

Acute Effects of Fasting on Cognitive Performance: A Systematic Review and Meta-Analysis

Participant-visible method context. Results, discussion, captions, and the target figure are excluded.

Literature Search Strategy We searched the following online databases for relevant literature: SCOPUS, Web of Science (see the Supplemental Material for a complete list of the indices included), Embase, ProQuest, APA PsycInfo, PubMed, and CINAHL. This search covered articles from 1923 to February 15, 2025. In addition, the first 700 relevant citations from Google Scholar were screened. To identify gray literature, we searched ProQuest for unpublished theses and dissertations and reviewed preprints hosted on the Open Science Framework, PubMed preprints, ResearchSquare, SSRN, and preprint.org. One preprint that met inclusion criteria was retained. Search terms were organized into two conceptual categories: “fasting” and “cognition.” To capture lexical variability and interdisciplinary perspectives (e.g., from nutritional science, medicine, and psychology), we used a wide range of synonyms and related terms (e.g., “fasting” or “hunger” for fasting; “cogniti” or “mental” for cognition). Commonly used task names (e.g., “Stroop

task”) were also included (see Table 1 for the complete list of terms). Terms within each category were combined using Boolean OR operators, and the two categories were connected using AND. In addition to a structured keyword search, we used several supplementary strategies to maximize coverage. We screened the reference lists of all included articles and examined the first 10 citations of each article on Google Scholar. To further identify relevant unpublished or in-progress work, we also announced the search on the European Health Psychology Society’s social media platform. When included studies lacked necessary data, we contacted authors directly for clarification or additional information.

Study Selection Duplicate records were automatically removed using the Systematic Review Accelerator Deduplicator (Clark et al., 2020). Eligible studies compared cognitive performance between fasted and satiated conditions using standardized cognitive tasks. Studies included both between-subjects and within-subjects designs (e.g., crossover or repeated-measures), involving healthy participants. We did not impose restrictions based on publication year or geographic location. Studies were excluded if they involved clinical or patient groups (including participants with a body mass index [BMI] > 30), relied on artificial meals (e.g., protein shakes) as the satiation condition, or lacked a clearly defined fasting protocol—such as studies combining fasting with physical exercise. We also excluded studies on religious fasting (e.g., Ramadan), due to the unique confounds discussed elsewhere in the article. Poor methodological quality was not an exclusion criterion and was instead evaluated as a moderator. Screening was conducted based on titles and abstracts, with full texts and supplementary materials reviewed when eligibility was unclear. Abstracts in languages other than English were translated using Google Translate. Of 38 non-English studies, two met inclusion criteria, but neither provided extractable effect size data. Attempts to contact the corresponding authors were unsuccessful. Reviewers were not blinded to author identities or institutional affiliations.

Data Extraction From each included study, we extracted all available accuracy and response time measures from cognitive tasks. Where possible, we also extracted variables used in moderator analyses, including fasting duration, cognitive domain, experimental design, mean participant age, proportion of female participants, baseline BMI, use of food-related stimuli, and sample size (see the Supplemental Material for studies that met inclusion criteria but for which necessary information could not be extracted). If only a range of fasting durations was reported, the mean was used; if only a minimum was stated, that value was used. Similarly, for time-of-day data, we used the midpoint of the reported range. Our preregistration included plans to examine the time between breaking the fast and cognitive testing, but none of the included studies reported this information. Only four studies employed multiday fasting protocols (i.e., intermittent fasting), which was insufficient to evaluate the number of fasting cycles as a moderator. Blood glucose and mood were also omitted as moderators due to low reporting frequency (10 and 11 studies, respectively).

FASTING AND COGNITIVE PERFORMANCE

Table 1 Keywords Used in Systematic Search Fasting

Cognition

fasting “intermittent fasting” “alternate-day fasting”/“alternate day fasting” “periodic fasting” “time restricted feeding” “time restricted diet” “time restricted fasting” “time restricted eating” “food tim*” (for food timing, food time) “food abstinence” “food restriction” “food deprivation” “calori*” restriction” (for calorie restriction, caloric restriction) “Intermittent calori*” restriction” “energy deprivation” “continuous energy restriction” “intermittent energy restriction” “intermittent restrictive diet” “extreme diet” hunger/hungry hypoglycemia starvation

cogniti* (for cognitive, cognition) mental “executive funct*” (for executive function) “executive control” attention vigilance “attentional control” “inhibitory control” memory “working memory” recall learning processing perception planning reasoning “spatial ability” “decision making” psychomotor “problem solving” intelligence expertise distraction fatigue “psychological task” “psychological test” “behavioral task”/“behavioral task” “behavioral test”/“behavioral test” performance “reaction time” Stroop

Moderator coding involved a degree of subjectivity, which we addressed through double coding by two independent raters. The first coder was the lead author; the second was a trained postgraduate assistant. We calculated Cohen’s κ for categorical variables and intraclass correlation coefficients (ICC) for continuous ones. Reliability was acceptable to high for all key moderators: cognitive domain ($\kappa = .78$), fasting duration (ICC = .94), experimental design ($\kappa = .93$), and use of pictures of food as stimuli ($\kappa = .76$). ICC values above .9 can be interpreted as excellent reliability (Koo & Li, 2016); κ values above .61 indicate substantial agreement (Landis & Koch, 1977).

Risk of Bias Risk of bias was rated as “low,” “medium,” or “high” using the Cochrane Collaboration’s guidelines for both between- and withinsubjects designs (Higgins et al., 2011). Criteria included randomization procedures, adherence to intervention, carryover effects (in crossover trials), missing data, measurement quality, and selective reporting. Each study was independently rated by the first author and a second rater not otherwise involved in the project. When multiple studies appeared in one article, they were assessed individually. Agreement between raters was substantial: 36 ratings matched exactly, and 35 differed by only one level (e.g., “medium” vs. “high”). In all such cases, the more conservative (i.e., lower quality) rating was used in subsequent analyses. One case showed a

discrepancy of more than one level (“low” vs. “high”) and was resolved through discussion. Full individual ratings are reported in Tables S4 and S5.

Analysis We used Hedges’s g to calculate standardized mean differences (SMDs) between fasted and satiated groups or conditions, correcting for small sample size bias (Hedges, 1981). The primary outcome of this meta-analysis is the pooled SMD across studies, reflecting performance differences at a single time point. Positive SMDs indicate higher performance by satiated participants; negative SMDs represent higher performance by fasted participants. Reported standard deviations were used to compute each effect size. Following this, we conducted moderator analyses and sensitivity checks. We fitted Bayesian hierarchical models to the data. These types of models offer several advantages: First, they do not require specifying unknown covariances between effect sizes; second, they account for dependency among effect sizes from the same study (Cheung, 2019); and third, they allow for a probabilistic quantification of uncertainty around the pooled estimate (Voorre, 2016; Williams et al., 2018). Effect sizes were included at the first (lowest) level of the model, each assigned its own intercept. This approach accommodates the unknown correlations between multiple measures reported within

BAMBERG AND MOREAU

the same study, such as response time and error rate, which cannot be estimated directly from summary statistics. At the second level, effect sizes were nested within studies, allowing for pooled estimates at the study level. The highest level of the model was then aggregated across studies to yield the overall meta-analytic estimate. Estimates were weighted by their standard error, giving more influence to studies with larger sample sizes. We specified weakly informative priors for the overall intercept (α , normal distribution with $M = 0$, $SD = 1$), standard deviation (σ , half-Cauchy distribution with scale = 0.5), and metaregression coefficients when applicable (β , normal distribution with $M = 0$, $SD = 10$). Sensitivity analyses for prior specifications are reported in the Supplemental Material. Models were implemented in R (v4.0.2) using the brms package (Bürkner, 2017). Posterior distributions were sampled using the default No-U-Turn Sampler across four chains with 4,000 iterations each (2,000 warm-up). As measures of heterogeneity, we report the estimated between-study and between-estimate standard deviations (τ) from the Bayesian model. We also report Cochrane’s Q and I^2 based on a frequentist version of the hierarchical model to align with conventional meta-analytic reporting standards (Cheung, 2019). Note that we report 95%

credible intervals (95% CrI) of highest posterior density, as opposed to frequentist 95% confidence intervals; the number of effective samples for the posterior median point estimate (Bulk Effective Sample Size [ESS]) and 95% CrI (Tail ESS) as well as R diagnostics, which indicates how well the four chains mixed. R values between 1.00 and 1.02 were deemed acceptable (Brooks & Gelman, 1998).

Galioto & Spitznagel, 2016), and (c) the most frequently assessed domains in the included studies. This categorization allowed us to test whether fasting effects vary systematically across fundamental cognitive systems while maintaining compatibility with existing literature. Several studies used pictures of food as stimuli in their experiments. We distinguished between four types of studies for the subgroup analysis: first, studies that did not use any food-related stimuli (no food); second, studies that included both food-related and neutral stimuli (e.g., landscape pictures), for which we extracted effect sizes only for the neutral stimuli (neutral control); third, studies that used both types of stimuli but reported effect sizes jointly for both categories (mixed); and finally, studies that used only food-related stimuli (food only). In addition, we fitted the main model separately for estimates from no food or neutral control stimuli and mixed and food only stimuli. As a follow-up, we calculated a multiple regression with an interaction between cognitive domain and food stimuli. Studies using between-subjects, crossover, and repeated-measures designs were included in the meta-analysis. We tested the role of these differences in study design in a subgroup analysis. The study populations differed in age and included underage participants. To test whether the difference between conditions or groups was associated with participants' age, we entered the mean age of a sample as a predictor in a metaregression model. Last, we assessed whether participants' BMI, measured at the start of the study, had an influence. Other moderator analyses (including publication year and sex) are reported in the Supplemental Material.

Moderator Analyses

Sensitivity Analyses

We tested whether the difference between fasted and satiated conditions can be explained by a group of categorical moderators in subgroup analyses as well as continuous moderators in metaregressions, as preregistered. To test whether a linear dose–response relation between duration of the fast and the difference in performance in the fasted and satiated state exists, we performed a metaregression with fasting duration as the independent variable. We tested a linear relation since it is the most parsimonious assumption without prior arguments for a more complex functional relation. Notwithstanding, we reported the results of a quadratic model in the Supplemental Material. Few fasting durations were longer than 48 hr (see Figure S1 for a histogram). We also reported the regression after removing these outliers to better represent the majority of fasting durations. In addition, we performed a metaregression with time of day as a predictor. We separated the effect sizes depending on what cognitive domain the used task mainly tests to analyze differences between cognitive domains. For example, effect sizes from Go/No-Go and Eriksen flanker tasks are both in the inhibitory control subgroup. The categorization of cognitive domains (attention, cognitive flexibility, general executive function, inhibitory control, long-term memory, perceptual accuracy, and working memory) was based on the following: (a) established neuropsychological frameworks of cognitive architecture (see, e.g., the unity/diversity model of executive functions by Miyake et al., 2000), (b) previous meta-analyses examining fasting or nutritional interventions (Adolphus et al., 2016;

To measure the out-of-sample predictive performance of our main model, we performed approximate leave-one-out crossvalidation with Pareto-smoothed importance sampling (Vehtari et al., 2017). Since the Pareto-k diagnostic values were not satisfactory (Vehtari et al., 2024), we also performed leave-one-groupout calculations by, one after the other, removing each study from the data and estimating the pooled effect. Moreover, we refit our main model with different prior specifications to assess the influence of the chosen priors on the posterior estimates (see the Supplemental Material).

Metabiases We report several bias measures based on frequentist concepts such as statistical significance. We used a trim-and-fill approach to test for asymmetries in the effect sizes (Duval & Tweedie, 2000; Rücker et al., 2011) and the precision effect test–precision effect estimate with standard error (PET-PEESE) method for reporting publication bias as it provides a test of asymmetry with little bias compared to other meta-analytic methods (Stanley & Doucouliagos, 2014). When there is evidence of a true effect, the PEESE is thought to be a more accurate measure; in the absence of evidence of a true effect, the PET is presumed more adequate. Results were reported accordingly (Stanley & Doucouliagos, 2014). Funnel plots showing estimate precision (inverse of the standard error) against effect sizes were used to visualize and evaluate the trim-and-fill procedure.

FASTING AND COGNITIVE PERFORMANCE

Furthermore, a p curve analysis can be found in the Supplemental Material (Simonsohn et al., 2014).

Transparency and Openness This meta-analysis follows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines according to the PRISMA-P checklist and was preregistered at PROSPERO with the ID CRD42021272822 (Page et al., 2021). Deviations from the registered protocol are reported in Table S1 and the study repository (<https://doi.org/10.17605/OSF.IO/NB5MJ>). The data are available in the study repository as well (Bamberg & Moreau, 2021). The statistical analysis was performed in R primarily using the brms package (Bürkner, 2017); the analysis script was based on the publicly available template provided in Moreau and Gamble (2022) and is available in the study repository.