Investigating the Influence of Hormonal Contraceptive Use on Perseverance in Cognitively Demanding Tasks

Andrea Dioni Munksgaard (201907480@post.au.dk)
School of Communication and Cognition, University of Aarhus,
Jens Chr. Skous Vej 2, 8000 Aarhus, Denmark

Maria Nissen Byg (201906706@post.au.dk)
School of Communication and Cognition, University of Aarhus,
Jens Chr. Skous Vej 2, 8000 Aarhus, Denmark

Abstract

A limited but growing body of research suggests unintended cognitive side effects associated with the use of hormonal contraceptives. This study investigates the association between hormonal contraceptive use and perseverance in cognitively challenging tasks among Danish females. The study attempts a conceptual replication of Bradshaw and colleagues (2020) and hypothesizes that women using hormonal contraceptives spend less time on solving cognitively demanding tasks compared to naturally cycling women. As an extension of this, the study furthermore hypothesizes that women using hormonal contraceptives produce fewer correct responses in cognitively demanding tasks compared to their naturally cycling counterparts. Using a cross-sectional repeated-measures design, the experiment was conducted on 66 participants. All data was analyzed through multilevel Bayesian models with crossed varying effects. The study does not find sufficient evidence in favor of any of the hypotheses.

Keywords: *Hormonal contraception, Perseverance, Performance, Bayesian modelling, Naturally cycling women, Anagrams, HMT-(S)*

Introduction (MB, AM)[1]

More than 100 million women worldwide use hormonal contraceptives (HCs) (B. A. Pletzer & Kerschbaum, 2014). The main purpose for most women is to avoid unwanted pregnancy, but birth control can also provide other benefits such as lessening period cramps, bettering acne, reducing the risk of certain types of cancer, and is furthermore employed as treatment for certain hormonal disorders (Hawkshead, 2013).

All in all, the pill seems like a wonder-cure for a plethora of things - so what is the issue? In recent years, the scientific community has seen a surge in work exploring the effects of HC use on cognition (Bradshaw et al., 2020), with findings suggesting a variety of implications. Areas include, but are not limited to, learning, memory, exercise, and neurological structural changes (Bianchini et al., 2018; Herzberg et al., 2017; Nielsen et al., 2013; B. A. Pletzer & Kerschbaum, 2014). The nature of the different findings vary, with some expressing negative implications, others positive. A pressing issue is the lack of conclusive evidence regarding the unintended consequences of HC use, be they physical, physiological or cognitive. Many recent findings are either contradicted by other papers, inconclusive, or suffering from methodological pitfalls (Warren et al., 2014). The lack of consistent and methodologically sound findings makes it difficult for women to assess the potential influence that HC use may have on them: Knowledge that is necessary in making an informed choice about their bodies and brains.

This paper will limit itself to implications related to perseverance and cognitive performance. However, we cannot stress enough that general implications may go well beyond the scope of what is explored in this paper, potentially leading to broad societal and social consequences.

[1] Abbreviations of the co-authors' names. The responsible author is henceforth listed first for each section

The Area of Interest (AM, MB)

Recent research suggests that HC use is associated with decreased perseverance in cognitively demanding tasks. In a paper by Bradshaw, Mengelkoch, and Hill (2020), it was found that women who use HCs spent significantly less time on cognitively challenging tasks compared to their naturally cycling counterparts. This difference in perseverance led to decreased performance in the tasks, which comprised of both anagram-solving and GRE[2] questions. As the researchers have expressed the need for such studies to be replicated (Bradshaw et al., 2020), it is with this appeal in mind that we conduct our research. In doing so, we aim at contributing to the growing body of research on perseverance and performance implications of HC use.

In the following sections, an explanation of the biological workings of both the menstrual cycle and HCs will be presented. After this, the paper will be exploring different findings associated with decreased perseverance and performance.

The Menstrual Cycle (MC) (MB, AM)

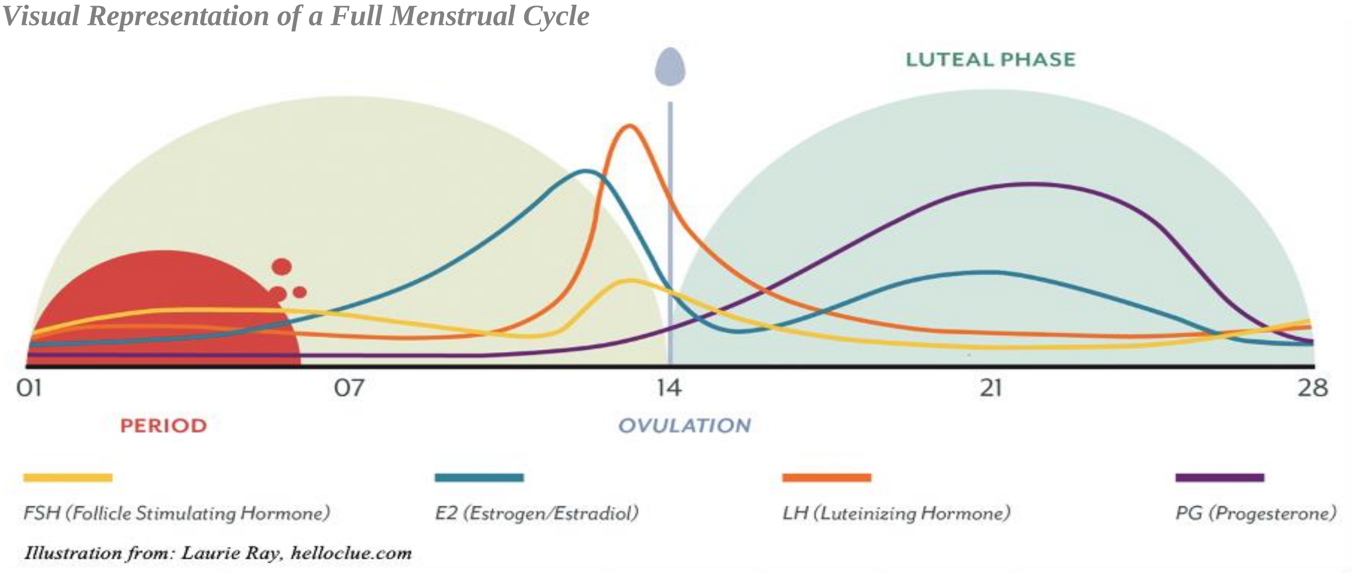
The typical text-book menstrual cycle lasts for 28 days before it repeats itself. However, most women do not experience an exact 28-day cycle, with averages varying between 21 and 35 days (Harlow & Ephross, 1995). The MC consists of three distinct phases; the follicular phase, ovulation and the luteal phase. The inner workings and cognitive implications differ throughout these three phases, as does hormone levels (Hawkshead, 2013).

The follicular phase marks the preovulatory phase of the MC and typically lasts 13-15 days in a 28 day cycle, starting on the first day of the period (Harlow & Ephross, 1995). The follicular phase is set off when the endometrial lining, which has been built up in preparation for pregnancy, is shed (presuming no presence of a fertilized egg). It is this shedding of the endometrial lining that results in menstrual

[2] Standardized test questions (Graduate Record Examinations) used in American high schools

Maria Nissen Byg (201906706)
Andrea Dioni Munskgaard (201907480)

Visual Representation of a Full Menstrual Cycle



bleeding (Hill, 2019, p. 27). After the menstruation is over and bleeding has stopped, the latter part of the f ollicular phase sets in, where the pituitary gland releases follicle-stimulating hormones (FSH), triggering the production of eggs (Watson, 2018). During the early follicular phase, the concentration of sex hormones such as progesterone and estrogen are generally at nadir level, though estrogen spikes in the last few days of this phase, leading up to ovulation. This spike of estrogen causes the pituitary gland to release luteinizing hormones (LH) that start the ovulation phase (Bianchini et al., 2018). Ovulation will occur approximately 10-12 hours after the peak of luteinizing hormones, and a mature egg will be released from the follicle, ready for fertilization (Ray, 2018).

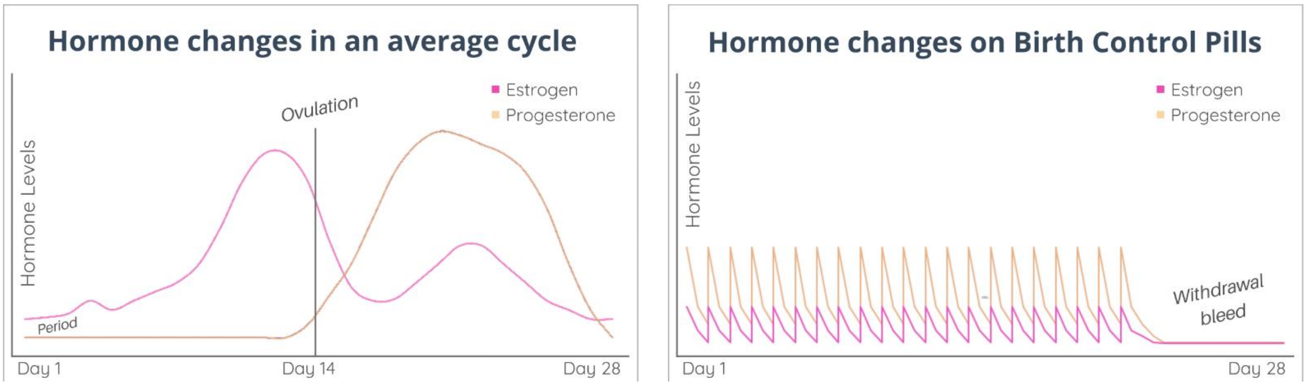
The last part of the menstrual cycle is the luteal phase, which prepares the body for either receiving a fertilized egg or getting ready for a new cycle. The recently vacated follicle, or “corpus luteum”, prompted by LH, begins producing large amounts of progesterone and smaller amounts of estrogen in order for a possible pregnancy to develop. If the released egg has not been fertilized by the late luteal phase, progesterone and estrogen levels will return to baseline and the menstrual cycle will repeat (Druet, 2017).

Cognitively, research suggests that the luteal phase, with its relative peaks in both estrogen and progesterone, may be associated with better verbal abilities as well as impaired visuo-spatial abilities for women tested during this part of their cycle (Hampson, 1990; Hausmann et al., 2000; Maki et al., 2002). However, findings are not always found (Rosenberg & Park, 2002), and recent reviews conclude that any cycle-dependent changes in cognitive functions are small compared to other factors (B. Pletzer et al., 2019).

Thus, the entire menstrual cycle can be seen as a feedback loop between FSH, estrogen, progesterone, and LH. This loop also involves the hypothalamus, and, as mentioned, the pituitary gland, but in short, it is this loop that is broken by the active progestin in the oral contraceptives.

Oral Contraceptives (OC) (AM, MB)
Most OCs are combination pills, containing both estrogen and progestins, a form of artificial progesterone usually derived from testosterone. It is these progestins that act as the main mechanism in OCs (Hill, 2019, p. 83-84). A typical pack of birth control will contain 21 combination pills and 7 sugar pills. The seven day rest period allows for the pituitary-ovarian axis to regain activity, and for the period to start (Bianchini et al., 2018).

A less popular choice of OCs are progestin-only pills. The mechanics of progestin-only pills are similar to that of combination pills, since progestin is the main workhorse in preventing pregnancy. However, many women rely on combination pills with both progestin



Illustrations from: Rach, rachcondon.com

and estrogen, since they prefer how they feel on both hormones together (Hill, 2019, p. 80).

There are currently three different generations of OCs in Denmark, each with different progestins and side effects[3] . For the sake of simplicity, the following explanation covers how progestins, regardless of specific kind, work in general. OCs, simply speaking, work by mimicking the mid-luteal phase where the concentration of progesterone and estrogen is at its highest. The artificially created constant high levels of sex hormones prevent FSH and LH from being released in large quantities, effectively stopping ovulation from occurring, which eliminates the possibility of pregnancy (Hill, 2019, p. 78-79). Sarah Hill (2019) refers to this phenomenon as "hormonal deja-vú", pointing to the repetition of hormonal spikes that occur each day.

Intrauterine Devices (IUDs) (MB, AM)
IUDs are similar to OCs in that the main active agent is progestin, specifically levonorgestrel. After the IUD has been inserted in the uterus, small amounts of progestin are excreted locally day-to-day. With only a small amount of hormones entering the bloodstream, estrogen and progesterone levels for women with IUDs are similar to that of naturally cycling women. The implant thus works through the small amount of released progestin in the area of interest, which makes the cervical mucus too thick for sperm to enter through (Bedsider, 2017). Generally speaking, IUDs are seen as a less invasive method of hormonal contraception. However, it is not uncommon for studies to find similar effects for women on IUDs, as they do for women on the pill (Bradshaw et al., 2020; Skovlund et al., 2016).

Implications (AM, MB)
Decreased perseverance has implications relating to a broad range of areas. Some research finds perseverance to be predictive of success benchmarks in a wide variety of areas (Duckworth et al., 2007), including job performance, academic achievement and learning. With more women than ever enrolling in higher education and undertaking ambitious job positions (*studyinternational.com*, 2018), the notion that perseverance may be negatively affected by the use of HCs is to be taken seriously. While the association between HC use and perseverance has yet to be explored in full, research regarding the link between hormones and other cognitive areas with a possible mediating effect on perseverance have been carried out. In the following sections, areas of research with relevant implications on perseverance will be presented.

[3] A short explanatory overview from of the three generations as well as their differences can be found in appendix 1, from Hill (2019, p. 83)

Cortisol (MB, AM)

Due to its impact in regulating the human stress-response, cortisol is sometimes referred to as “nature’s built-in alarm system”. Although the hormone is generally considered harmful in too great quantities (Coffman, 2020), it has several important implications that positively impact and regulate human behavior when released in the right amounts (Hoyt et al., 2016). A typical and healthy rise in cortisol happens both during situations where a negative impact is felt, but also in situations of importance, e.g. during high arousal or excitement, marking that something consequential is happening. Such acute stress responses have been shown to facilitate working memory and learning (Yuen et al., 2009).

With cortisol's impact in regulating human behavior, it is notable that users of HCs have been shown to exhibit cases of hypercortisolism. Hypercortisolism is a state in which long-term exposure to high levels of cortisol causes the body to shut down its response to this hormone (Hertel et al., 2017; Kirschbaum et al., 1999).

Spikes in cortisol are caused by activation of the hypothalamic-pituitary-adrenal (HPA) axis, which – along with activation of the sympathetic nervous system – is a key process in response to stress. While dynamic bursts of the HPA axis are important, chronic activation of cortisol signaling has the potential of harming the body (Hill, 2019, p. 152). Because of this, a chronic activation of the HPA axis will lead the body to release proteins that inactivate some of said cortisol. Furthermore, the body will lower the response of the pituitary and adrenal glands to signals telling them to produce more cortisol (Herman et al., 2016). Consequently, cortisol spikes in response to stress are dampened.

While states of hypercortisolism as described above typically occur in people having been under chronic stress for an extended amount of time, similar dampened cortisol spikes are observed in women on HCs (Kirschbaum et al., 1996). While research has yet to pinpoint what makes women’s HPA axes go into overdrive when on HCs, findings reliably show evidence of said altered and dampened cortisol response in relation to stress (Mordecai et al., 2017).

With cortisol being a broad modulator of both body and brain, a wide array of implications is to be expected when the response of this hormone gets dampened. Structurally, cortisol overdrive can lead to shrinkage of brain mass, especially in the hippocampus, which is most commonly associated with learning and memory (Conrad, 2008), as well as motivation and emotion (Frodl et al., 2006). Lower hippocampal volume is indeed observed in pill-taking women compared to naturally cycling women, hinting at possible unintended consequences of HC use other than that of regulating the MC (Hertel et al., 2017).

Other than the measurable decrease in brain mass, hypercortisolism can also be measured behaviorally: If reaching a shutdown of dynamic changes in cortisol response, negative side effects such as decreased ability for coping, learning, and adapting to new environments are to be expected (Conrad, 2010; Kinlein et al., 2015). Furthermore, the absence of high-arousal flagging of events, such as threats or opportunities, may leave women’s brains feeling unstimulated. Together, these consequences are potential modulators of perseverance.

Performance, Academic Life and Self-Control (AM, MB)

On a population-specific level, research suggests that academic performance may be negatively influenced as a consequence of the altered hormone level of HC users and the dampened stress response that follows (Hill 2019, p. 161). For instance, hypercortisolism can lead to cell proliferation in the hippocampus, an area often linked to learning and memory (Bradshaw et al., 2020).

In 2018, Gregory et al., conducted a study investigating the relation between HC use, depression and academic performance among college students. Findings suggest that HC use leads to increased odds of reporting academic performance issues in the youngest age group. Moreover, HC use was found to increase the odds of being diagnosed with depression for all age groups, which in turn was found to increase odds of academic performance issues for all ages groups. This effect could be imagined to be further fueled by a lack of self-control. Reported associations between self-control and HC use indicate that HC use is associated with a decrease in self-control (Gregory et al., 2018; Zethraeus et al., 2017), an area highly linked to perseverance.

Women using HCs have demonstrated a difference in performance compared to naturally cycling women in a wide variety of tasks, i.e. verbal memory and fluency, as well as visuospatial tasks. Some studies demonstrated significantly lowered performance on verbal fluency tasks compared to naturally cycling women (Griksiene & Ruksenas, 2011), with the performance being lowest for users of androgenic oral contraceptives (see appendix 1). It should be noted that results on visuospatial tasks are inconsistent across studies, as other studies have found HC users to have improved ability (Cicinelli et al., 2011; Wharton et al., 2008) or no change in such tasks (Mordecai et al., 2008).

Considering the lack of consistent findings, combined with the limited scope of most studies investigating consequences of HC use, we imagine that any real-world effects would presumably be small. While much research finds areas linked to perseverance to be widely predictive of success, Bradshaw and colleagues (2020) still remain the main contributors to research specifically on perseverance. Therefore, studies attempting replication are of the essence.

Combined, the above possible unintended consequences of HC use, as well as the lack of consistent findings regarding cognitive effects and implications of HCs, lead us to our own research.

The Current Study (MB, AM)

We set out to create an experimental design that is a conceptual replication of the Bradshaw et al. 2020 paper. This is with the purpose of investigating whether the relationship between HC use, perseverance, and performance in cognitively demanding tasks is reproduced in a Danish setting.

In the original study, researchers made a plea for investigations like this to be carried out with different sample sizes and possibly different tasks. As previously mentioned, the original study had anagram-solving (both solvable and unsolvable) and GRE questions as their tasks, as there is plenty of evidence suggesting that language, logic, and math tasks are associated with exercise of cognitive control (Gilhooly & Fioratou, 2009). In order to test whether the effect on perseverance is found in other tasks, we opted for including solvable anagrams and IQ-matrix puzzles as our two tasks. By keeping one original task, we would be able to compare results more directly, of course keeping in mind that the anagrams were inherently different due to language differences. Furthermore, we chose not to include unsolvable anagrams, despite them being a measure of persistence during perceived failure (Ventura et al., 2013), as we wanted to keep a study design that would mimic real life tasks (which usually have a solution). The IQ-puzzle task was deemed different enough from the original GRE task to investigate possible effect-replication when new tasks were included, while still being largely driven by similar logical reasoning.

With the aforementioned implications and these methodological considerations in mind, we settled on investigating the association between HC use and perseverance with the following hypotheses:

H1: Women using hormonal contraceptives spend less time on solving cognitively demanding tasks compared to naturally cycling women. Specifically:

- $H1_{P1}$: Women using hormonal contraceptives spend less time on solving anagrams compared to naturally cycling women.

- $H1_{P2}$: Women using hormonal contraceptives spend less time on solving IQ-puzzles compared to naturally cycling women.

H2: As an extension of the above hypotheses, we expect women using hormonal contraceptives to have fewer correct responses in cognitively demanding tasks compared to naturally cycling women. Specifically:

- $H2_{P1}$: We expect women using hormonal contraceptives to have fewer correct responses in anagram tasks compared to naturally cycling women.
- $H2_{P2}$: We expect women using hormonal contraceptives to have fewer correct responses in IQ-puzzle tasks compared to naturally cycling women.

Materials and methods

Participants (AM, MB)
Participants were primarily recruited through online groups strictly for women, though none of the groups had any theme of, or affiliation with, anything related to menstruation or HCs. All participants were naive to the true purpose of the study prior to participation. The sample of participants consisted of 94 Danish-speaking females. Of the sample, 66 females qualified for analysis based on our inclusion criteria, with 29 of these falling in the naturally cycling group (M_{age} = 23.87, SD_{age} = 3.98) and 37 in the HC using group (M_{age} = 22.78, SD_{age} = 3.43).

In the naturally cycling group, 13 participants were in the follicular phase, 3 were in the ovulatory phase, and 13 were in the luteal phase. In the HC using group, 17 had an IUD, 8 used the mini-pill (progestin-only pill), and 11 were on combined birth control. Additionally, one person who recently took an emergency contraceptive (day-after-pill) was included in this group.

Among the excluded participants were naturally cycling women who had been off HCs for less than a year (n = 11), women having recently been pregnant (n = 3), women with an attentional deficit disorder (n = 9), as well as participants reporting having a hormonal disease (n = 5).

Procedure (MB, AM)
The experimental procedure was run as an online study using a combination of surveys and experiments. During recruitment, each participant was given written instructions on how to follow through from their own computer. Experiments were coded through the PsychoPy Builder and run via Pavlovia (Peirce et al., 2019). Here, participants were presented with two types of tasks, anagram solving and IQ-puzzles.

The anagram task consisted of 9 unique word scrambles, each with at least one correct solution that would spell out a Danish word. The IQ-puzzle task consisted of 6 puzzles of the short form Hagen Matrices Test, with permission granted from its originator (Heydasch et al., 2013). Examples of the tasks can be seen in appendix 2-3. For both types of tasks, accuracy and response time were recorded for each trial.

After having been exposed to the two experimental tasks, participants were redirected to a subsequent questionnaire hosted by Qualtrics (www.qualtrics.com). Among other demographic questions, participants were asked to share their history of HC use, including type (e.g. combined birth control pill, hormonal IUD, skin implant, etc.), generation, and brand name. While current users of HCs were prompted to report how long they had been on birth control, naturally cycling women were asked whether or not they had been off HCs for more than a year. Additionally, naturally cycling women were asked about their typical cycle length, the typical predictability of this, as well as how far they currently were in their cycle, allowing for rough estimates of menstrual phases.

Subsequent to participation, the true purpose of the study was revealed to participants.

Data analysis (AM, MB)

The analysis was conducted using the statistical software environment R (Core Team, 2019) and RStudio (Team 2020) with use of the packages tidyverse (Wickham et al., 2019), brms (Bürkner, 2018), rstan (Stan

Development Team, 2020), and rethinking (McElreath, 2020).

Initial Data Processing (AM, MB)

In coding the accuracy column, denoting correct and incorrect responses, we assessed the answers submitted by participants and allowed some additional correct solutions that were not anticipated prior to data collection (see appendix 4).

Approximate menstrual phases were coded by calculating a proportion representing how far along each person was in their cycle as a measure of days, then assigning the corresponding phase (follicular, ovulatory or luteal), with the ovulatory phase assumed to lie in the middle of participants' typical cycle lengths, at 45-55 percent through.

The density of the data revealed a positive skew of log-normal character for the response time outcome, and a binomial distribution for the accuracy measure, however with more density in correct answers compared to incorrect answers. A decision was made to settle on models of these families.

Across tasks, outliers in response time were found. Since the outcome of interest is perseverance, a decision was made to not exclude these outliers in the analysis, as participants could plausibly spend long durations of time on the cognitively demanding tasks – which means that an exclusion of such participants could alter the question under investigation when comparing perseverance between the two groups. However, any response time exceeding an upper limit of 10 minutes (600 seconds) was changed to this cutoff to avoid extreme data points being too influential (n = 4).

Additionally, for the sake of clarity, data was run through analysis both with and without outlier removal, which yielded no difference in the findings. The reporting here thus includes all data points, cut off at a response time of 10 minutes.

Preliminary analyses (MB, AM)

Preliminary analyses were run to assure correct modelling and inference in the main analysis. Tests were carried out both to assess potential differences in response time between women using different types of HCs, and to assess if women in different menstrual phases should be grouped together or not.

To investigate potential variation and differences within HC using women, a multilevel Bayesian model with a log-normal distribution was built, with the following syntax:

Response Time ~ 0 + HC Type
+ (1 | Participant ID)

Response time served as a continuous outcome variable. The different types of HCs reported by participants in the study was a categorical predictor. Varying effects for participants were added to account for individual variability. A wide prior was set that covered two standard deviations of the response time distribution to allow for any effect to show itself unhindered. Considering largely overlapping credibility intervals, the analysis showed no difference in response time for women using different types of HC. As such, HC users were pooled together as one group in the main investigation. This finding and choice coincides with the methodology from the Bradshaw et al. paper (2020).

For relevant checks, prior definitions, and summaries, see appendix 5. Results are reported in table 1.

Table 1

	Estimate	Est. Error	Lower CI	Upper CI
IUD	21.54s	1.14s	16.78s	27.66s
Combined birth-control	27.93s	1.19s	20.29s	38.47s
Mini-pill	29.08s	1.22s	19.69s	42.95s
Emergency contraception	19.89s	1.68s	7.46s	54.05s

Table 1: Results from the preliminary analysis of HC type

A multilevel Bayesian model relying on the same conceptual principles as the above model was applied to examine whether women should be allowed varying intercepts in the main analysis based on their menstrual phase at the time of participation. The syntax was as follows:

Response Time ~ 0 + Menstrual Phase
+ (1 | Participant ID)

Again, response time was the outcome variable, while menstrual phase was a categorical predictor. Varying intercepts per participant were added as varying effects. The setting of the beta priors was done using the same reasoning as in the above model. The analysis showed differences in response times across menstrual phases that were deemed sufficiently different to warrant implementing varying intercepts for menstrual phase in upcoming models. Results are reported in table 2.

Table 2

	Estimate	Est. Error	Lower CI	Upper CI
Follicular phase	19.89s	1.14s	15.33s	25.79s
Ovulatory phase	38.47s	1.32s	22.42s	66.69s
Luteal phase	27.66s	0.14s	21.54s	36.23s
Hormonal group	24.78s	1.08s	21.12s	29.08s

Table 2: Results from the preliminary analysis of menstrual phases

For relevant checks, prior definitions, and summaries, see appendix 5.

Response Time Analyses (AM, MB)

To investigate whether response time (as a substitute for perseverance) in both tasks would be shorter for HC using women compared to naturally cycling women ($H1_{p1}$ and $H1_{p2}$), two multilevel Bayesian regression models with a log-normal distribution were built. The models both featured the same formula, with the data being a subset of either the anagram or IQ tasks:

Response Time ~ 1 + HC use
+ Education + (1 | Menstrual Phase)
+ (1 | Menstrual Phase: Participant ID)

Once again, the continuous variable response time served as the outcome, while HC use and education were categorical predictors. Crossed varying effects were added for menstrual phase and ID to account for individual and group-level variability, as per the findings of the preliminary analyses.

All priors were set using a log-normal scale. For the anagram response time model, an informed prior with a mean of -.54 and standard deviation of 1 for the HC use beta estimate was implemented based on effect sizes from the Bradshaw and colleagues (2020) paper.

In the IQ-puzzle response time model, the prior for HC use was based on an informed prior from the GRE tests from the Bradshaw and colleagues (2020) paper. The informed mean at .21 was kept, while the standard deviation was expanded from 0.44 to 0.7. We settled on the wider error parameter since the GRE test and IQ-puzzles were deemed sufficiently different from each other that we needed to allow for a substantially different effect in the IQ-puzzle as opposed to the GRE testing.

For both models, priors for education were set with a mean parameter of 0 and a standard deviation of .5, as we did not expect effects in any specific direction. Model quality was assessed both by prior and posterior predictive checks, by inspecting chains, and through prior posterior update checks. Hypotheses were tested by assessing both evidence ratios and posterior probabilities from a hypothesis check.

For relevant checks, prior definitions and summaries, see appendix 6.

Accuracy Analyses (MB, AM)

To investigate whether performance (accuracy) in both tasks would be lower for HC using women compared to naturally cycling women ($H2_{p1}$ and $H2_{p2}$), two additional multilevel Bayesian regression models were employed. The data had a binomial distribution, so the Bernoulli family was applied for both models. Once again, the models were identical, except for which subset of data was used. The syntax was as follows:

Accuracy ~ 1 + HC use
+ Education + (1 | Menstrual Phase)

+ (1 | Menstrual Phase: Participant ID)

The binomial/categorical variable accuracy was used as the outcome, where 1 indicated a correct answer and 0 indicated a wrong answer. The predictor variables were chosen based on the same reasoning as in the response time models.

All informed priors were converted to a log-odds scale. For the anagram accuracy model, a wide, unexpecting prior was set for HC use with a mean of 0 and a standard deviation of .8 as no effect was expected due to the lack of response time findings in the original paper. The large error estimate was chosen since we wanted to allow for a possible effect to show itself, if it existed.

For the IQ-puzzle accuracy model, an informed prior was employed from the Bradshaw et al. (2020) paper, with a mean of .37 and standard deviation of 1.43.

Model quality for accuracy models was assessed identically to the response time models.

For relevant checks, prior definitions, and summaries, see appendix 7.

Full reasonings and considerations for all models can be found in appendix 8.

Results (AM, MB)

We did not find sufficient evidence in favor of $H1_{p1}$, $H1_{p2}$. The beta estimates for both response time (perseverance) models did show a miniscule effect in support of the hypotheses, but with wide and overlapping credibility intervals, the already small effect was not deemed reliable. Beta estimates are reported in table 3.

findings. Hypothesis testing of $H1_{p1}$ produced an evidence ratio of .5 and a posterior probability of .33. Hypothesis testing of $H1_{p2}$ produced an evidence ratio of 1.24 and a posterior probability of .55.

No sufficient evidence was found in favor of $H2_{p1}$ and $H2_{p2}$ either. Beta estimates for both accuracy models showed a miniscule effect that went against the hypothesis, where HC using women appeared to have a higher probability of answering correctly. Once again, model assessment produced wide and overlapping credibility intervals. Hypothesis testing of $H2_{p1}$ produced an evidence ratio of 2.19 and a posterior probability of .69. Hypothesis testing of $H2_{p2}$ produced an evidence ratio of 1.54 and a posterior probability of .61.

See table 4 for estimates and further details.

Figure 1

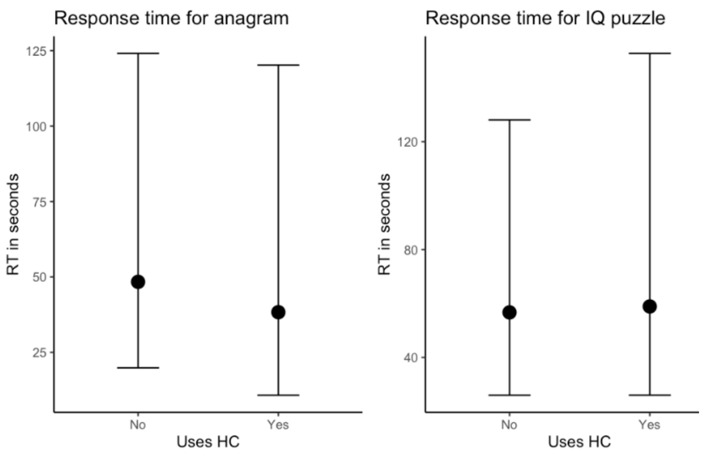


Figure 1: Conditional effects plots from the response time (perseverance) models

Table 3

	Anagram Tasks						IQ-puzzle Tasks					
	Estimate	Est. error	Lower CI	Upper CI	ER	Post. prob	Estimate	Est. error	Lower CI	Upper CI	ER	Post. prob
Intercept	22.4s	1.57s	9.21s	56.26s	-	-	42.5s	1.49s	19.5s	95.6s	-	-
HC use	17.28s	1.84s	4.62s	55.15s	0.5	0.33	45.2s	1.52s	19.89s	111.1s	1.24	0.55

Table 3: Results from the response time (perseverance) models reported in seconds

	Anagram Tasks					IQ-puzzle Tasks				
	Prob. of correct answer	Lower CI (probability)	Upper CI (probability)	ER	Post. prob	Prob. of correct answer	Lower CI (probability)	Upper CI (probability)	ER	Post. prob
Naturally Cycling Women (Mellemlang Education)	65.7%	29.3%	89.7%	-	-	62.7%	33.6%	81.9%	-	-
HC Using Women (Mellemlang Education)	70%	29.9%	91.6%	2.19	0.69	66.4%	16.9%	94.4%	1.54	0.61

Table 4: Results from the accuracy (performance) models reported in probabilities

Figure 2

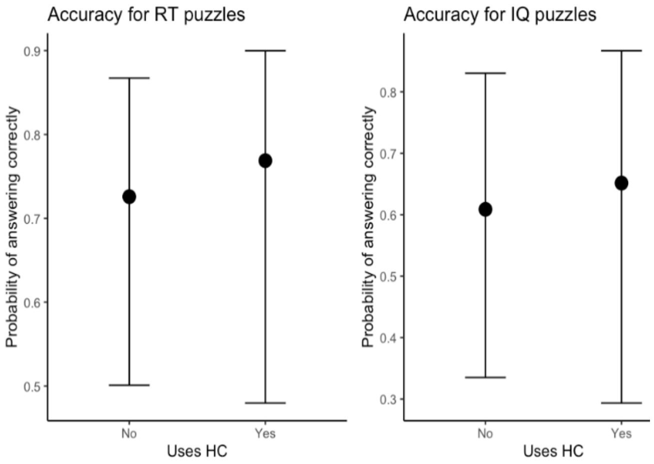


Figure 2: Conditional effects plots from the accuracy (performance) models

Power analysis (Post Hoc) (MB, AM)

A Bayesian power analysis was conducted using the effect sizes from the Bradshaw and colleagues (2020) paper. For each of the task types, we followed the approach of Solomon Kurz (2019) and simulated data with the current study's number of participants in the two experimental groups in 100 different seeds. An overview of the results is reported in table 5.

In the anagram task, simulating an effect of -.48 for the HC using women with a SD of 1 yielded a power of 45 from running each of the simulated datasets through the Bayesian formula. Thus, only 45 of the 100 simulations produced a 95% Bayesian credibility interval that did not straddle 0.

In the logic type task, simulating an effect of -.35 for the HC using women with a SD of 1 yielded a power of 24 after running each of the simulated datasets through the Bayesian formula, meaning that only 24 of the 100 simulations produced a 95% Bayesian credibility interval that did not straddle 0.

For this study to have sufficient power of approximately 80 percent, a total minimum of 70 participants would be necessary in each of the two groups for the anagram tasks. For the logic tasks to approximate 80 percent power, a total minimum of 120 participants in each condition would be necessary.

Table 5

	This Study		Goal-Standard Study	
	IQ Tasks	Anagram Tasks	IQ Tasks	Anagram Tasks
Power	24	45	79	79
Number of HC using participants	37	37	120	70
Number of naturally cycling participants	29	29	120	70
Total number of participants	66	66	240	140

Table 5: Power for different numbers of participants using a Bayesian power analysis

Discussion

Discussion of Results (MB, AM)
In this study, we tested whether HC use would predict decrements in perseverance and performance in cognitively demanding tasks, specifically word scrambles and IQ-puzzles. We found no evidence in favour of a difference between the two groups of women in any of the task types, neither for proportion of accurate answers, nor for the time spent trying to figure out the correct solution.

While beta estimates for the response time models did seem to yield small effects in favor of H1-hypotheses, the large and overlapping credibility intervals indicate that the small effects found are likely insignificant. Furthermore, seeing as non-converted credibility intervals straddle 0 and thus produce plausible effects of beta in both directions, no conclusive evidence as to the true effect of HC can be drawn. This is further supported by the evidence ratio for both H1-hypotheses providing less than moderate evidence in favor of the hypotheses, e.g. with the evidence ratio for $H1_{p1}$ falling below 1.

This lack of results might be explained by the insufficient power of the study, or they may simply be an indication of no association between decreased perseverance and HC use.

Seeing as we found no support for the H1-hypotheses, the lack of evidence supporting the H2-hypotheses was to be expected, as the two sets of hypotheses are interlinked. The evidence ratios for accuracy models also provided less than moderate evidence, while evidence ratios for the null hypothesis were markedly larger.

Had there been findings in any of the response time models, a mediated analysis of accuracy via response time would have been fitting in investigating whether response time is a mediator of performance. Future studies should seek to incorporate such a link between the two hypotheses in the case that any findings are obtained in the perseverance models, as to properly investigate the relationship between performance via perseverance.

Limitations of The Study (AM, MB)
While our research contributes to a growing body of work investigating possible cognitive implications of HC use, for better and for worse (Egan & Gleason, 2012; Griksiene & Ruksenas, 2011; Mordecai et al., 2008), and while any lack of findings have great scientific value, we must stress the urge for further investigation with appropriate sample sizes that satisfy the criteria of sufficient power. With our simulations in the Bayesian power analysis leaving only sufficient credibility intervals in 45 out of 100 cases for the expected effect size in anagram tasks, and only 24 out of 100 cases for the expected effect size in logic-type tasks, as based on the findings from the Bradshaw and colleagues paper (2020), evidence points to the need of further research drawing from a larger participant pool.

Additionally, the density distribution of our outcome measure for the accuracy performance of participants, as well as the probabilities of answering correctly as obtained by the model, revealed a tendency of answering correctly well above 60 percent for all tasks and group combinations, however with broad credibility intervals. This suggests that our choice of tasks may not have been cognitively demanding enough. Specifically, in the anagram task, the two groups of participants were left with a probability of answering the tasks correctly at 65.7 and 70 percent respectively. While the study that we set out to replicate (Bradshaw et al., 2020) used word scrambles containing 9-10 letters for both solvable and unsolvable anagrams, our number of letters in the word scrambles used ranged from 4-8 and did not contain any unsolvable tasks. Such unsolvable tasks could likely be a significant driver of perseverance, and future studies should thus consider increasing the difficulty of the word scrambles used, as well as adding problems with no solution.

Likewise, the logic type tasks in the Bradshaw and colleagues paper (2020) seems to have been balanced at a more challenging skill-level, with participants on average answering 3.21 out of 8 GRE math problems correctly in the HC using group, while on average answering 3.88 correctly in the naturally cycling group, resulting in an average performance measure of 40.13 and 48.5 percent for the two groups respectively. In our study, with the use of the logic-type IQ puzzles from the short form Hagen Matrices Test (Heydasch et al., 2013), the density distribution of the accuracy outcome suggests more correct answers than incorrect answers.

The probabilities of answering correctly as obtained from our model are 62.7 and 66.4 percent for the two groups respectively, however with broad credibility intervals.

Future studies should thus consider using more challenging logic-type tasks, possibly of a sort that does not entail multiple choice options, although there are benefits and drawbacks in both approaches: This is because the Hagen Matrices Test as used in this study has the benefit of being a validated test protocol, which furthermore makes it directly comparable to other studies that apply it.

Together, these considerations suggest that future studies should consider using parts of the Hagen Matrices Test requiring more logical reasoning than the puzzles from the short form as used in this study.

Issues in The Current Research Field (MB, AM)

As mentioned multiple times throughout this paper, research on HCs is challenged in several regards. A thorough systematic review on 22 eligible studies was conducted by Annabelle Warren in 2014. She found limited sample size to be a significant limitation of the studies included, with an overall mean of 24 HC users and many studies with 10 or fewer participants. Considering the lack of power with 70+ participants in our study, we can only imagine the inadequate power in many of these studies.

While double-blind placebo-controlled randomised controlled designs would be to prefer, they are inherently impractical, as the risk of unwanted pregnancy makes placebo-contraceptives incredibly unethical. Instead, most studies including ours employ a cross-sectional design where women are studied in two seperate groups. Crossover within participants study designs studying the same sample population off and on HCs could counterbalance this, but would require a substantial amount of both time, funding and supervision with health officials.

Furthermore, Pletzer & Kerschbaum (2014) point to the issue of “survivor effect” in research on HC, meaning that only women who continue the intake are included in most studies. Thus, women who stop using HCs due to negative experiences with topics such as emotion, learning, memory or perhaps perseverance are typically not included in analysis. Our study may have fallen victim to survivors bias as well, with 31 participants out of 37 using HCs for more than two years - of those, 24 answered that they had been using HCs for more than three years, the “highest” option available in the survey.

Overall, as Warren herself puts it, HC research suffers from, “*insufficient high-quality research”*. Additionally, she remarks that *“the diversity of findings and high risk of bias in the available studies make it difficult to draw meaningful conclusions”* (Warren, 2014, section Discussion, paragraph 1).

Future Directions (AM, MB)

We believe this specific study, as well as most methodologically challenged or weakly-powered studies, to be relevant contributions to current research, even after taking their limitations into account. Considering the magnitude of HC using women worldwide, as well as the increased public interest on the subject matter, it should seem that now is as good a time as ever to start pushing for awareness, as well as well-powered, methodologically sound and design-robust studies to further advance the knowledge in the field.

Were we to conceptually reproduce this study at a later time with the discussed limitations and considerations in mind, we imagine that it would include better and more reliable measures of both HC use and perseverance. For instance, by measuring cortisol and other hormone levels in undergraduate students before, during, and after taking a tough test, and then investigating perseverance in this setting, we believe more reliable results could be obtained. With this approach, it would be more methodologically sound to properly detect whether any association between HC use and perseverance is due to actual differences in cortisol and/or sex-hormone levels.

Conclusion (MB, AM)

This paper sought to examine how perseverance and performance are associated with the use of hormonal contraceptives. Four multilevel Bayesian models with crossed varying effects were built, two investigating perseverance across tasks and two investigating performance across tasks. None of the analyses produced sufficient evidence for either of the hypotheses. A power analysis was conducted post. hoc., revealing that the study had insufficient power at 45% and 24% for anagram and IQ-puzzle tasks, respectively. Since both response time models showed miniscule effects in favor of H1-hypotheses, although without conclusive evidence, we believe the study should be re-run utilizing better difficulty-adjusted stimuli and sufficient power in terms of larger participant pools in the future.

References

Bedsider. (2017). *All about hormones!* Bedsider. https://www.bedsider.org/features/317-all-about hormones

Bianchini, F., Verde, P., Colangeli, S., Boccia, M., Strollo, F., Guariglia, C., Bizzarro, G., & Piccardi, L. (2018). Effects of oral contraceptives and natural menstrual cycling on environmental learning. *BMC Women's Health*, *18*. https://doi.org/10.1186/s12905-018-0671-4

Bradshaw, H. K., Mengelkoch, S., & Hill, S. E. (2020). Hormonal contraceptive use predicts decreased perseverance and therefore performance on some simple and challenging cognitive tasks. *Hormones and Behavior*, *119*, 104652. https://doi.org/10.1016/j.yhbeh.2019.104652

Bürkner P (2018). "Advanced Bayesian Multilevel Modeling with the R Package brms." *The R Journal*, 10(1), 395–411. doi: 10.32614/RJ-2018-017.

Cicinelli, E., De Tommaso, M., Cianci, A., Colacurci, N., Rella, L., Loiudice, L., Cicinelli, M. V., & Livrea, P. (2011). Oral contraceptive therapy modulates hemispheric asymmetry in spatial attention. *Contraception*, *84*(6), 634–636. https://doi.org/10.1016/j.contraception.2011.03.016

Coffman, J. A. (2020). Chronic stress, physiological adaptation and developmental programming of the neuroendocrine stress system. *Future Neurology*, *15*(1), FNL39. https://doi.org/10.2217/fnl-2019-0014

Conrad, C. D. (2008). Chronic Stress-Induced Hippocampal Vulnerability: The Glucocorticoid Vulnerability Hypothesis. *Reviews in the Neurosciences*, *19*(6), 395–411.

Conrad, C. D. (2010). A critical review of chronic stress effects on spatial learning and memory. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *34*(5), 742–755. https://doi.org/10.1016/j.pnpbp.2009.11.003

Druet, A. (2017). *What is ovulation?* https://helloclue.com/articles/cycle-a-z/ovulation-101-what-is-it-how-does-it-work

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and Passion for Long-Term Goals. *Journal of Personality and Social Psychology*, *92*(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Egan, K. R., & Gleason, C. E. (2012). Longer Duration of Hormonal Contraceptive Use Predicts Better Cognitive Outcomes Later in Life. *Journal of Women's Health*, *21*(12), 1259–1266. https://doi.org/10.1089/jwh.2012.3522

Frodl, T., Schaub, A., Banac, S., Charypar, M., Jäger, M., Kümmler, P., Bottlender, R., Zetzsche, T., Born, C., Leinsinger, G., Reiser, M., Möller, H.-J., & Meisenzahl, E. M. (2006). Reduced hippocampal volume correlates with executive dysfunctioning in major depression. *Journal of Psychiatry and Neuroscience*, *31*(5), 316–325.

Gilhooly, K. J., & Fioratou, E. (2009). Executive functions in insight versus non-insight problem solving: An individual differences approach. *Thinking & Reasoning*, *15*(4), 355–376. https://doi.org/10.1080/13546780903178615

Gregory, S. T., Hall, K., Quast, T., Gatto, A., Bleck, J., Storch, E. A., & DeBate, R. (2018). Hormonal Contraception, depression, and Academic Performance among females attending college in the United States. *Psychiatry Research*, *270*, 111–116. https://doi.org/10.1016/j.psychres.2018.09.029

Griksiene, R., & Ruksenas, O. (2011). Effects of hormonal contraceptives on mental rotation and verbal fluency. *Psychoneuroendocrinology*, *36*(8), 1239–1248. https://doi.org/10.1016/j.psyneuen.2011.03.001

Hampson, E. (1990). Variations in sex-related cognitive abilities across the menstrual cycle. *Brain and Cognition*, *14*(1), 26–43.

Harlow, S. D., & Ephross, S. A. (1995). Epidemiology of menstruation and its relevance to women's health. *Epidemiologic Reviews*, *17*(2), 265–286. https://doi.org/10.1093/oxfordjournals.epirev.a036193

Hausmann, M., Slabbekoorn, D., Van Goozen, S. H., Cohen-Kettenis, P. T., & Güntürkün, O. (2000). Sex hormones affect spatial abilities during the menstrual cycle. *Behavioral Neuroscience*, *114*(6), 1245.

Hawkshead, B. (2013). *HORMONAL EFFECTS ON COGNITION IN ORAL CONTRACEPTIVE USERS AND NATURALLY-CYCLING WOMEN*. 55.

Herman, J. P., McKlveen, J. M., Ghosal, S., Kopp, B., Wulsin, A., Makinson, R., Scheimann, J., & Myers, B. (2016). Regulation of the hypothalamic-pituitary-adrenocortical stress response. *Comprehensive Physiology*, *6*(2), 603–621. https://doi.org/10.1002/cphy.c150015

Hertel, J., König, J., Homuth, G., Van der Auwera, S., Wittfeld, K., Pietzner, M., Kacprowski, T., Pfeiffer, L., Kretschmer, A., Waldenberger, M., Kastenmüller, G., Artati, A., Suhre, K., Adamski, J., Langner, S., Völker, U., Völzke, H., Nauck, M., Friedrich, N., & Grabe, H. J. (2017). Evidence for Stress-like Alterations in the HPA-Axis in Women Taking Oral Contraceptives. *Scientific Reports*, *7*(1), 14111. https://doi.org/10.1038/s41598-017-13927-7

Herzberg, S. D., Motu'apuaka, M. L., Lambert, W., Fu, R., Brady, J., & Guise, J.-M. (2017). The Effect of Menstrual Cycle and Contraceptives on ACL Injuries and Laxity: A Systematic Review and Meta-analysis. *Orthopaedic Journal of Sports Medicine*, *5*(7), 2325967117718781. https://doi.org/10.1177/2325967117718781

Heydasch, T., Haubrich, J., & Renner, K.-H. (2013). The Short Version of the Hagen Matrices Test (HMT-S): 6-Item Induction Intelligence Test. *methods, data, analyses*, *7*(2), 26. https://doi.org/10.12758/mda.2013.011

Hill, S. (2019). *This is your brain on birth control: The surprising science of women, hormones, and the law of unintended consequences*. Penguin.

Hoyt, L. T., Zeiders, K. H., Ehrlich, K. B., & Adam, E. K. (2016). Positive upshots of cortisol in everyday life. *Emotion (Washington, D.C.)*, *16*(4), 431–435. https://doi.org/10.1037/emo0000174

Kinlein, S. A., Wilson, C. D., & Karatsoreos, I. N. (2015). Dysregulated Hypothalamic–Pituitary–Adrenal Axis Function Contributes to Altered Endocrine and Neurobehavioral Responses to Acute Stress. *Frontiers in Psychiatry*, *6*. https://doi.org/10.3389/fpsyt.2015.00031

Kirschbaum, C., Kudielka, B. M., Gaab, J., Schommer, N. C., & Hellhammer, D. H. (1999). Impact of Gender, Menstrual Cycle Phase, and Oral Contraceptives on the Activity of the Hypothalamus-Pituitary-Adrenal Axis. *Psychosomatic Medicine*, *61*(2), 154–162.

Kirschbaum, C., Platte, P., Pirke, K.-M., & Hellhammer, D. (1996). Adrenocortical Activation Following Stressful Exercise: Further Evidence for Attenuated Free Cortisol Responses in Women Using Oral Contraceptives. *Stress Medicine*, *12*(3), 137–143. https://doi.org/10.1002/(SICI)1099-1700(199607)12:3<137::AID-SMI685>3.0.CO;2-C

Kurz, A. S. (2019, July 18). *Bayesian power analysis: Part I. Prepare to reject H_0 with simulation.* A. Solomon Kurz. https://solomonkurz.netlify.app/post/bayesian-power-analysis-part-i/

Maki, P. M., Rich, J. B., & Rosenbaum, R. S. (2002). Implicit memory varies across the menstrual cycle: Estrogen effects in young women. *Neuropsychologia*, *40*(5), 518–529.

McElreath, R. (2020). Statistical Rethinking: A Bayesian Course with Examples in R and Stan (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781315372495.

Mordecai, K. L., Rubin, L. H., Eatough, E., Sundermann, E., Drogos, L., Savarese, A., & Maki, P. M. (2017). Cortisol reactivity and emotional memory after psychosocial stress in oral contraceptive users. *Journal of Neuroscience Research*, *95*(1–2), 126–135. https://doi.org/10.1002/jnr.23904

Mordecai, K. L., Rubin, L. H., & Maki, P. M. (2008). Effects of menstrual cycle phase and oral contraceptive use on verbal memory. *Hormones and Behavior*, *54*(2), 286–293. https://doi.org/10.1016/j.yhbeh.2008.03.006

Nielsen, S. E., Segal, S. K., Worden, I. V., Yim, I. S., & Cahill, L. (2013). Hormonal contraception use alters stress responses and emotional memory. *Biological Psychology*, *92*(2), 257–266. https://doi.org/10.1016/j.biopsycho.2012.10.007

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pletzer, B. A., & Kerschbaum, H. H. (2014). 50 years of hormonal contraception—Time to find out, what it does to our brain. *Frontiers in Neuroscience, 8*. https://doi.org/10.3389/fnins.2014.00256

Pletzer, B., Harris, T.-A., Scheuringer, A., & Hidalgo-Lopez, E. (2019). The cycling brain: Menstrual cycle related fluctuations in hippocampal and fronto-striatal activation and connectivity during cognitive tasks. *Neuropsychopharmacology*, *44*(11), 1867–1875. https://doi.org/10.1038/s41386-019-0435-3

Ray, L. (2018). *The menstrual cycle, explained*. https://helloclue.com/articles/cycle-a-z/the-menstrual-cycle-more-than-just-the-period

Rosenberg, L., & Park, S. (2002). Verbal and spatial functions across the menstrual cycle in healthy young women. *Psychoneuroendocrinology*, *27*(7), 835–841.

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Skovlund, C. W., Mørch, L. S., Kessing, L. V., & Lidegaard, Ø. (2016). Association of Hormonal Contraception With Depression. *JAMA Psychiatry*, *73*(11), 1154–1162. https://doi.org/10.1001/jamapsychiatry.2016.2387

Stan Development Team (2020). “RStan: the R interface to Stan.” R package version 2.21.2, http://mc-stan.org/.

Ventura, M., Shute, V., & Zhao, W. (2013). The relationship between video game use and a performance-based measure of persistence. *Computers & Education*, *60*(1), 52–58. https://doi.org/10.1016/j.compedu.2012.07.003

Warren, A. M., Gurvich, C., Worsley, R., & Kulkarni, J. (2014). A systematic review of the impact of oral contraceptives on cognition. *Contraception*, *90*(2), 111–116. https://doi.org/10.1016/j.contraception.2014.03.015

Watson, S. (2018, August 17). *Stages of Menstrual Cycle: Menstruation, Ovulation, Hormones, Mor*. Healthline. https://www.healthline.com/health/womens-health/stages-of-menstrual-cycle

Wharton, W., Hirshman, E., Merritt, P., Doyle, L., Paris, S., & Gleason, C. (2008). Oral contraceptives and androgenicity: Influences on visuospatial task performance in younger individuals. *Experimental and Clinical Psychopharmacology*, *16*(2), 156–164. https://doi.org/10.1037/1064-1297.16.2.156

Maria Nissen Byg (201906706)
Andrea Dioni Munskgaard (201907480)

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686

Women are outnumbering men at a record high in universities worldwide—Study International. (n.d.). Retrieved May 25, 2021, from https://www.studyinternational.com/news/record-high-numbers-women-outnumbering-men-university-globally/

Yuen, E. Y., Liu, W., Karatsoreos, I. N., Feng, J., McEwen, B. S., & Yan, Z. (2009). Acute stress enhances glutamatergic transmission in prefrontal cortex and facilitates working memory. *Proceedings of the National Academy of Sciences*, *106*(33), 14075–14079. https://doi.org/10.1073/pnas.0906791106

Zethraeus, N., Dreber, A., Ranehill, E., Blomberg, L., Labrie, F., von Schoultz, B., Johannesson, M., & Hirschberg, A. L. (2017). A first-choice combined oral contraceptive influences general well-being in healthy women: A double-blind, randomized, placebo-controlled trial. *Fertility and Sterility*, *107*(5), 1238–1245. https://doi.org/10.1016/j.fertnstert.2017.02.120

Maria Nissen Byg (201906706)
Andrea Dioni Munskgaard (201907480)

Appendix

Appendix 1

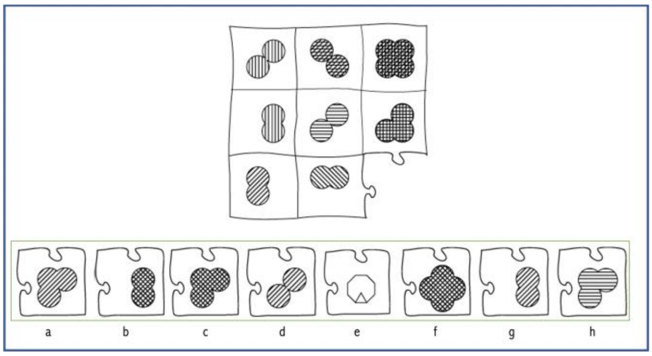
Table of differences between generations of birth control pills and the progestins used:

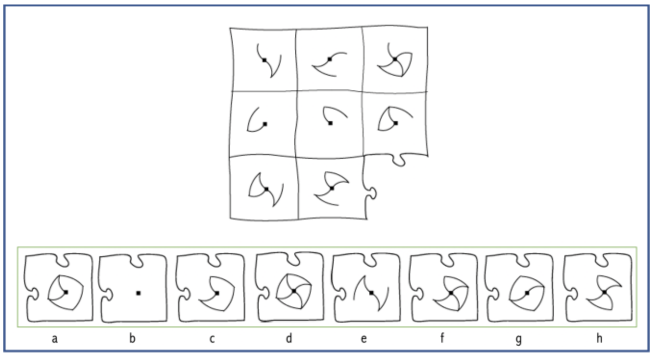
Generation	Characteristics	Currently in use in Denmark
First	Derived from testosterone. Androgenic (produce side-effects from binding to testosterone receptors), e.g. breakouts, weight gain, hair growth and possibly masculinizing effects on the brain such as decreased verbal fluency and increased performance on mental rotation tasks.	No
Second	Derived from testosterone. Androgenic (produce side-effects from binding to testosterone receptors), e.g. breakouts, weight gain, hair growth and possibly masculinizing effects on the brain such as decreased verbal fluency and increased performance on mental rotation tasks.	Yes

Third	Derived from testosterone, but molecules are modified to produce fewer masculinizing effects than the first couple of generations. However, as the testosterone-related side effects decrease, the risk of blood clots increases.	Yes
Fourth (Dienogest)	Derived from testosterone. Here, the modified chemical structure blocks the effect of testosterone in the body, meaning that the progestin becomes anti-androgenic. Associated with an increased risk of blood clots and decreased libido.	Yes
Fourth (Drospirenone)	Derived from spironolactone. The chemical structure blocks the effect of testosterone in the body more so than the other fourth generation progestin derived from testosterone. Associated with an increased risk of blood clots and decreased libido.	Yes

Appendix 2

Examples of IQ-puzzle tasks:





Appendix 3

List of words included in anagram tasks, as well as the scrambles participants received:

ANISPLEP	appelsin [orange]
LSEPJ	spejl [mirror]
STEH	hest [horse]
MKRAAE	kamera [camera]
MLOBST	blomst [flower]
DBGOOR	ordbog [dictionary]
LKDSAR	skrald [trash]
LKECYR	cykler [bikes]
PIARP	papir [paper]

Appendix 4

Accepted solutions in anagram tasks, post-data-collection:

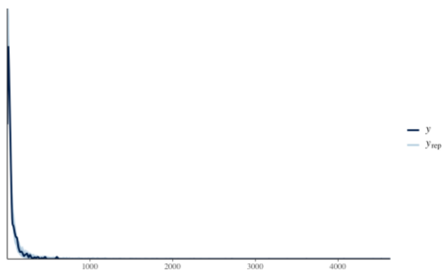
Anagram	Correct solution	Accepted solution	Reasoning
DBGOOR	ordbog [dictionary]	bogrod [disordered books/a mess of books]	Since compound nouns are common in Danish, it is not unlikely that this would be a word used in daily speech. This spelling is apparently also the name of a well-known Harry Potter character. Since it was not made clear in the instructions whether solutions could be names, we counted this as accepted.
ANISPLEP	appelsin [orange]	apelsin	This misspelling of the word was accepted, as the solution was deemed clearly correct despite the typing error.

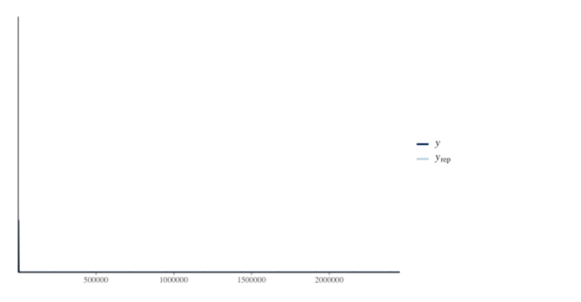
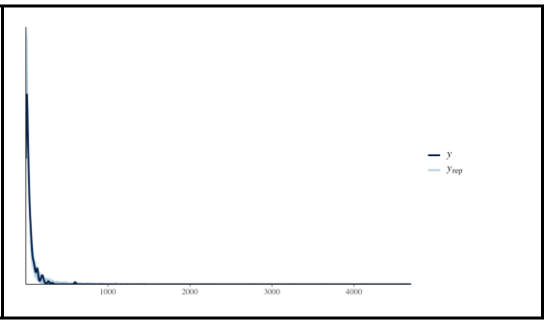
Appendix 5

Priors for HC differences and menstrual phases

HC differences	Menstrual phases
HC type: Normal(3.22, 1.2) SD: Normal(1.2, .6) Sigma: Normal(1.2, .6)	Menstrual phase: Normal(3.21, 1.2) Sigma: Normal(1.2, 0.6) SD: Normal(1.2, 0.6)

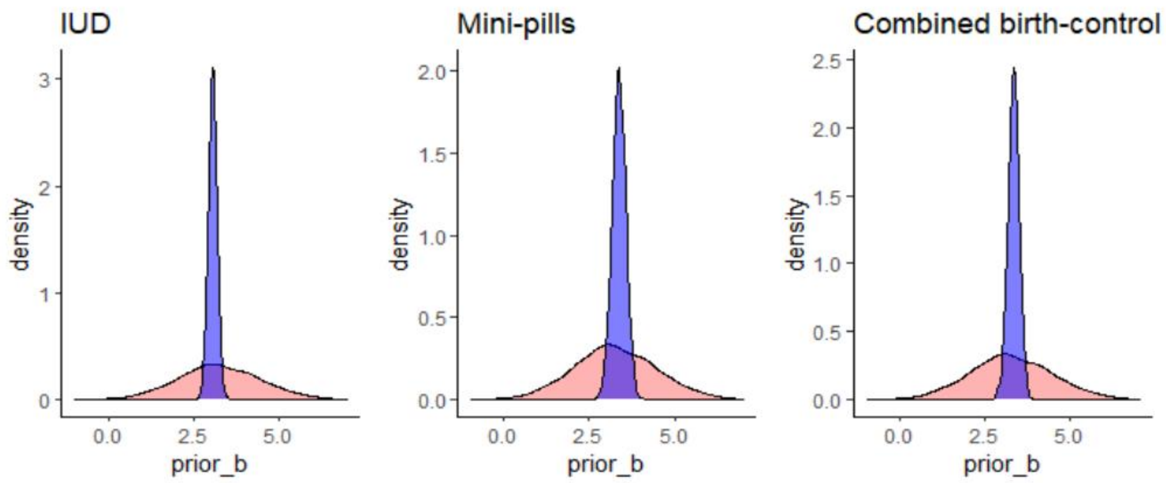
Prior and posterior predictive plots

	m0	m1
Differences between HC type		

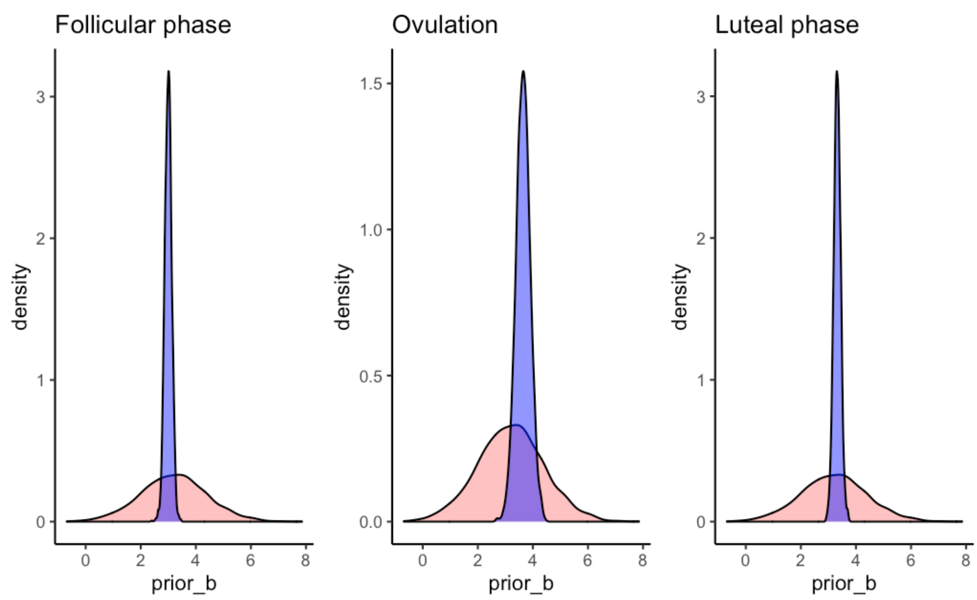
Differences in menstrual phase		

Prior posterior update checks

HC-types



Menstrual phases

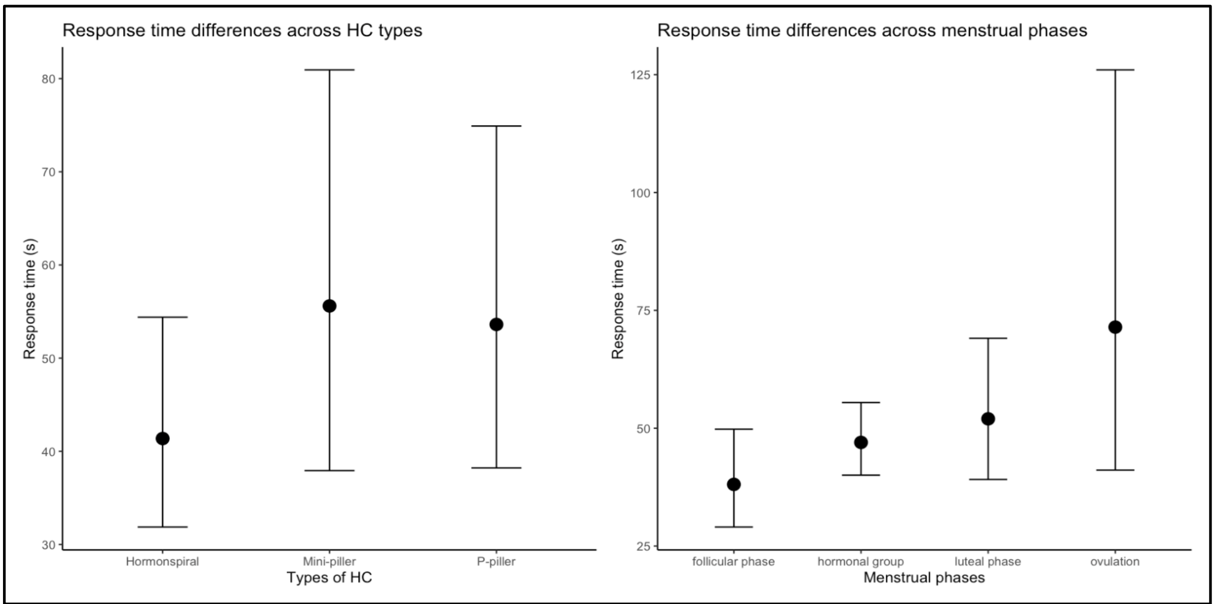


Full summary

Summary for HC-type differences				
	Estimate	Estimated error	Lower CI	Upper CI
IUD	21.54s	1.14s	16.78s	27.66s
Combined birth-control	27.93s	1.19s	20.29s	38.47s
Mini-pill	29.08s	1.22s	19.69s	42.95s
Emergency contraception	19.89s	1.68s	7.46s	54.05s

Summary for menstrual phase differences				
	Estimate	Estimated error	Lower CI	Upper CI
Follicular phase	19.89s	1.14s	15.33s	25.79s
Ovulatory phase	38.47s	1.32s	22.42s	66.69s
Luteal phase	27.66s	0.14s	21.54s	36.23s
Hormonal group	24.78s	1.08s	21.12s	29.08s

Conditional effects plots

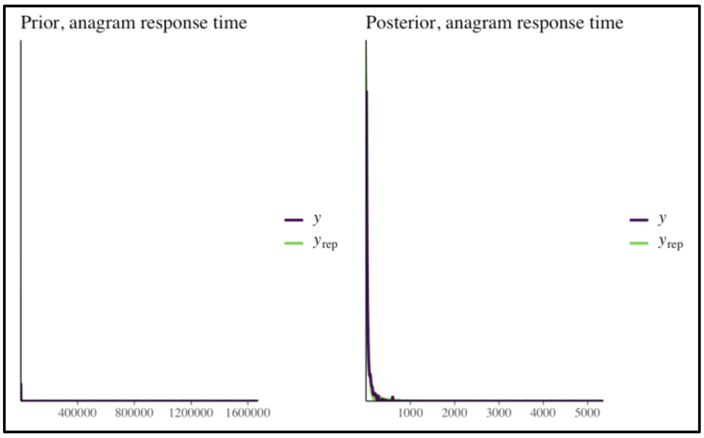


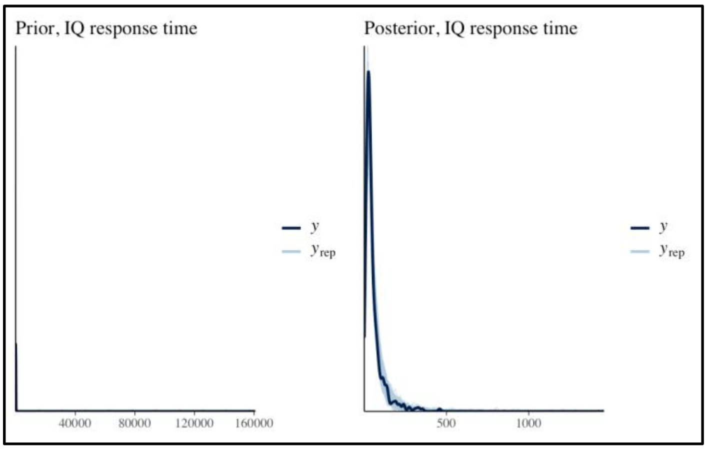
Appendix 6

Priors for anagram and IQ response time models

Anagram	IQ
Intercept: Normal(2.95, 1) Use of HC: Normal(-0.54, 1) Education level: Normal(0, .5) SD: Normal(1, .5) Sigma: Normal(1, .5)	Intercept: Normal(3.61, 0.87) Use of HC: Normal(0.21, 0.7) Education level: Normal(0, .5) SD: Normal(1, .5) Sigma: Normal(1, .5)

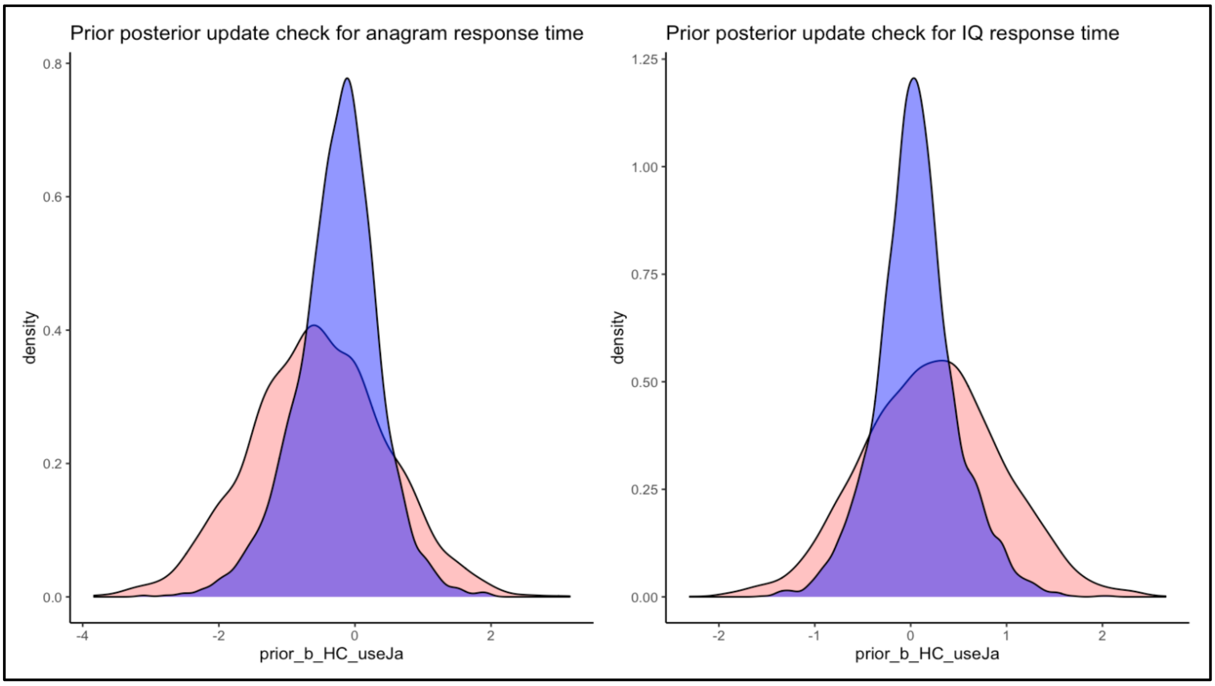
Predictive prior and posterior checks for response time models



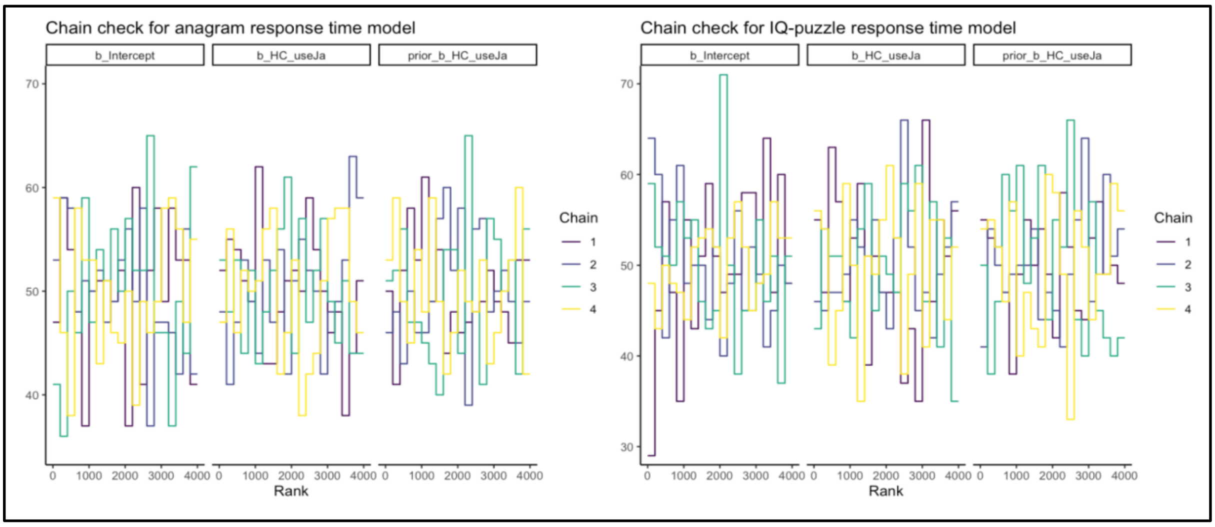


Maria Nissen Byg (201906706)
Andrea Dioni Munskgaard (201907480)

Prior posterior update checks for response time models



Chain checks for response time models



Full summaries on a log scale

Anagram

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	3.11	0.45	2.22	4.03	1.00	1935	2351
HC_useJa	-0.26	0.61	-1.58	0.90	1.00	2356	2423
EducationGymnasieluddannelse	0.11	0.26	-0.41	0.62	1.00	1865	2541
EducationKort	-0.14	0.28	-0.70	0.39	1.00	1774	2451
EducationLang	-0.03	0.31	-0.65	0.56	1.00	2217	2316
EducationMellemlang	-0.04	0.25	-0.53	0.44	1.00	1699	2532

IQ-puzzles

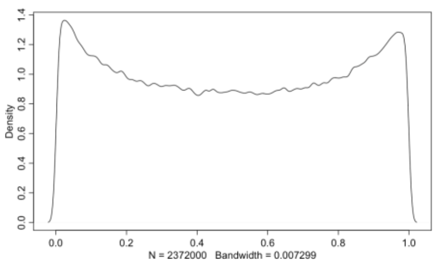
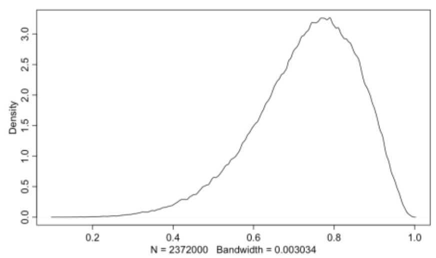
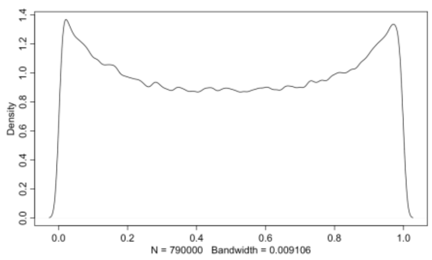
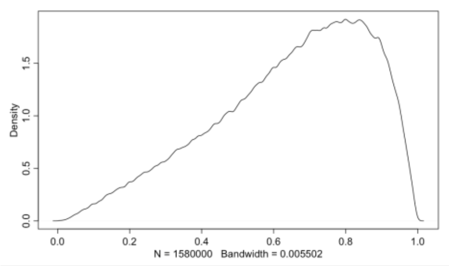
```
Population-Level Effects:
                             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept                        3.75      0.40     2.97     4.56 1.00     1670     2222
HC_useJa                         0.06      0.42    -0.76     0.96 1.00     1986     2222
EducationGymnasieluddannelse     0.04      0.26    -0.49     0.54 1.00     1361     1960
EducationKort                   -0.31      0.28    -0.85     0.23 1.00     1524     2333
EducationLang                   -0.05      0.31    -0.65     0.55 1.00     1772     2542
EducationMellemlang             -0.13      0.25    -0.63     0.34 1.00     1476     2046
```

Appendix 7

Priors for anagram and IQ accuracy models

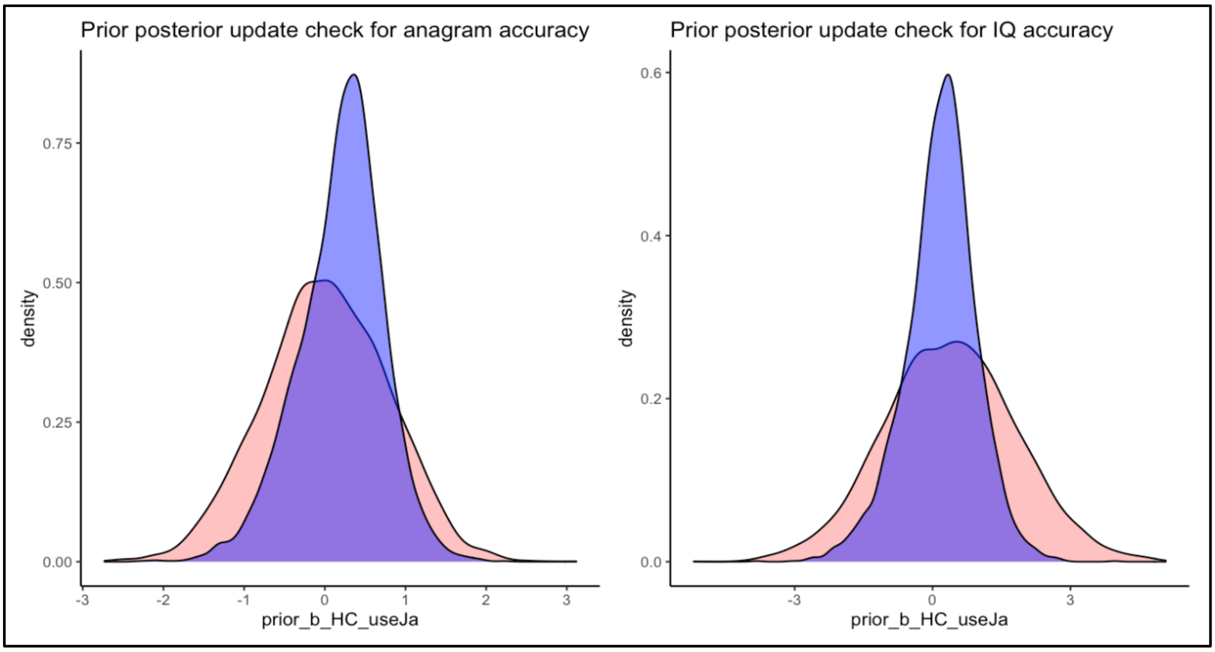
Anagram	IQ
Intercept: Normal(0, 1) Use of HC: Normal(0, .8) Education level: Normal(0, .5) SD: Normal(1, .5)	Intercept: Normal(0, .8) Use of HC: Normal(.37, 1.43) Education level: Normal(0, .5) Menstrual phase: Normal(0, .5) SD: Normal(1, .5)

Lin-pred checks for accuracy models

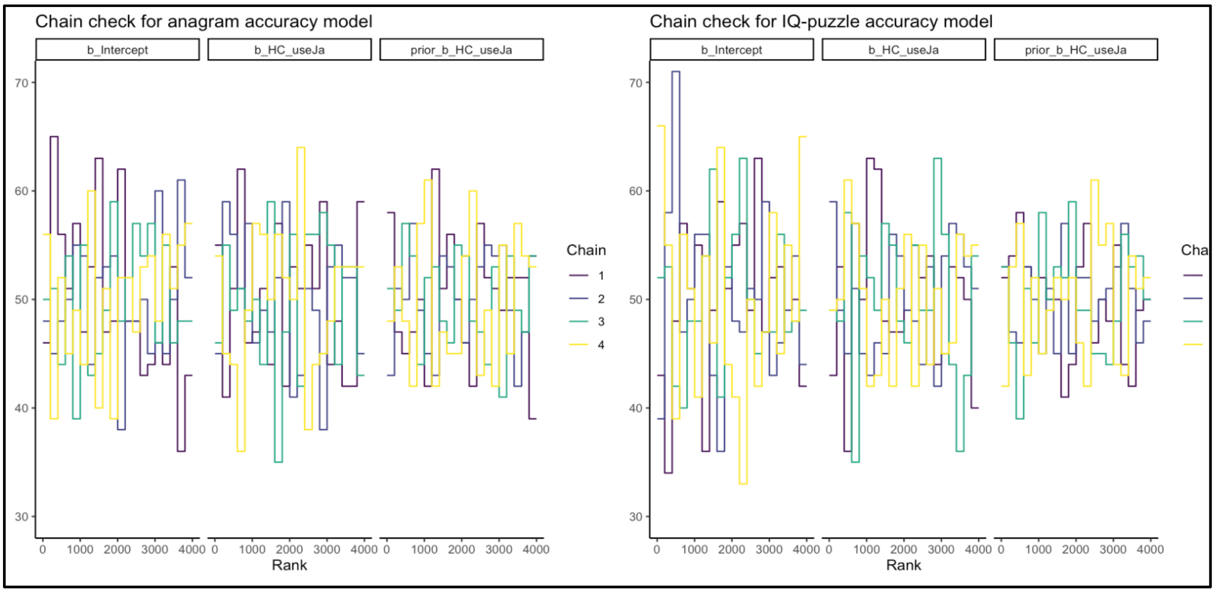
	m0-models	m1-models
accuracy ~ anagram group (open prior)		
accuracy ~ IQ group (informed prior)		

Maria Nissen Byg (201906706)
Andrea Dioni Munskgaard (201907480)

Prior posterior update checks for accuracy models



Chain checks for accuracy models



Full summaries for anagram and IQ-puzzle accuracy models on a log-odds scale

Anagram accuracy

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	3.11	0.45	2.22	4.03	1.00	1935	2351
HC_useJa	-0.26	0.61	-1.58	0.90	1.00	2356	2423
EducationGymnasieluddannelse	0.11	0.26	-0.41	0.62	1.00	1865	2541
EducationKort	-0.14	0.28	-0.70	0.39	1.00	1774	2451
EducationLang	-0.03	0.31	-0.65	0.56	1.00	2217	2316
EducationMellemlang	-0.04	0.25	-0.53	0.44	1.00	1699	2532

Maria Nissen Byg (201906706)
Andrea Dioni Munskgaard (201907480)

IQ-puzzles accuracy

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	0.45	0.56	-0.68	1.51	1.00	2910	2901
HC_useJa	0.16	0.78	-1.44	1.65	1.00	2694	2255
EducationGymnasieluddannelse	0.36	0.38	-0.37	1.10	1.00	3061	3102
EducationKort	-0.50	0.39	-1.25	0.25	1.00	3501	2774
EducationLang	0.32	0.42	-0.53	1.12	1.00	4422	2907
EducationMellemlang	0.07	0.34	-0.60	0.72	1.00	3105	3003

Appendix 8

Full reasoning and explanation of the analyses can be found here:

https://docs.google.com/document/d/1wKFYLOlIYP0dHVxcY5NVvGvAszcms62CRAH3yG2hIps/edit?usp=sharing