cycle in individuals who menstruate and do not use*

hormonal birth control?" The results showed that general anxiety levels significantly differed across menstrual phases (regular cycles). Therefore, the null hypothesis stating that mean anxiety levels do not vary across phases can be rejected. The findings suggest that the menstrual cycle affects anxiety, with higher anxiety during menstruation and lower anxiety after menstruation during the rest of the follicular phase. Anxiety during the luteal phase appears moderate, showing no significant difference from the other phases.

The sub-research question of this study was *"How does affective Theory of Mind performance of naturally cycling individuals differ between the mid-follicular and the mid-to-late luteal phase of the menstrual cycle?"* To answer this, affective ToM performance was measured using adapted versions of the Adult FPRT completed by participants during both phases across two cycles. Affective ToM performance remained stable across phases. For all three measures (total ToM scores, Empathy ratios, and Understanding Inappropriateness ratios), there were no significant differences between the two phases. This means the null hypothesis that affective ToM performance does not differ between these phases cannot be rejected. The findings indicate that affective ToM remains consistent across the menstrual cycle, even when ovarian hormone levels and anxiety fluctuate. However, only two phases were examined, and no hormone levels were measured, so definitive conclusions about the full cycle cannot be drawn. Nevertheless, this study takes an important step toward exploring whether the menstrual cycle affects ToM and social cognition.

Together, the results provide new insight into anxiety levels and social functioning across the menstrual cycle. Although individual patterns sometimes differed, the overall findings support the idea that different cycle phases are associated with changes in anxiety but not necessarily with differences in affective ToM performance.

Acknowledgements

It is important to mention that this research would not have been possible without the participants. Their consistency and commitment made it successful. Many of you shared positive responses regarding the importance of this topic, which motivated me to keep going even when I felt stuck. Furthermore, I would like to thank my primary supervisor, Dr. Tessa Verhoef, for providing guidance throughout the thesis process and for answering my technical questions. I also want to express my appreciation to my secondary supervisor, Dr. Liisa Hantsoo, for her expertise, thoughtful advice, and for taking the time to respond to my questions related to this topic. Not to be forgotten are the friends who read and commented on my thesis, and everyone who provided ongoing support and encouragement. I also appreciate their patience as I obsessively checked Qualtrics every hour of the day for months to see whether participants were still completing the surveys. Thank you.

References

1. United Nations, Department of Economic and Social Affairs, Population Division (2024). World Population Prospects 2024, Online Edition: https://population.un.org/wpp/
2. Rohatgi, A., Dash, S. (2023). Period poverty and mental health of menstruators during COVID-19 pandemic: Lessons and implications for the future. Frontiers in Global Women's Health, 4, 1128169. https://doi.org/10.3389/fgwh.2023.1128169
3. Unicef. Menstrual hygiene: Gender inequality, cultural taboos and poverty can cause menstrual health needs to go unmet. Retrieved August 2025 from: https://www.unicef.org/wash/
4. Schoep, M. E., Nieboer, T. E., van der Zanden, M., Braat, D. D. M., Nap, A. W. (2019). The impact of menstrual symptoms on everyday life: a survey among 42,879 women. American Journal of Obstetrics & Gynecology, 220(6), 569.e1–569.e7. https://doi.org/10.1016/j.ajog.2019.02.048
5. Sharp, G. C., De Giorgio, L. (2023). Menarche, Menstruation, Menopause and Mental Health (4M): a consortium facilitating interdisciplinary research at the intersection of menstrual and mental health. Frontiers in Global Women's Health, 4. https://doi.org/10.3389/fgwh.2023.1258973
6. Critchley, H. O. D., Babayev, E., Bulun, S. E., Clark, S., Garcia-Grau, I., Gregersen, P. K., Kilcoyne, A., Kim, J.-Y. J., Lavender, M., Marsh, E. E., Matteson, K. A., Maybin, J. A., Metz, C. N., Moreno, I., Silk, K., Sommer, M., Simon, C., Tariyal, R., Taylor, H. S., Wagner, G. P., Griffith, L. G. (2020). Menstruation: science and society. American Journal of Obstetrics & Gynecology, 223(5), 624–664. https://doi.org/10.1016/j.ajog.2020.06.004
7. Kumar, M., Singh, S. (2025). Menstrual cycle: An overview. Int J Clin Obstet Gynaecol 2025;9(1):177-180.https://www.doi.org/10.33545/gynae.2025.v9.i1c.1581
8. Del Río, J. P., Alliende, M. I., Molina, N., Serrano, F. G., Molina, S., Vigil, P. (2018). Steroid Hormones and Their Action in Women's Brains: The Importance of Hormonal Balance. Frontiers in Public Health, 6, 141. https://doi.org/10.3389/fpubh.2018.00141
9. Wieczorek, K., Targonskaya, A., Maslowski, K. (2023). Reproductive Hormones and Female Mental Wellbeing. Women, 3(3), 432-444. https://doi.org/10.3390/women3030033
10. Briscoe, H., Vickers-Graver, B., Cherukat, M., Jones, C., Surtees, A. (2024). The link between anxiety and theory of mind in children: A meta-analysis. Journal of Affective Disorders, 367, 530–544. https://doi.org/10.1016/j.jad.2024.08.171
11. Maner, J. K., Miller, S. L. (2014). Hormones and social monitoring: Menstrual cycle shifts in progesterone underlie women's sensitivity to social information. Evolution and Human Behavior, 35(1), 9–16. https://doi.org/10.1016/j.evolhumbehav.2013.09.001
12. Dvash, J., Shamay-Tsoory, S. G. (2014). Theory of Mind and Empathy as Multidimensional Constructs: Neurological Foundations. Topics in Language Disorders 34(4):p 282-295. https://doi.org/10.1097/TLD.0000000000000040
13. Thiyagarajan, D. K., Basit, H., Jeanmonod, R. (2024). Physiology, Menstrual Cycle. In: StatPearls. Treasure Island (FL): StatPearls Publishing. https://www.ncbi.nlm.nih.gov/books/NBK500020/
14. Braidwood, E. (2024, February 27). Menstrual cycle phases: Why it's so much more than your period. Flo. Retrieved August 2025 from: https://flo.health/menstrual-cycle/health/menstrual-cycle-phases.
15. What is menopause? (2024, October 16). National Institute on Aging. Retrieved August 2025 from: https://www.nia.nih.gov/health/menopause/what-menopause

16. Watson, S. (2023, March). Stages of the menstrual cycle. Healthline. Retrieved August 2025 from: https://www.healthline.com/health/womens-health/stages-of-menstrual-cycle#common-issues
17. Bull, J.R., Rowland, S.P., Scherwitzl, E.B. et al. (2019) Real-world menstrual cycle characteristics of more than 600,000 menstrual cycles. npj Digit. Med. 2, 83. https://doi.org/10.1038/s41746-019-0152-7
18. Villines, Z. (2023). What to know about irregular periods. Retrieved October 2025 from: https://www.medicalnewstoday.com/articles/178635
19. Telfer, N. (2021). Clue. What's "normal"?: menstrual cycle length and variation. Retrieved August 2025 from: https://helloclue.com/articles/cycle-a-z/what's-normal-menstrual-cycle-length-and-variation
20. Mount Sinai Health System. (2019). Menstrual disorders. Retrieved November 2025 from: https://www.mountsinai.org/health-library/report/menstrual-disorders
21. Cornforth, T. (2023). Body changes during the menstrual cycle. Verywell Health. Retrieved August 2025 from: https://www.verywellhealth.com/the-menstrual-cycle-3520919#citation-8
22. Madormo C. (2024, May 20). When ovulation occurs in the menstrual cycle. Verywell Health. Retrieved August 2025 from: https://www.verywellhealth.com/when-does-ovulation-occur-8633617
23. Mesen, T. B., Young, S. L. (2015). Progesterone and the luteal phase: a requisite to reproduction. Obstetrics and gynecology clinics of North America, 42(1), 135–151. https://doi.org/10.1016/j.ogc.2014.10.003
24. Campbell, N. A., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., Orr, R. B. (2020). Biology: A global approach (12th ed., Global ed.). Pearson. ISBN: 9781292341750
25. Reed, B. G., Carr, B. R. (2018). The Normal Menstrual Cycle and the Control of Ovulation. In: Feingold K. R., Ahmed S. F., Anawalt B., et al., editors. Endotext. South Dartmouth (MA): MDText.com, Inc. https://www.ncbi.nlm.nih.gov/books/NBK279054/
26. Dias Da Silva, I., Wuidar, V., Zielonka, M., Pequeux, C. (2024). Unraveling the Dynamics of Estrogen and Progesterone Signaling in the Endometrium: An Overview. Cells, 13(15), 1236. https://doi.org/10.3390/cells13151236
27. Wharton, W., Gleason, C. E., Olson, S., Carlsson, C. M., Asthana, S. (2012). Neurobiological underpinnings of the estrogen–mood relationship. Current Psychiatry Reviews, 8(3). https://doi.org/10.2174/157340012800792957
28. Normal Menstrual cycle | UCSF Center for Reproductive Health. UCSF. Retrieved September 2025 from: https://crh.ucsf.edu/about-fertility/normal-menstrual-cycle
29. Felman, A. (2025, June 24). What to know about anxiety. Retrieved August 2025 from: https://www.medicalnewstoday.com/articles/323454
30. Spielberger, C. D., Gorsuch R. L., Lushene R. E. (1972), The State-Trait Anxiety Inventory (STAI). PaloAlto, CA: Consulting Psychologists Press. PDF retrieved from: https://www.advancedassessments.co.uk/resources/Mental-Health-Test.pdf
31. Premack D, Woodruff G. (1978) Does the chimpanzee have a theory of mind? Behavioral and Brain Sciences. 1978;1(4):515-526. https://doi.org/10.1017/S0140525X00076512
32. Sebastian, C. L., Fontaine, N. M. G., Bird, G., Blakemore, S., De Brito, S. A., McCrory, E. J. P., Viding, E. (2012) Neural processing associated with cognitive and affective Theory of Mind in adolescents and adults, Social Cognitive and Affective Neuroscience, Volume 7, Issue 1, Pages 53–63, https://doi.org/10.1093/scan/nsr023
33. Raimo, S., Cropano, M., Roldán-Tapia, M. D., Ammendola, L., Malangone, D., Santangelo, G. (2022). Cognitive and Affective Theory of Mind across Adulthood. Brain Sciences, 12(7), 899. https://doi.org/10.3390/brainsci12070899
34. Stone, V. E., Baron-Cohen, S., Knight, R. T. (1998). Frontal lobe contributions to theory of mind. Journal of Cognitive Neuroscience, 10(5), 640–656. https://doi.org/10.1162/089892998562942. FPRT PDF retrieved from: https://docs.autismresearchcentre.com/tests/FauxPas_Adult.pdf
35. Mackenzie, A. C. L., Chung, S., Hoppes, E., Mickler, A. K., Cartwright, A. F. (2024). Measurement of changes to the menstrual cycle: A transdisciplinary systematic review evaluating measure quality and utility for clinical trials. PLOS ONE, 19(7), e0306491. https://doi.org/10.1371/journal.pone.0306491
36. Reynolds, T. A., Makhanova, A., Marcinkowska, U. M., Jasienska, G., McNulty, J. K., Eckel, L. A., Nikonova, L., Maner, J. K. (2018). Progesterone and women's anxiety across the menstrual cycle. Hormones and Behavior, 102, 34–40. https://doi.org/10.1016/j.yhbeh.2018.04.008
37. Gurski, D., Gurski, Y. Flo Health, Inc. Period tracker, ovulation tracker, pregnancy app. https://flo.health/

38. Hantsoo, L., Rangaswamy, S., Voegtline, K., Salimgaraev, R., Zhaunova, L., Payne, J. L. (2022, August) Premenstrual symptoms across the lifespan in an international sample: data from a mobile application. Arch Womens Ment Health 25, 903–910. https://doi.org/10.1007/s00737-022-01261-5
39. Nillni, Y. I., Toufexis, D. J., Rohan, K. J. (2011). Anxiety sensitivity, the menstrual cycle, and panic disorder: A putative neuroendocrine and psychological interaction. Clinical Psychology Review, 31(7), 1183–1191. https://doi.org/10.1016/j.cpr.2011.07.006
40. Green, S. A., Graham, B. M. (2022). Symptom fluctuation over the menstrual cycle in anxiety disorders, PTSD, and OCD: A systematic review. Archives of Women's Mental Health, 25, 71–85. https://doi.org/10.1007/s00737-021-01187-4
41. Jarvis, A. L., Keage, H. A. D., Wong, S., Weightman, M., Stephens, R. G. (2023). Evidence for a multidimensional account of cognitive and affective theory of mind: A state-trace analysis. Memory & Cognition, 52(3), 525–535. https://doi.org/10.3758/s13421-023-01481-9
42. Baez, S., Tangarife, M. A., Davila-Mejia, G., Trujillo-Güiza, M., Forero, D. A. (2023). Performance in emotion recognition and theory of mind tasks in social anxiety and generalized anxiety disorders: a systematic review and meta-analysis. Frontiers in Psychiatry, 14. https://doi.org/10.3389/fpsyt.2023.1192683
43. Foulds, C., Khudiakova, V., Surtees, A., et al. (2025, June 18). Do state and trait general and social anxiety affect theory of mind? PREPRINT (Version 1). Research Square. https://doi.org/10.21203/rs.3.rs-6855230/v1
44. Derntl, B., Kryspin-Exner, I., Fernbach, E., Moser, E., Habel, U. (2008). Emotion recognition accuracy in healthy young females is associated with cycle phase. Hormones and Behavior, 53(1), 90–95. https://doi.org/10.1016/j.yhbeh.2007.09.006
45. Osório, F. L., De Paula Cassis, J. M., De Sousa, J. P. M., Poli-Neto, O., Martín-Santos, R. (2018). Sex Hormones and Processing of Facial Expressions of Emotion: A Systematic Literature review. Frontiers in Psychology, 9. https://doi.org/10.3389/fpsyg.2018.00529
46. Rafiee, Y., Stern, J., Ostner, J., Penke, L., Schacht, A. (2023). Does emotion recognition change across phases of the ovulatory cycle? Psychoneuroendocrinology, 148, 105977. https://doi.org/10.1016/j.psyneuen.2022.105977
47. Jang, D., Lybeck, M., Cortes, D. S., Elfenbein, H. A., Laukka, P. (2024) Estrogen predicts multimodal emotion recognition accuracy across the menstrual cycle. PLOS ONE 19(10): e0312404. https://doi.org/10.1371/journal.pone.0312404
48. Derntl, B., Hack, R. L., Kryspin-Exner, I., Habel, U. (2013). Association of menstrual cycle phase with the core components of empathy. Hormones and Behavior, 63(1), 97–104. https://doi.org/10.1016/j.yhbeh.2012.10.009
49. Decety, J., Jackson, P. L. (2004). The Functional Architecture of Human Empathy. Behavioral and Cognitive Neuroscience Reviews, 3(2), 71-100. https://doi.org/10.1177/1534582304267187
50. Kimmig, A.-C. S., Wildgruber, D., Wendel, S.-M. U., Sundström-Poromaa, I., Derntl, B. (2021). Friend vs. Foe: Cognitive and affective empathy in women with different hormonal states. Frontiers in Neuroscience, 15. https://doi.org/10.3389/fnins.2021.608768
51. Mier, D., Lis, S., Neuthe, K., Sauer, C., Esslinger, C., Gallhofer, B. and Kirsch, P. (2010), The involvement of emotion recognition in affective theory of mind. Psychophysiology, 47: 1028-1039. https://doi.org/10.1111/j.1469-8986.2010.01031.x
52. Mentese, B., Kutlu, N. (2023). Menstrual cycle phase influences on neurocognitive functions: a comuterised Psychometric assessment. Article DOI: 10.47739/2578-3718/1105. https://www.jscimedcentral.com/jounal-article-info/JSM-Sexual-Medicine/Menstrual-Cycle-Phase-Influences-on-Neurocognitive-Functions%3A-A-Comuterised-Psychometric-Assessment-11034#
53. Flores-Ramos, M., Silvestri Tomassoni, R., Guerrero-López, J. B., Salinas, M. (2017). Evaluation of trait and state anxiety levels in a group of peri- and postmenopausal women. Women & Health, 58(3), 305–319. https://doi.org/10.1080/03630242.2017.1296059
54. Born, C., Seitz, N.-N., Grunze, H., Vieta, E., Dittmann, S., Seemüller, F., Amann, B. (2009). Preliminary results of a fine-grain analysis of mood swings and treatment modalities of bipolar I and II patients using the daily prospective life-chart-methodology. Acta Psychiatrica Scandinavica, 120, 474–480. https://doi.org/10.1111/j.1600-0447.2009.01412.x
55. Nordholt, S., Garrison, P., Aichhorn, W., Ochs, M., Schiepek, G. (2024). Pattern transitions in diary data of MDD patients: A mixed-methods multiple case study of psychotherapy dynamics. Frontiers in Psychology, 15. https://doi.org/10.3389/fpsyg.2024.1259610
56. Dodell-Feder, D., Lincoln, S. H., Coulson, J. P., Hooker, C. I. (2013). Using fiction to assess mental state understanding: A new task for assessing theory of mind in adults. PLoS ONE, 8(11), e81279. https://doi.org/10.1371/journal.pone.0081279

57. Gregory, C., Lough, S., Stone, V., Erzinclioglu, S., Martin, L., Baron-Cohen, S., Hodges, J. R. (2002). Theory of mind in patients with frontal variant frontotemporal dementia and Alzheimer's disease: Theoretical and practical implications. Brain, 125(4), 752–764. https://doi.org/10.1093/brain/awf079
58. Đorđević, J., Pavlović, A., Mihajlović, G., Hinić, D., Vojvodić, J., Živanović, M., Pavlović, D. (2025). Assessing theory of mind abilities in schizophrenia and bipolar disorder: A psychometric study of the Faux Pas test in Serbian. Psihologija, 57(4), 373–391. https://doi.org/10.2298/PSI220824006D
59. Rademaker, N. (2025). Tracking anxiety across the menstrual cycle. https://beyondthemoodswings.netlify.app/
60. JASP - a fresh way to do statistics. (2025). JASP - Free and User-Friendly Statistical Software. https://jasp-stats.org/
61. Bobbitt, Z. (2021). The three assumptions of a repeated measures ANOVA. Statology. Retrieved October 2025 from: https://www.statology.org/repeated-measures-anova-assumptions/
62. Bobbitt, Z. (2022). The three assumptions made in a paired T-Test. Statology. https://www.statology.org/paired-t-test-assumptions/
63. Cohen, J. (1992). A power primer. Psychological Bulletin, 112(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155
64. Isernia, S., Rossetto, F., Blasi, V., Massaro, D., Castelli, I., Ricci, C., Shamay-Tsoory, S., Marchetti, A., Baglio, F. (2023). Measuring cognitive and affective Theory of Mind with the Italian Yoni task: normative data and short versions. Curr Psychol 42, 23519–23530. https://doi.org/10.1007/s12144-022-03457-5

A Appendix: Anxiety Levels (STAI-S)

A.1 STAI-S Statements

For the STAI-S, participants are asked to rate how they felt at the moment using a 1-4 scale (1 = "Not at all", 2 = "A little", 3 = "Somewhat", 4 = "Very much so") for each of the following statements [30]:

1. I feel calm
2. I feel secure
3. I feel tense
4. I feel strained
5. I feel at ease
6. I feel upset
7. I am presently worrying over possible misfortunes
8. I feel satisfied
9. I feel frightened
10. I feel uncomfortable
11. I feel self-confident
12. I feel nervous
13. I feel jittery
14. I feel indecisive
15. I am relaxed
16. I feel content
17. I am worried
18. I feel confused
19. I feel steady
20. I feel pleasant

A.2 Mean STAI-S Scores Per Participant

Table A1 and Table A2 present the average STAI-S scores per menstrual phase for participants with regular and irregular cycles, respectively. These values represent each participant's mean state anxiety level within the corresponding menstrual phase, based on their daily survey responses.

Participant ID	Menstrual	Follicular incl.	Follicular excl.	Luteal
P001	53.62	48.48	42.92	37.68
P002	60.60	53.41	50.14	59.36
P003	57.36	55.48	54.45	52.62
P004	31.44	33.04	36.25	34.82
P005	42.67	39.53	36.00	37.71
P006	49.00	46.48	45.29	45.82
P008	35.70	31.92	29.56	31.50
P009	55.08	44.58	34.08	40.30
P010	51.60	47.46	44.35	52.21
P011	31.00	29.21	27.79	26.46
P013	63.00	45.61	34.35	59.64
P014	53.17	51.21	50.78	49.69
P017	32.40	33.07	33.42	37.50
P019	33.13	37.90	42.38	41.88
P020	45.13	43.41	42.23	47.43
P022	43.14	37.28	35.64	50.84
P024	35.00	32.92	30.09	31.93
P026	25.80	22.88	21.55	24.21
P027	40.38	39.38	39.16	40.54
P028	67.11	44.87	30.57	58.29

Table A1: Average STAI-S scores per menstrual phase for participants with regular cycles. The "follicular incl." phase includes menstrual days, while the "follicular excl." phase excludes them. STAI-S scores can range from 20 (low anxiety) to 80 (high anxiety).

Participant ID	Menstrual	Follicular incl.	Follicular excl.	Luteal
P007	37.20	37.20	37.20	36.36
P012	46.29	46.56	46.78	49.79
P015	48.33	47.69	47.61	58.07
P021	38.14	36.42	36.03	35.36

Table A2: Average STAI-S scores per menstrual phase for participants with irregular cycles. The "follicular incl." phase includes menstrual days, while the "follicular excl." phase excludes them. STAI-S scores can range from 20 (low anxiety) to 80 (high anxiety).

B Appendix: Theory of Mind Tests

B.1 Adult Faux Pas Recognition Test

The following instructions were given to the participants during the Theory of Mind tests in Qualtrics:

"This test contains 10 short stories designed to assess Affective Theory of Mind (ToM). Affective ToM refers to the ability to understand how others feel.
For each story, carefully read the text and then answer the questions that follow. These questions may ask you to identify whether someone said something awkward or inappropriate (a faux pas), how the characters might feel about what happened, and whether they understood

each other's feelings.
You're absolutely welcome to go back to the story while answering the questions. There's no need to rely on memory alone. It's completely understandable to miss a detail the first time. Feel free to revisit the story whenever you need to."

This is a faux pas story (Story 2) from the Adult Faux Pas Recognition Test:
Helen's husband was throwing a surprise party for her birthday. He invited Sarah, a friend of Helen's, and said, "Don't tell anyone, especially Helen."
The day before the party, Helen was over at Sarah's, and Sarah spilled some coffee on a new dress that was hanging over her chair.
"Oh!" said Sarah, "I was going to wear this to your party!"
"What party?" said Helen.
"Come on," said Sarah, "Let's go see if we can get the stain out."

For each short story, the test contains 8 questions, such as:

1. Did anyone say something they shouldn't have said or something awkward?
2. Who said something they shouldn't have said or something awkward?
3. Why shouldn't he/she have said it, or why was it awkward?
4. Why do you think he/she said it?
5. Did Sarah remember that the party was a surprise party?
6. How do you think Helen felt?
7. In the story, who was the surprise party for?
8. What got spilled on the dress?

These questions belong to Story 2 from the Adult Faux Pas Recognition Test [34]. Questions 7 and 8 are the control questions for every story. They test whether the participant understood the story's content [34]. Participants who answered "no" to the first question (faux pas detection) of a story were not shown questions 2–6. In that case, they only answer questions 1, 7, and 8.

B.2 Test composition

The story composition of the four Theory of Mind test subsets used in this study is listed below.

- Subset 1: story 2, 14, 8, 12, 10, 7, 1, 19, 5, 16.
- Subset 2: story 6, 15, 11, 9, 13, 4, 20, 18, 3, 17.
- Subset 3 (deviated version of subset 1): story 7, 10, 14, 8, 16, 5, 1, 19, 12, 2.
- Subset 4 (deviated version of subset 2): story 9, 18, 4, 6, 20, 17, 11, 3, 13, 15.

Story numbering refers to the original Adult Faux Pas Recognition Test [34]. Each subset contained five control stories and five faux pas stories.

B.3 Phase Assignment

Table B1 gives an overview of the menstrual cycle phases during which each participant completed the four Theory of Mind tests for two consecutive cycles. This is under the assumption that the mid-follicular phase occurred approximately on cycle days 7-10 and the mid-to-late luteal phase on days 21-24. The resulting phase assignments were used to link Theory of Mind test scores to menstrual phases for data analysis.

Participant ID	Test 1	Test 2	Test 3	Test 4
P001	Follicular	Luteal	Follicular	Luteal
P002	Follicular	Luteal	Follicular	Luteal
P003	Luteal	Follicular	Luteal	Follicular
P004	Luteal	Follicular	Luteal	Follicular
P005	Follicular	Luteal	Follicular	Luteal
P006	Luteal	Follicular	Luteal	Follicular
P008	Luteal	Follicular	Luteal	Follicular
P010	Follicular	Luteal	Follicular	Luteal
P013	Luteal	Follicular	Luteal	Follicular
P017	Luteal	Follicular	Luteal	Follicular
P019	Luteal	Follicular	Luteal	Follicular
P022	Luteal	Follicular	Luteal	Follicular
P024	Follicular	Luteal	Follicular	Luteal
P025	Follicular	Luteal	Follicular	Luteal
P026	Luteal	Follicular	Luteal	Follicular
P027	Luteal	Follicular	Luteal	Follicular
P028	Luteal	Follicular	Luteal	Follicular

Table B1: Overview of the menstrual cycle phases during which each participant completed the four affective Theory of Mind (ToM) tests. Each participant completed a subset of stories during cycle days 7-10 (assumed mid-follicular phase) and cycle days 21-24 (assumed mid-to-late luteal phase) across two cycles.

B.4 Total Theory of Mind Score Per Participant

Table B2 shows the Theory of Mind score per test for each participant. Each total score reflects the overall Theory of Mind performance for that specific test, based on the scoring system of the Adult Faux Pas Recognition Test.

Participant ID	Test 1	Test 2	Test 3	Test 4
P001	0.88	0.72	0.84	0.80
P002	0.68	0.76	0.66	0.71
P003	0.90	0.86	0.84	0.86
P004	0.83	0.66	0.62	0.56
P005	0.75	0.90	0.70	0.67
P006	0.86	0.75	1.00	0.98
P008	0.88	0.80	0.90	0.88
P010	0.84	0.62	0.80	0.88
P013	0.96	1.00	0.84	1.00
P017	0.86	0.90	0.95	0.88
P019	0.90	0.72	0.86	0.88
P022	0.97	0.88	0.87	0.86
P024	0.94	0.82	0.94	0.72
P025	1.00	0.96	0.98	0.94
P026	1.00	0.96	0.98	0.94
P027	0.96	0.90	0.96	0.96
P028	0.90	0.94	0.90	0.98

Table B2: Overview of the total Theory of Mind (ToM) score per test for each participant. The total score is the average of five ratios describing different aspects of the test: Faux Pas Detection, Understanding Inappropriateness, Intentions, Belief, and Empathy. The minimum possible total score is 0, and the maximum is 1.