FACULTY OF PSYCHOLOGY AND EDUCATIONAL SCIENCES

Department of Experimental Psychology

MALADAPTIVE PERSEVERATION: THE IMPACT OF IRREFUTABLE BELIEF SYSTEMS ON INSIGHTFUL REASONING

Student number: 02109759

Supervisor: Prof. Dr. Marc Brysbaert

Research proposal for Research Project Experimental Psychology
Lecturer: Prof. Dr. Louisa Bogaerts

Academic year: 2024 - 2025

GHENT UNIVERSITY

Abstract

Belief systems that resist contradictory evidence can make individuals less open to changing their perspective, possibly affecting their ability to effectively navigate complex problem domains. To explore this, 60 participants took part in a 55-minute experiment where they temporarily adopted either refutable or irrefutable reasoning styles prior to insight tasks. After exposure to rigid cognitive frameworks, participants had more difficulty with insightful reasoning where representational change was a prerequisite for finding the solution, as indicated by an overall increase in reaction time. A second analysis found that participants who were more drawn to irrefutable belief systems had an overall poorer performance on the insight tasks. These findings suggest that frequent engagement with such beliefs may hinder the mental flexibility needed to adaptively restructure ideas. Given the importance of repetitive negative thinking across many psychopathologies, future research could potentially benefit from incorporating the degree of falsifiability as a novel transdiagnostic variable into contemporary etiological models.

Keywords: irrefutable reasoning, constraint relaxation, maladaptive perseveration

Introduction

Maladaptive Perseveration

Theoretical models have long delineated neurocognitive components, such as bottom-up limbic hyperactivity and top-down attenuated cognitive control, as key mechanisms underlying repetitive negative thoughts (Disner et al., 2011; Koster et al., 2011; Hallion et al., 2022). Beyond these dysfunction-based accounts, cognitive perseveration has also been linked to motivated reasoning (Kunda, 1990; Caddick & Feist, 2021), where individuals are incentivized to process information in a biased manner that supports their pre-existing beliefs or desires. Relatedly, confirmation bias—the tendency to predominantly engage with information that confirms one's existing beliefs—has been shown to create self-reinforcing loops by amplifying congruent evidence and discounting disconfirming data (Klayman, 1995; Modgil et al., 2021). Brosschot et al. (2006) demonstrated that cognitive perseveration in negative mental states can mediate the development of chronic pathogenic conditions by inducing a continuous psychophysiological burden. Therefore, it remains vital to consider novel predictors when creating holistic models with incremental validity compared to traditional etiological models.

Since Karl Popper revolutionized the field of rational inquiry by introducing the concept of falsifiability as the single most important demarcation criterion for science, no experimental research has yet been conducted that investigates whether unfalsifiability can moderate cognitive perseveration. Given that some belief systems can be set within parameters that preclude any disconfirmation (i.e., unfalsifiable), those beliefs could become impervious upon presentation of contradictory evidence, as this information would be rendered epistemically irrelevant to preserve internal coherence (Popper, 1934; Boudry & Braeckman, 2010). Consequently, cognitive perseveration may occur, not solely from a dysfunctional neurocognitive system, motivated reasoning, or confirmation bias, but as an inherent feature of the axiomatic framework on which the belief system rests.

De Facto Unfalsifiability

Boudry & Braeckman (2010) discussed two recurring types of irrefutable reasoning that may achieve invulnerability against rational criticism and empirical disconfirmation: a) An epistemic defense mechanism is a structural (i.e., proposition-inherent) feature of the belief system itself that allows for the reframing of disconfirming accounts to ensure consistency within the belief system. b) An immunizing strategy refers to a flexible (i.e., proposition-independent) feature that is applied externally to the belief system, allowing disconfirming accounts to be dismissed, discredited, or reinterpreted to preserve the belief's validity.

Some examples of modern belief systems that are often criticized for being too reliant on irrefutable reasoning are mono- and polytheistic doctrines (Dawkins, 2006), various postmodernist branches (Knight, 2020; Pluckrose & Lindsay, 2021), Freudian psychoanalytic frameworks (Boudry, 2007), alternative medicinal practices (Ernst & Singhs, 2008), conspiracy beliefs (Boudry, 2022), Scientology (Kent, 1999), and simulation hypothesis (Wolpert, 2024).

In view of their established role in shaping contemporary epistemological frameworks, introducing this construct of unfalsifiability into an experimental set-up may offer deeper insight into how some individuals get stuck in their own thought bubble. To make the design more feasible, a more practical variant of irrefutability is introduced: "A belief system, as a proposition, is 'de facto' unfalsifiable (DFU) whenever there are proposition-inherent or proposition-independent epistemic features that effectively decrease its probability to be refuted (Boudry, 2007)".

Higher-order Priming Paradigm

To operationalize the impact of DFU belief systems on cognitive perseveration in an experimental setting, a novel priming paradigm will be employed where, after temporarily adopting DFU belief systems, changes in insightful reasoning abilities will be measured. Priming refers to the process by which exposure to a stimulus influences the processing of and/or response to a subsequent stimulus. This primarily occurs through the retention of neural activation, making

related information or cognitive strategies more likely to influence the subsequent response pathway (Horner & Henson, 2008; Dai et al., 2023; Hao et al., 2024).

Beyond priming that facilitates posterior perceptual processing, an interest in higher-order priming effects has been growing. These can occur in a wide array of cognitively complex domains, such as moral judgement (Cameron et al., 2017), belief in free will or determinism (Genschow et al., 2022), cooperation (Drouvelis et al., 2015), risk perception (Wadhera & Kakkar, 2020), emotion regulation style (Dewitte, 2011), pro-environmental attitudes (Kim et al., 2020), among others. An experimental approach using unfalsifiable reading material has only been conducted once by Friesen et al. (2015). They successfully demonstrated that prior exposure to unfalsifiable reasoning patterns can increase religious conviction and steepen the ideology polarization slope.

Insightful Reasoning

Insight is frequently conceptualized as a transformative process where an individual abruptly transitions from a state of not knowing how to solve a problem to knowing how to solve a problem (Gilhooly & Murphy, 2005; Sheth et al., 2009; Stuyck et al., 2021). Insightful reasoning, according to Representational Change Theory (Knoblich et al., 1999; Osuna-Mascaro & Auersperg, 2021), involves two cognitive restructuring mechanisms: a) Constraint relaxation entails inhibiting previously used successful problem-solving strategies that allow for an exploration of new feature dimensions to be incorporated into the current problem representation. b) During chunk decomposition, perceptual units are broken down, allowing new configurations of chunks to emerge. If those two types of restructuring do not occur, an impasse period is reached in which no further progress can be made.

The link between insightful reasoning and DFU reasoning lies in the importance of re-evaluating the validity of the current representation. That is, if an individual habitually disregards performance failure (as indicated by contrary evidence) as a signal to reconsider the validity of their belief-based representation, then the necessary cognitive trigger for restructuring may

become desensitized. As a result, this induced desensitization may generalize to their overall problem-solving strategies that could be reflected by a prolonged impasse period during insightful reasoning for which constraint relaxation is crucial for finding the solution. In this way, DFU reasoning may start to suppress the adaptive flexibility required for representational change.

Methods and Materials

Design and Procedure

60 natively Dutch psychology students from Ghent University were recruited for a 55-minute Psychopy experiment for which they could obtain a credit to complete a course. To avoid uncontrolled differences between participants and preserve sufficient statistical power, a within-subject design was employed. The experiment consisted of 24 trials, partially randomized into six blocks, where each block contained four trials of the same condition type. All trials comprised three distinct elements (fig. 1):

1) The first element was a natural reading paradigm, where participants were instructed to temporarily simulate how ‘person A’ would respond to critique by ‘person B’: The participants were given 90 seconds to read an in-first-person text of a cognitive schema (Reinecke et al., 2013) of a fictional character (person A) and a critique of that cognitive schema by an antagonist (person B). In that same time period, they had to answer a multiple-choice question with three answer options containing three types of possible reactions by ‘person A’ to the critique of ‘person B’. The correct answer is always the one with the strongest alignment to how ‘person A’ would respond.

An example that was used of a DFU cognitive schema is the widespread conspiracy belief that the very absence of evidence for elitist corruption actually confirms its presence. This would be based on the assumption that powerful, elitist groups would simply use their resources to hide any incriminating proof (Mehl et al., 2025). This inverted logic, which tampers with conventional frameworks for evidence-based belief updating, could trap someone in a thought loop. In contrast, an example that was used of a refutable cognitive schema concerned the effectiveness of a fictional

mRNA vaccine in slowing the spread of a viral disease. It was formulated so that the claim makes a clear, testable prediction that can be empirically supported or disproven.

Trials for which no correct response was given on the multiple-choice question were assumed to reflect a failure to temporarily adopt the respective cognitive schema. Overall performance on the multiple choice-questions was high enough such that the priming effect could be tested for the majority of trials (M=80.11%; SD=18.27%). A logistic mixed-effects model for accuracy showed that the multiple-choice questions for DFU blocks were slightly easier (Δ=14.12%; p=0.00426). The reaction times did not significantly differ, as estimated using a linear mixed-effects model (Δ=1.76s; p=0.322). Random intercepts and slopes were included for participants, and random intercepts for task items.

Several steps were undertaken to ensure that the only systematic difference between conditions could arise from the independent variable (i.e., manipulation of falsifiability). Obinwanne & Brandtner (2024) illustrated that large language models (LLM) can substantially outperform traditional machine learning methods to extract complex features from texts. Therefore, a blind LLM-based validation (model='o1') was used to evaluate four potential confounders and the manipulation of interest on a 7-point Likert scale. In total, 30 in-first-person texts of three to five sentences were evaluated on valence, arousal, reading complexity, political neutrality and the degree of falsifiability (Dreisbach & Fischer, 2012; Shenhav et al., 2013).

The 24 most suitable texts were selected (i.e., 12 for each condition type) based on these ratings, such that the only significant difference between DFU and F texts was found in the degree of falsifiability (d=-5.83; p=1.797e-12; see supplementary materials). The four potential confounders showed small and non-significant effect sizes: valence (d=-0.193; p=0.640), arousal (d=0.227; p=0.584), reading complexity (d=0.120; p=0.771), and political neutrality (d=-0.082; p=0.843). A visual analog was displayed for each text to indicate the remaining time to answer the multiple-choice question.

2) Each of the 24 priming natural reading components was followed by an insight task. Increasing the number of trials would be beneficial for the statistical power, but with the inherently lengthy nature of insightful reasoning, this would be challenging. Therefore, classical insight tasks, in combination with a modern approach (i.e., remote associates' tasks), are deemed the most viable option to achieve an optimal balance between internal validity and sufficient statistical power. After each multiple-choice question answer, the participant was given either 120 or 20 seconds to complete the insight task for the classical and modern approach, respectively. For this second component, a visual analog was displayed once again to indicate the remaining time.

A comprehensive literature scanning resulted in 56 classical insight tasks of sufficient quality. A strict selection method was applied to remove the 44 most infeasible insight tasks based on five criteria: 1. An average accuracy below 40% and above 80% 2. An average response time higher than 120 seconds 3. A digital formatting incompatibility (e.g., if answers cannot be typed in a centralized response box) 4. A high probability to be known beforehand 5. To increase the homogeneity of the set of classical insight tasks, only verbal classical insight tasks were selected. This process yielded a final set of 12 feasible classical insight tasks (see supplementary table 1).

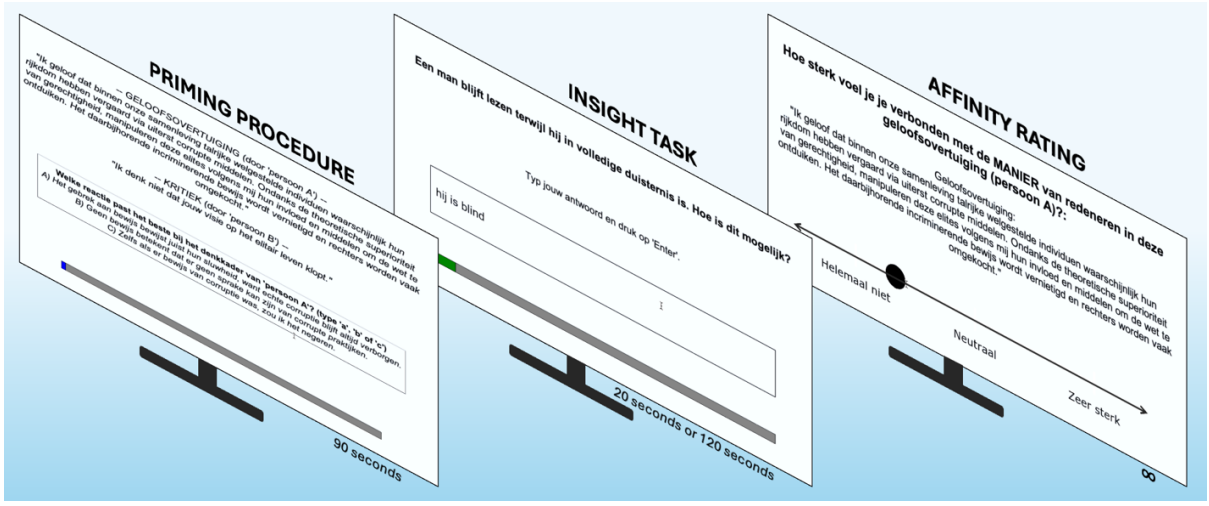
A modern approach to assess insight reasoning is the remote associates' test (RAT), crafted by Mednick & Halpern (1962). In this task, participants are presented with three seemingly unrelated cue words and must find one target word. Remote connections must be made within a semantic search space to find the solution. A 22-item Dutch variant was introduced by Chermahini et al. (2012), employing Item Response Theory to gauge item difficulty and discriminability. To obtain 12 modern insight tasks, the 10 items with the highest difficulty level were removed for the final set (see supplementary table 2). The participants were given a maximum of 20 seconds to enter their answer in a centered response box. Both task types were balanced across condition.

Huang et al. (2019) supported the applicability of Representational Change Theory for RAT by manipulating the sequential position of the keyword to alter the constraint on the semantic search space. Furthermore, Davelaar (2015) found that successfully solved items exhibited a pattern of anti-clustering where more between-patch transitions are made than expected by chance. The proposed priming mechanism still holds because the decreased proclivity to shift between feature spaces during DFU reasoning could instigate a reduction in the amount of switching between semantic patches during the RAT.

3) Each trial ended with a question about the overall affinity towards the cognitive schema of 'person A'. A 7-point slider was displayed with a Likert scale (1=very low affinity; 7=very strong affinity). The initial cognitive schema of 'person A' was displayed. No time constraint was used for this final component. After each block of four trials, a short obligatory break of ten seconds was inserted to reduce carry-over effects between blocks and deal with potentially depleting attention and motivation levels.

Figure 1

Exemplary sequence of a trial where its three constituents are illustrated



Note. Participants were instructed to read an in-first-person text of a belief system, which was either refutable or irrefutable; followed by a critique on that belief. Then, they had to answer a multiple-choice question where the correct answer is the option with the strongest aligning counter reaction. Afterwards, participants were given either a classical insight task or a remote

associates' task. Finally, a 7-point slider was presented where the overall affinity towards the belief system could be scored.

Measures and Statistical Analysis

The classical and modern insight tasks were evaluated in a standard binary manner (0=incorrect; 1=correct). Speed and accuracy of the classical and modern insight tasks were computed into a single balanced integration (BI) score (Liesefeld & Janczyk, 2019), which is designed to be relatively insensitive to speed-accuracy trade-offs. The usage of BI-score has been validated for within-subject designs (Liesefeld & Janczyk, 2022). Its formula is given by:

$$\mathrm{BI}_i = \frac{(1-w)\cdot z(\mathrm{acc}_i) - w \cdot z(\mathrm{RT}_i)}{\sqrt{(1-w)^2 + w^2}}$$

,where *w* is a weight parameter that was initialized at 0.5, *z* indicates the standardization for both accuracy scores and reaction times, and the denominator normalizes the weight difference.

The first analysis tested whether there is a significant priming effect. Since the effect relies on successful adoption of the cognitive schema, indicated by correctly answering the multiple-choice question, all incorrect trials were excluded if no correct response had been given earlier in that block. Four mixed-effects models were tested, crossed for unintegrated performance measures and task types, using a maximal random effects structure. Significance was evaluated at a Bonferroni-corrected threshold of 0.0125. Two additional mixed-effects models were included for both task types, with fixed effects for experimental condition, task type, and their interaction, using a threshold of 0.025. The analysis concluded with a single model integrating all task types and measures, using maximal random effects. Logistic mixed-effects models were, given the binary evaluation, used for the accuracies. Linear mixed-effects models were used for reaction times and BI scores, which are both continuous outcome variables.

The second analysis tested whether higher levels of affinity for DFU reasoning would be associated with lower insight performance. The affinity responses for the (12) DFU and (12) F cognitive schemata were averaged, which resulted in a 'DFU affinity index' and an 'inverted F

affinity index'. Subtracting the F affinity scores from 8 inverted them such that an 'overall affinity index' could be estimated using all the cognitive schemata. Eight Pearson correlation tests were run to examine whether the two latent disposition estimates ('DFU' and 'inverted F') are significantly associated with aggregated accuracy scores and reaction times for both classical and modern task types. A Bonferroni-corrected significance threshold of 0.00625 were used for the eight statistical tests. Two additional Pearson correlation tests were run to estimate the effect of both indices on the overall insight performance, as measured by the BI scores, with a Bonferroni-corrected significance threshold of 0.025. Finally, mixed-effects models with maximal random structure were fitted separately for all three performance measures.

All analyses, except for the mixed-effects models of the second analysis, were preregistered on Open Science Framework (see link1 to OSF). Mixed-effects analyses were conducted in R; correlation analyses and visualization in Python (see link2 to Github repository)

Results

Analysis 1: Higher-order Priming Effect

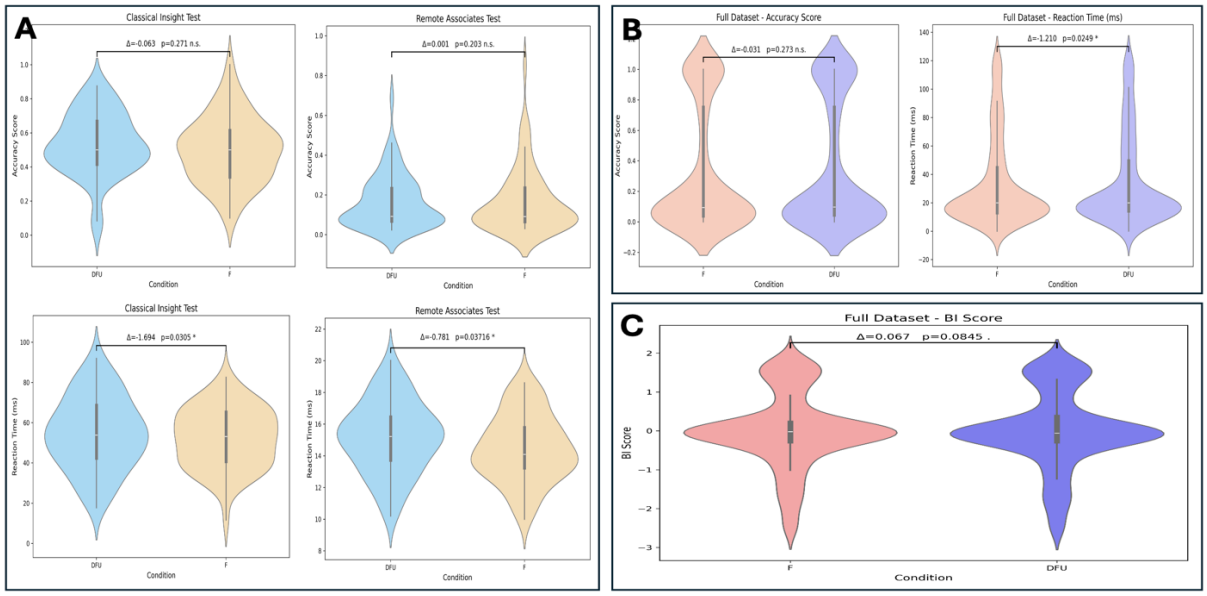
In total, 127 (8.82%) incorrect trials were excluded from the analysis for which no priming effect could have occurred. One participant was excluded from the analysis for miscomprehension of the experiment where all responses to the multiple-choice question were given at the time the insight response was requested. The mean accuracy for the classical insight tasks was 54.25% (*SD*=49.76%), for which no significant difference between conditions was found (Δ=0.063; *p*=0.271). Participants showed significantly higher reaction times for the DFU condition compared to the F condition (Δ=1.69s; *p*=0.0305) with a mean reaction time of 53.63 seconds (*SD*=34.71s). A similar pattern of significance was observed for the modern insight tasks, although the mean accuracy was noticeably lower (*M*=10.34%; *SD*=30.49%). Again, no significant differences were found between the two conditions for accuracy (Δ=-0.083%; *p*=0.203), whereas reaction times

(*M*=14.79s; *SD*=46.48s) did reach significance (*Δ*=0.781s; *p*=0.0372). No effects survived the Bonferroni-corrected significance threshold of 0.0125 (fig. 2).

After combining the data from both task types into a single dataset, a mixed-effects model was tested for both performance measures. The model where the dependent variable was accuracy showed, once again, no significant differences between the two conditions (*B*=0.202; *p*=0.273). For reaction times, the model marginally survived the Bonferroni-correct p-value of 0.025 (*B*=5.404; *p*=0.0249). Finally, the linear mixed-effects model with the integrated performance measure (i.e., BI score) for all task type data did not reach the significance threshold of 0.05 (*B*=0.11; *p*=0.0845).

Figure 2

Results of higher-order priming effect analysis



Note. A) No significant differences in accuracy between conditions were found for both task types. However, reaction times were significantly higher in the DFU condition for both classical and modern insight tasks. B) A similar significance pattern was found for the full dataset where only reaction times significantly differed across conditions. C) Finally, an integrating approach to estimate the effect on overall insight performance resulted in marginal non-significance.

In a post-hoc analysis, four models were fitted on the reaction time data for each task item separately. All the models (i.e., ex-Gaussian distribution, normal distribution, chi-squared distribution and polynomial distribution) demonstrated the substantial variance in performance,

which is to be expected for insight tasks (see supplementary figure 1). Similar substantial variance patterns were observed across participants and condition items (see supplementary figure 2).

Analysis 2: Association between DFU Affinity and Insight Performance

After aggregating all affinity scores for each participant and condition, no significant differences were found between DFU and inverted F affinity scores (Δ=-0.039; p=0.765) when using a paired samples t-test. No significant correlation was found between the two affinity indices (r=0.007; p=0.957) for the Pearson correlation test. When correlating the DFU affinity index scores with the four performance measures, the only significant association was found for the reaction times of the classical insight tasks (r=0.309; p=0.017). For the inverted F affinity index scores, both the accuracy of the classical insight tasks (r=0.300; p=0.021) and reaction time of the modern insight tasks (r=0.315; p=0.015) were significant. None of the eight Pearson correlation tests survived the conservative Bonferroni-corrected significance threshold of 0.00625 (fig. 3).

Subsequently, the accuracies and reaction times were integrated and aggregated across both task types to compute an overall BI score for each participant. A moderate negative correlation was found between these BI scores and the DFU affinity scores (r=-0.469; p=0.0002). Furthermore, a weak negative correlation was found for the inverted F affinity scores; however this association did not reach the Bonferroni-corrected significance threshold of 0.025 (r=-0.242; p=0.067).

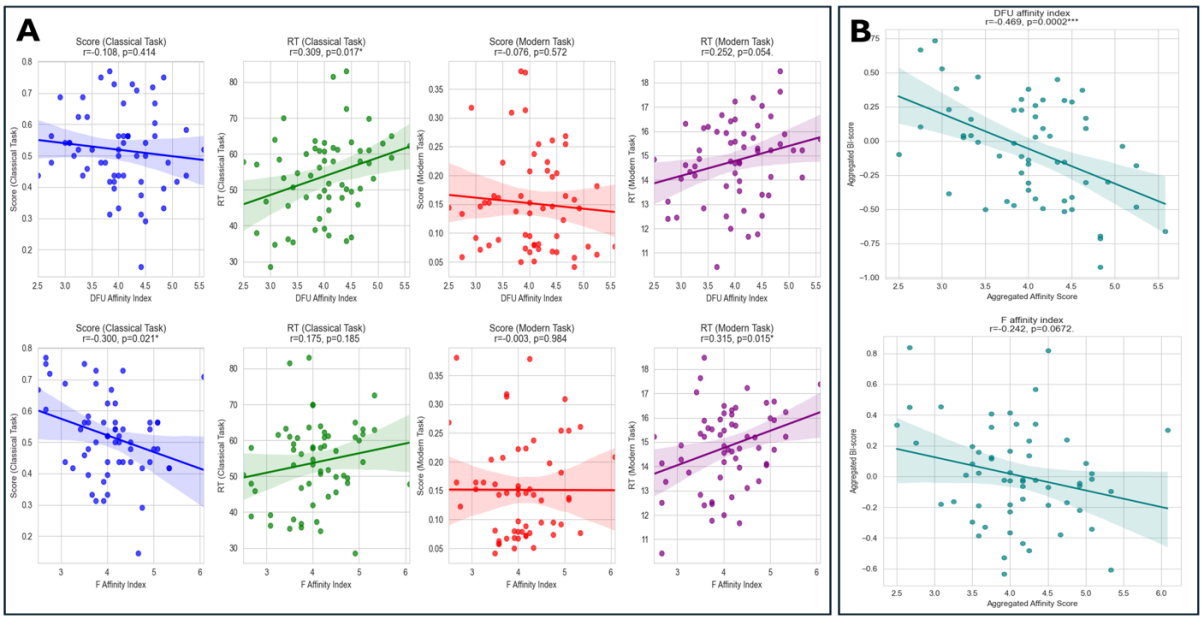
Separate mixed-effects models for the three performance measures were then tested with random intercepts for participants, task items and condition items. An estimation of verbal intelligence was incorporated as a fixed effect to control for the possibility that the association between DFU reasoning might actually come from a general association between verbal intelligence and exposure to DFU belief systems. Before participants commenced the experiment, a self-perceived estimate of verbal intelligence was obtained on a 7-point Likert scale (Furnham & Grover, 2020). Higher affinity levels were significantly associated with lower insight performance

for all three models (see supplementary figure 3): accuracy (B=0.0268; p=1.38e-08), reaction time (B=2.4276; p=2e-16), and BI score (B=-0.1146; p=3.14e-15).

In a post-hoc analysis, factor analysis was used to determine whether aggregation by averaging was a valid approach for estimating the underlying latent disposition scores. Although the factor analysis revealed a strong correlation (r>0.70) between the index scores where aggregation by averaging was used and the regressed factor scores, the extracted latent factors explained only a small proportion of variance across the affinity scores. This was consistent among separate analyses: 8% for the 12 DFU items, 10% for the 12 inverted F items, and 14% for the combined 24 items (see supplementary figure 4).

Figure 3

Analysis results of association between affinity scores and insight performance



Note. A) For the DFU affinity index, a significant association was found only with reaction times on the classical insight tasks. In contrast, the inverted F affinity index showed significant correlations with both the accuracy of classical insight tasks and the reaction times of modern insight tasks. However, none of the eight Pearson correlation tests remained significant after applying Bonferroni correction. B) Overall insight performance (i.e., BI score) was moderately and significantly correlated with DFU affinity index, whereas its association with the inverted F index was weaker and fell just short of significance.

Discussion

This paper investigated how both acute (i.e., through higher-order priming effect) and presumed chronic exposure (i.e., through DFU association) to irrefutable belief systems influence cognitive performance on problem domains where overcoming fixation and restructuring entrenched representations are essential to find the solutions. The increasing digital proliferation of pseudoscientific ideas on social media platforms (Faddoul et al., 2020; Chavda et al., 2022; Impey, 2024) raises pressing questions about their influence on how individuals process and respond to conflicting information. Both of the observations align with the assertion that the engagement with unfalsifiable beliefs can negatively impact insightful reasoning. However, caution about the quality of the research materials is warranted.

Two types of insight tasks were used to investigate the impact of irrefutable belief systems on maladaptive cognitive perseveration: 12 classical insight tasks and 12 remote associates' items. Increasing the total number of trials per condition would have improved statistical power. However, alternative cognitive tasks with relatively shorter time frames capable of capturing the proposed mechanism where an internal failure is attributed to the current problem representation would likely compromise the internal validity of the findings. Of the available cognitive executive functioning tasks that assess representational updating abilities, the Wisconsin card sorting task most closely aligns with the proposed priming mechanism (Friedman & Robbins, 2021). In this task, participants must initially infer sorting rules based on multiple hidden features, which requires representational updating and hypothesis testing. However, once these fixed features are discovered, the task primarily measures cognitive flexibility through strategy switching in response to changing reinforcement contingencies, rather than continued feature-based exploration. Therefore, classical insight tasks, in combination with the remote associates' items, are deemed the most viable option to achieve an optimal balance between internal validity and statistical power.

From both a theoretical and correlational perspective, classical insight tasks and remote associates' tasks are similar with respect to four properties: a) Since they have a limited solution space, they both measure convergent creative thinking abilities (Hommel, 2012; Colzato et al., 2012). b) An

impasse is caused by an initial and dominant, yet incorrect, problem representation (Bowden & Jung-Beeman, 2003). c) Constraint relaxation advantageously affects the likelihood of solving the insight task (Davelaar, 2015; Huang et al., 2019; Osuna-Mascaró & Auersperg, 2021). d) The majority of successful restructuring is accompanied by a sudden 'AHA!' experience in which the individual becomes aware of the correct problem representation in a way that feels subjectively compelling and self-evident (Danek et al., 2018; Stuck et al., 2022).

Effect of Priming Procedure on Insight Performance

The experimental effect on overall insight performance (i.e., BI score) trended towards the predicted pattern, yet did not reach statistical significance. Looking at both measures of insight performance separately, a similar pattern was found across both task types: Reaction times significantly increased after exposure to rigid cognitive frameworks, which aligns with the hypothesis of the prolonged impasse period. In contrast, accuracy measures did not significantly differ between the two experimental conditions. This absence of an effect on accuracy may reflect the inherent nature of insight problems, where individuals typically provide an answer only when they are confident in its correctness (Stuck et al., 2021). As a result, the accuracies are less likely to vary significantly across conditions, especially for longer time periods where guessing strategies are typically avoided.

Moreover, these findings prompt further reflection on the operational definition of 'irrefutability'. Although the DFU items were validated using a large language model to distinguish them from F items on the basis of falsifiability, the texts could still have embodied an unequal amount of subtle higher-order linguistic nuances. These nuances may have introduced confounding influences that affected participants' problem-solving processes in ways that were not entirely controlled. Another interpretation of these results could be that the observed 14.12% higher rate of correct responses on the multiple-choice questions in the DFU condition may have triggered a differential reallocation of cognitive resources, resulting in a subtle downstream benefit for subsequent insight task engagement.

From a methodological perspective, the reliance on a relatively brief priming paradigm, comprising maximally 90 seconds of reading and responding, might have limited the depth of

internalization of the respective DFU or F schemas. Future research could benefit from implementing longer exposure times, or employing between-subjects designs. Furthermore, the academic context in which the experiment took place raises questions regarding its external validity. All participants were psychology students who may have engaged with the material in a more analytical manner than a general population sample would. As such, future studies should consider replicating the findings in more ecologically valid settings and with a more diverse group of participants to ensure that the observed effect properly generalizes beyond artificial environments such as computer experiments. Relatedly, future research might opt for including an additional validation protocol or using alternative experimental designs to ensure that the participants properly internalized the respective cognitive schemata. The design in the current study was methodologically limited in this regard

Association between DFU affinity and Insight Performance

For a second method to examine how irrefutable belief systems might impact problem solving abilities, affinity scores were obtained for all the presented belief systems in the experiment. The primary assumption on which this cross-sectional examination is based is that stronger affinities likely coincide with stronger internalization and repeated exposure to those belief systems. To estimate the general affinity towards DFU reasoning, aggregation by averaging was used to derive a single estimate of the latent dispositions of the participants. The results of the initial eight Pearson correlation tests partially supported the hypothesis that higher DFU levels might be associated with poorer insight performance. All of the eight correlations trended towards the predicted results, three of which had a p-value below the conventional significance threshold of 0.05. None of them survived the conservative Bonferroni correction that statistically controls for multiple comparisons.

Secondly, the two insight performance measures were integrated into a single BI score for each participant, which was then correlated separately with both of the affinity indices. Results demonstrated a moderate Pearson correlation of -0.469 ($p<0.001$) for the DFU affinity index and a marginally non-significant correlation ($r=-0.242$; $p=0.067$) for the inverted F affinity index. These findings were further complemented by a post-hoc mixed-effects analysis that used the non-aggregated data and

self-perceived intelligence as a covariate. All the three performance measures were significantly associated with poorer insight performance (p<0.001). Together observations corroborate the hypothesis that higher levels of affinity towards DFU belief systems can negatively impact the ability to adaptively change cognitive perspectives.

In this study, self-perceived verbal intelligence was included as a covariate to control for its relationship with affinity scores. This methodological decision is based on the observation that certain institutions that tend to oppose pseudoscientific belief systems tend to be primarily constituted by individuals with higher levels of verbal intelligence (Ritchie & Tucker-Drob, 2018; Izzaty & Setiawati, 2019). Given that verbal insight itself moderately correlates with general verbal intelligence (Dechaume et al., 2024), this represents an important confounding factor that could distort the interpretations of the actual relationship between DFU affinity and insight performance. A noteworthy limitation should be recognized in light of prior findings by Furnham and Grover (2020), who reported only a weak correlation between self-perceived intelligence and objective intelligence estimates. Future research with more resources and less financial constraints should consider using standardized estimates of verbal intelligence.

Furthermore, statistical range restriction may be a concern given that no affinity index scores were obtained in the extreme lower (1–2) or upper (6–7) ranges. This limitation potentially constrains the generalizability of the findings, particularly for populations exhibiting more pronounced or radical affinities towards certain belief systems or ideologies. Consequently, especially for future clinical research, it may be advantageous to adopt qualitative, in-depth case studies rather than relying solely on quantitative nomothetic methods. This alternative approach could yield a more nuanced portrayal of the dynamics at play and better address the challenges of recruiting a large number of participants with these extreme affinity scores.

Given that the LLM-based validation method of the reading material resulted in a substantial effect size (d=-5.83; p=1.797e-12) between the two conditions for the degree of falsifiability, it

seems reasonable to assume that a simple averaging technique might suffice to estimate the two latent dispositions. Since the post-hoc factor analysis did not confirm this assumption, future research should consider using psychometrically-sound items that have been validated by conventional evaluation frameworks such as Item Response Theory (Yigiter & Boduroglu, 2024; Chen et al., 2021).

Moreover, contrary to the expectation that university students would exhibit an overall stronger liking towards scientific falsificationist frameworks, the raw affinity scores did not significantly differ between conditions (p=0.765). Given that also no significant correlation was found between the two affinity indices (p=0.957), these null findings might be interpreted by looking at DFU reasoning styles as strongly context-dependent. This might imply, once again, that future research should opt for perhaps rather qualitative use case studies where idiosyncratic belief systems are modelled independently of other belief systems.

Implications and Future Directions

Both the experimental and cross-sectional investigation point to the assertion that the adoption of irrefutable belief systems can hinder the mental flexibility needed to adaptively restructure ideas. Given the importance of repetitive negative thinking (Ehring & Watkins, 2009; Zagaria et al., 2023) across many psychopathologies, future research could potentially benefit from incorporating the degree of falsifiability as a novel transdiagnostic variable into contemporary etiological models. Furthermore, a heightened awareness from psychotherapists concerning this topic could be vital to guide clientele towards the discovery of their state (Overholser & Beale, 2023).

Apart from clinical implications, Boudry (2022) posits that the key ingredient causing a spiral towards radical conspiracies is the inherent irrefutability of the theories themselves. With the rise of pseudoscience on various media platforms (Zhang et al., 2021; Impey, 2024), social networking companies might benefit from including the degree of falsifiability as a criterion in

their content flagging algorithms (Katsaros et al., 2023; Gomes & Sultan, 2024) to prevent users from spiraling into distressing thought bubbles or radicalizing harmful behavior.

Importantly, more research should be conducted on the specific neurocognitive mechanisms behind this effect of irrefutability on impaired cognitive performance to change mental representations. An interesting avenue for neuroscientific research would be to use electroencephalography to investigate whether event-related potential paradigms could corroborate the proposed mechanism of having a desensitized trigger that decreases the overall likelihood to adaptively restructure mental representations. A promising approach would be to examine the feedback-related negativity (Huang & Yu, 2014; Wang et al., 2020) in designs where participants are required to respond to contrary evidence. Peters et al. (2025) showed that incongruent evaluative feedback about self-views elicits more mid- and late-stage frontal processing that subsequently drove rapid behavioral updating. The present findings would be further substantiated if attenuated neural responses and reduced behavioral adjustments were observed in individuals with higher levels of DFU affinity when confronted with conflicting self-relevant information.

A further line of inquiry may consider functional magnetic resonance imaging studies employing representational similarity analysis (Kriegeskorte et al., 2008; Matheson et al., 2023) to test whether, and in what manner, poorer working memory updating abilities are found in individuals with stronger affinities towards irrefutable reasoning styles. According to the cascade-of-control model (Banich, 2019), the dorsal anterior cingulate cortex monitors the performance of a problem representation, which is maintained by the dorsolateral prefrontal cortex projecting to posterior uni- and multimodal association regions (MacDonald et al., 2000; Chiang et al., 2014). Upon internal detection of performance failure, the probability of triggering representational change could vary across different levels of latent DFU affinity estimates. Finding a significantly stronger autocorrelation of multivariate time series data for brain regions that represent the time-

varying working memory content might corroborate the findings of this study. In that, individuals with elevated DFU levels may exhibit greater difficulty in exploring alternative representations.

Finally, the post-hoc factor and item-level analyses in this study illustrated that the current items may not have been ideal to estimate an underlying disposition towards overall irrefutable belief systems. Future studies should consider using items that have been psychometrically validated by conventional evaluation frameworks such as Item Response Theory (Yigiter & Erdem, 2024; Chen et al., 2021). The lack of a negative correlation between the DFU index and F index, which are two semantically opposing constructs, furthermore demonstrates the context-dependent nature of endorsing specific belief systems. This finding may suggest that item-wise aggregation procedures may be insufficient for obtaining latent disposition scores. Future research could benefit from qualitative, in-depth, comparative case studies that isolate and examine idiosyncratic belief systems within the sociocognitive and contextual frameworks of the individuals.

Conclusion

In conclusion, the results of this study provide preliminary evidence for the hypothesis that irrefutable belief systems can impair insightful reasoning by delaying the timepoint at which adaptive representational change is triggered. This effect is observed under both acute priming and chronic exposure investigation methods, particularly apparent in delayed reaction times of participants. To corroborate the observed effects, future research should consider using alternative approaches and psychometrically validated research materials.

AI acknowledgement section

AI tools, specifically OpenAI's 'gpt-4o' and 'o1-mini' models, were used to support the writing process of this report. Their contributions included rephrasing sentences for clarity and academic tone, generating alternative formulations, summarizing background literature, and improving the coherence of arguments in the discussion and future directions sections. All AI-assisted content was critically reviewed. The scientific reasoning, structure, and conclusions presented are entirely the result of the author's work.

Furthermore, as part of the study, an LLM-based validation technique was used to score 30 in-first-person texts on five dimensions. OpenAI's large language model 'o1' was used for this validation method. The Python script and prompt can be found in the Github repository in the file 'validate.py' under the directory 'RPEP_PROJECT/experiment/validation_materials/scripts'.

Bibliography

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. Clinical Psychology Review, *30*(2), 217-237 https://doi.org/10.1016/j.cpr.2009.11.004

Baddeley, A. D., & Hitch, G. (2017). Working memory. Exploring Working Memory, 43-79 https://doi.org/10.4324/9781315111261-780

Banich, M. T. (2009). *Executive Function: The Search for an Integrated Account. Current Directions in Psychological Science, 18*(2), 89-94. https://doi.org/10.1111/j.1467-8721.2009.01615.x

Banich, M. T. (2019). The Stroop effect occurs at multiple points along a Cascade of control: Evidence from cognitive neuroscience approaches. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.02164

Barbieri, F., Camacho-Collados, J., Espinosa Anke, L., & Neves, L. (2020). TweetEval: Unified benchmark and comparative evaluation for tweet classification. *Findings of the Association for Computational Linguistics: EMNLP 2020*. https://doi.org/10.18653/v1/2020.findings-emnlp.148

Baz, A., & Karagüzel, E.O. (2022). Comparison of early maladaptive schemas in obsessive-compulsive disorder patients, their siblings, and controls. *ALPHA PSYCHIATRY, 23*(4), 157-163. https://doi.org/10.5152/alphapsychiatry.2022.21565

Boudry, M. (2022). Why we should be suspicious of conspiracy theories: A novel demarcation problem. *Episteme, 20*(3), 611-631. https://doi.org/10.1017/epi.2022.3426

Boudry, M., & Braeckman, J. (2010). Immunizing strategies and epistemic defense mechanisms. *Philosophia, 39*(1), 145-161. https://doi.org/10.1007/s11406-010-9254-9

Boudry, M., Braeckman, J., & Buekens, F. (2007). De naakte keizers van de psychoanalyse : de immunisatiestrategieën van een pseudowetenschap. https://lib.ugent.be/nl/catalog/rug01:001396449

Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior Research Methods, Instruments, & Computers, 35*(4), 634-

639. https://doi.org/10.3758/bf03195543

Brosschot, J. F., Gerin, W., & Thayer, J. F. (2006). The perseverative cognition hypothesis: A review of worry, prolonged stress-related physiological activation, and health. *Journal of Psychosomatic Research, 60*(2), 113-124. https://doi.org/10.1016/j.jpsychores.2005.06.074

Brysbaert, M., & Debeer, D. (2023). How to run linear mixed effects analysis for pairwise comparisons? A tutorial and a proposal for the calculation of standardized effect sizes. https://doi.org/10.31234/osf.io/esnku

Caddick, Z. A., & Feist, G. J. (2021). When beliefs and evidence collide: Psychological and ideological predictors of motivated reasoning about climate change. *Thinking & Reasoning*, *28*(3), 428-464. https://doi.org/10.1080/13546783.2021.1994009

Cameron, C. D., Payne, B. K., Sinnott-Armstrong, W., Scheffer, J. A., & Inzlicht, M. (2017). Implicit moral evaluations: A multinomial modeling approach. *Cognition, 158*, 224-241. https://doi.org/10.1016/j.cognition.2016.10.013

Castro, E., Wray-Lake, L., & Cohen, A. K. (2022). Critical consciousness and wellbeing in adolescents and young adults: A systematic review. *Adolescent Research Review, 7*(4), 499-522. https://doi.org/10.1007/s40894-022-00188-3

Chavda, V. P., Sonak, S. S., Munshi, N. K., & Dhamade, P. N. (2022). Pseudoscience and fraudulent products for COVID-19 management. *Environmental Science and Pollution Research*, *29*(42), 62887-62912. https://doi.org/10.1007/s11356-022-21967-4

Chermahini, A. S., Hickendorff, M., & Hommel, B. (2012). Development and validity of a Dutch version of the remote associates task: An item-response theory approach. *Thinking Skills and Creativity, 7*(3), 177-186. https://doi.org/10.1016/j.tsc.2012.02.00327

Chiang, T., Lu, R., Hsieh, S., Chang, Y., & Yang, Y. (2014). Stimulation in the Dorsolateral prefrontal cortex changes subjective evaluation of percepts. *PLoS ONE, 9*(9), e106943. https://doi.org/10.1371/journal.pone.0106943

Colzato, L. S., Ozturk, A., & Hommel, B. (2012). Meditate to create: The impact of focused-attention

and open-monitoring training on convergent and divergent thinking. *Frontiers in Psychology, 3*. https://doi.org/10.3389/fpsyg.2012.00116

Dai, W., Yang, T., White, B. X., Palmer, R., Sanders, E. K., McDonald, J. A., Leung, M., & Albarracín, D. (2023). Priming behavior: A meta-analysis of the effects of behavioral and nonbehavioral primes on overt behavioral outcomes. *Psychological Bulletin*, *149*(1-2), 67-98. https://doi.org/10.1037/bul0000374

Danek, A. H., Williams, J., & Wiley, J. (2018). Closing the gap: Connecting sudden representational change to the subjective AHA! experience in insightful problem solving. *Psychological Research, 84*(1), 111-119. https://doi.org/10.1007/s00426-018-0977-8

Davelaar, E. J. (2015). Semantic search in the remote associates test. *Topics in Cognitive Science, 7*(3), 494-512. https://doi.org/10.1111/tops.12146

Dehaene, S., Changeux, J., & Naccache, L. (2011). The global neuronal workspace model of conscious access: From neuronal architectures to clinical applications. *Research and Perspectives in Neurosciences,* 55-84. https://doi.org/10.1007/978-3-642-18015-6_4

Dewitte, M. (2011). Adult attachment and attentional inhibition of interpersonal stimuli. *Cognition & Emotion, 25*(4), 612-625. https://doi.org/10.1080/02699931.2010.508683

Disner, S. G., Beevers, C. G., Haigh, E. A., & Beck, A. T. (2011). Neural mechanisms of the cognitive model of depression. *Nature Reviews Neuroscience, 12*(8), 467-28 477. https://doi.org/10.1038/nrn3027

Donoghue, T., & Voytek, B. (2022). Automated meta-analysis of the event-related potential (ERP) literature. *Scientific Reports, 12*(1). https://doi.org/10.1038/s41598-022-05939-9

Douglas, K. M., & Sutton, R. M. (2023). What are conspiracy theories? A definitional approach to their correlates, consequences, and communication. *Annual Review of Psychology, 74*(1), 271-298. https://doi.org/10.1146/annurev-psych-032420-031329

Dobson, K. S., & Dozois, D. J. (2008). Introduction. *Risk Factors in Depression*, 1-16. https://doi.org/10.1016/b978-0-08-045078-0.00001-0

Dreisbach, G., & Fischer, R. (2012). Conflicts as aversive signals. *PsycEXTRA Dataset.* https://doi.org/10.1037/e502412013-956

Drouvelis, M., Metcalfe, R., & Powdthavee, N. (2015). Can priming cooperation increase public good contributions? *Theory and Decision, 79*(3), 479-492. https://doi.org/10.1007/s11238-015-9481-4

Ehring, T., & Watkins, E. R. (2008). Repetitive negative thinking as a transdiagnostic process. *International Journal of Cognitive Therapy*, *1*(3), 192-205. https://doi.org/10.1521/ijct.2008.1.3.192

Ellaway, R. H. (2020). Postmodernism and medical education. *Academic Medicine, 95*(6), 856-859. https://doi.org/10.1097/acm.0000000000003136

Etkin, A., Büchel, C., & Gross, J. J. (2015). The neural bases of emotion regulation. *Nature Reviews Neuroscience, 16*(11), 693-700. https://doi.org/10.1038/nrn4044

Faddoul, M., Chaslot, G., & Farid, H. (2020). A longitudinal analysis of YouTube's promotion of conspiracy videos. arXiv. https://doi.org/10.48550/arXiv.2003.03318

Friedman N.P, Robbins T.W. (2021). The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology, 47*(1), 72-89. https://doi.org/10.1038/s41386-021-01132-0

Friesen, J. P., Campbell, T. H., & Kay, A. C. (2015). The psychological advantage of unfalsifiability: The appeal of untestable religious and political ideologies. *Journal of Personality and Social Psychology*, *108*(3), 515-529. https://doi.org/10.1037/pspp0000018

Genschow, O., Cracco, E., Schneider, J., Protzko, J., Wisniewski, D., Brass, M., & Schooler, J. (2021). Manipulating belief in free will and its downstream consequences: A meta-analysis. https://doi.org/10.31234/osf.io/quwgr

Gilhooly, K., & Murphy, P. (2005). Differentiating insight from non-insight problems. *Thinking*

& Reasoning, 11(3), 279-302. https://doi.org/10.1080/135467804420001 87

Gomes, A. B., & Sultan, A. (2024). Problematizing content moderation by social media platforms and its impact on digital harm reduction. *Harm Reduction Journal*, *21*(1). https://doi.org/10.1186/s12954-024-01104-9

Gonthier, C. (2022). Cross-cultural differences in visuo-spatial processing and the culture-fairness of visuo-spatial intelligence tests: An integrative review and a model for matrices tasks. *Cognitive Research: Principles and Implications, 7*(1). https://doi.org/10.1186/s41235-021-00350-w

Hakan, T. (2021). Philosophy of science and Black swan. *Child's Nervous System, 38*(9), 1655-1657. https://doi.org/10.1007/s00381-020-05009-3

Hallion, L. S., Wright, A. G., Joormann, J., Kusmierski, S. N., Coutanche, M. N., & Caulfield, M. K. (2022). A five-factor model of perseverative thought. *Journal of Psychopathology and Clinical Science, 131*(3), 235-252. https://doi.org/10.1037/abn0000737

Hao, J., Plangger, K., & West, D. (2024). Conceptualizing sustainable consumption priming: A scoping review. *Psychology & Marketing*, *41*(11), 2772-2788. https://doi.org/10.1002/mar.22083

Hauser, T. U., Iannaccone, R., Stämpfli, P., Drechsler, R., Brandeis, D., Walitza, S., & Brem, S. (2014). The feedback-related negativity (FRN) revisited: New insights into the localization, meaning and network organization. *NeuroImage, 84*, 159-168. https://doi.org/10.1016/j.neuroimage.2013.08.028

Horner, A., & Henson, R. (2008). Priming, response learning and repetition suppression. *Neuropsychologia*, *46*(7), 1979-1991. https://doi.org/10.1016/j.neuropsychologia.2008.01.018

Huang, P., Liu, C., & Chen, H. (2019). Examining the applicability of representational change theory for remote associates problem-solving with eye movement evidence. *Thinking Skills and Creativity, 31*, 198-208. https://doi.org/10.1016/j.tsc.2018.12.001

Huang, Y., & Yu, R. (2014). The feedback-related negativity reflects more or less prediction error in appetitive and aversive conditions. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00108

Joset, E., & Todd, B. (2015). Reward motivation enhances coding of task-set information in frontoparietal cortex. *Frontiers in Human Neuroscience, 9*. https://doi.org/10.3389/conf.fnhum.2015.217.00008

Kamarajan, C. (2019). Brain electrophysiological signatures in human alcoholism and risk. *Neuroscience of Alcohol*, 119-130. https://doi.org/10.1016/b978-0-12-813125-1.00013-1

Katsaros, M., Kim, J., & Tyler, T. (2023). Online content moderation: Does justice need a human face? *International Journal of Human–Computer Interaction*, *40*(1), 66-77. https://doi.org/10.1080/10447318.2023.2210879

Kent, S. (1999). Creation of 'Religious' Scientology. *Religious Studies and Theology*, *18*(2), 97-126. https://doi.org/10.1558/rsth.v18i2.97

Kim, E. J., Tanford, S., & Book, L. A. (2020). The effect of priming and customer reviews on sustainable travel behaviors. *Journal of Travel Research, 60*(1), 86-101. https://doi.org/10.1177/0047287519894069

Kincaid, J. P., Fishburne, J., Robert P., R., Richard L., C., & Brad S. (1975). Derivation of new readability formulas (Automated readability index, fog count and Flesch reading ease formula) for navy enlisted personnel. https://doi.org/10.21236/ada006655

Klayman, J. (1995). Varieties of confirmation bias. *Psychology of Learning and Motivation*, 385-418. https://doi.org/10.1016/s0079-7421(08)60315-1

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 25*(6), 1534-1555. https://doi.org/10.1037/0278-7393.25.6.153431

Korovkin, S., Vladimirov, I., Chistopolskaya, A., & Savinova, A. (2018). How working memory

provides representational change during insight problem solving. *Frontiers in Psychology, 9*. https://doi.org/10.3389/fpsyg.2018.01864

Koster, E. H., De Lissnyder, E., Derakshan, N., & De Raedt, R. (2011). Understanding depressive rumination from a cognitive science perspective: The impaired disengagement hypothesis. *Clinical Psychology Review, 31*(1), 138-145. https://doi.org/10.1016/j.cpr.2010.08.005

Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*. https://doi.org/10.3389/neuro.06.004.2008

Kube, T., Glombiewski, J. A., Gall, J., Touissant, L., Gärtner, T., & Rief, W. (2019). How to modify persisting negative expectations in major depression? An experimental study comparing three strategies to inhibit cognitive immunization against novel positive experiences. *Journal of Affective Disorders, 250*, 231-240. https://doi.org/10.1016/j.jad.2019.03.027

Kube, T., Kirchner, L., Gärtner, T., & Glombiewski, J. A. (2021). How negative mood hinders belief updating in depression: Results from two experimental studies. *Psychological Medicine*, *53*(4), 1288-1301. https://doi.org/10.1017/s0033291721002798

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480-498. https://doi.org/10.1037//0033-2909.108.3.480

Lee, C. S., Huggins, A. C., & Therriault, D. J. (2014). A measure of creativity or intelligence? Examining internal and external structure validity evidence of the remote associates test. *Psychology Of Aesthetics, Creativity, and the Arts, 8*(4), 446-460. https://doi.org/10.1037/a0036773

Liesefeld, H. R., & Janczyk, M. (2018). Combining speed and accuracy to control for speed-accuracy trade-offs(?). *Behavior Research Methods, 51*(1), 40-32 60. https://doi.org/10.3758/s13428-018-1076-x

Liesefeld, H. R., & Janczyk, M. (2022). Correction to: Same same but different: Subtle but consequential differences between two measures to linearly integrate speed and accuracy (LISAS vs. BIS). *Behavior Research Methods, 55*(3), 1511-1511. https://doi.org/10.3758/s13428-022-01963-9

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the Dorsolateral prefrontal and anterior Cingulate cortex in cognitive control. *Science, 288*(5472), 1835-1838. https://doi.org/10.1126/science.288.5472.1835

Mashour, G. A., Roelfsema, P., Changeux, J., & Dehaene, S. (2022). Conscious processing and the global neuronal workspace hypothesis. *Neuron, 105*(5), 776-798. https://doi.org/10.1016/j.neuron.2020.01.026

Matheson, H. E., Kenett, Y. N., Gerver, C., & Beaty, R. E. (2023). Representing creative thought: A representational similarity analysis of creative idea generation and evaluation. *Neuropsychologia*, *187*, 108587. https://doi.org/10.1016/j.neuropsychologia.2023.108587

McDonnell, M. D., & Abbott, D. (2009). What is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology. PLoS Computational Biology*, 5*(5), e1000348. https://doi.org/10.1371/journal.pcbi.1000348

Mednick, M. T., & Halpern, S. (1962). Remote associates test. *PsycTESTS Dataset.* https://doi.org/10.1037/t11859-000

Mehl, S., Rief, W., Soll, D., & Pytlik, N. (2025). Populist attitudes and belief in conspiracy theories: Anti-elitist attitudes and the preference for unrestricted popular sovereignty reduce the positive impact of an analytical thinking style on conspiracy beliefs. *BMC Research Notes*, *18*(1). https://doi.org/10.1186/s13104-025-07136-z

Millard, S. J., Bearden, C. E., Karlsgodt, K. H., & Sharpe, M. J. (2021). The prediction-error hypothesis of schizophrenia: New data point to circuit-specific changes in dopamine activity. Neuropsychopharmacology*, 47*(3), 628-640. https://doi.org/10.1038/s41386-021-01188-y

Misra, S. (2012). Randomized double blind placebo control studies, the "Gold standard" in intervention based studies. *Indian Journal of Sexually Transmitted Diseases and AIDS, 33*(2), 131. https://doi.org/10.4103/0253-7184.102130

Modgil, S., Singh, R.K., Gupta, S. *et al.* A Confirmation Bias View on Social Media Induced Polarisation During Covid-19. *Inf Syst Front 26*, 417–441 (2024). https://doi.org/10.1007/s10796-021-10222-9

Nyberg, L., & Eriksson, J. (2015). Working memory: Maintenance, updating, and the realization of intentions. *Cold Spring Harbor Perspectives in Biology, 8*(2), a021816. https://doi.org/10.1101/cshperspect.a021816

Öllinger, M., Jones, G., & Knoblich, G. (2013). The dynamics of search, impasse, and representational change provide a coherent explanation of difficulty in the nine-dot problem. *Psychological Research, 78*(2), 266-275. https://doi.org/10.1007/s00426-013-0494-8

OpenAI. (2023). GPT-4 Technical Report. Retrieved from https://openain.com/research/gpt-4
Osuna-Mascaró, A. J., & Auersperg, A. M. (2021). Current understanding of the "Insight" phenomenon across disciplines. *Frontiers in Psychology, 12.* https://doi.org/10.3389/fpsyg.2021.791398

Overholser, J. C., & Beale, E. (2023). The art and science behind socratic questioning and guided discovery: A research review. *Psychotherapy Research, 33*(7), 946-956. https://doi.org/10.1080/10503307.2023.2183154

Panitz, C., Endres, D., Buchholz, M., Khosrowtaj, Z., Sperl, M. F., Mueller, E. M., Schubö, A., Schütz, A. C., Teige-Mocigemba, S., & Pinquart, M. (2021). A revised framework for the investigation of expectation update versus maintenance in the context of expectation violations: The ViolEx 2.0 model. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.726432

Peters, A., Witte, J., Helming, H., Moeck, R., Straube, T., & Schindler, S. (2025). How positive and negative feedback following real interactions changes subsequent sender ratings. *Scientific Reports*, *15*(1). https://doi.org/10.1038/s41598-025-91750-1

Qiu, J., Li, H., Yang, D., Luo, Y., Li, Y., Wu, Z., & Zhang, Q. (2008). The neural basis of insight problem solving: An event-related potential study. *Brain and Cognition, 68*(1), 100-34 106. https://doi.org/10.1016/j.bandc.2008.03.004

Reinecke, A., Rinck, M., Becker, E. S., & Hoyer, J. (2013). Cognitive-behavior therapy resolves implicit fear associations in generalized anxiety disorder. *Behaviour Research and Therapy*, *51*(1), 15-23. https://doi.org/10.1016/j.brat.2012.10.004

Rustam, F., Khalid, M., Aslam, W., Rupapara, V., Mehmood, A., & Choi, G. S. (2021). A performance comparison of supervised machine learning models for COVID-19 tweets sentiment analysis. *PLOS ONE, 16*(2), e0245909. https://doi.org/10.1371/journal.pone.0245909

Samtani, S., & Moulds, M. L. (2017). Assessing maladaptive repetitive thought in clinical disorders: A critical review of existing measures. *Clinical Psychology Review, 53*, 14-28. https://doi.org/10.1016/j.cpr.2017.01.007

Shen, W., Tong, Y., Li, F., Yuan, Y., Hommel, B., Liu, C., & Luo, J. (2018). Tracking the neurodynamics of insight: A meta-analysis of neuroimaging studies. *Biological Psychology, 138*, 189-198. https://doi.org/10.1016/j.biopsycho.2018.08.018

Shen, W., Yuan, Y., Lu, F., Liu, C., Luo, J., & Zhou, Z. (2019). Unpacking impasse-related experience during insight. *The Spanish Journal of Psychology, 22*. https://doi.org/10.1017/sjp.2019.40

Shenhav, A., Botvinick, M., & Cohen, J. (2013). The expected value of control: An integrative theory of anterior Cingulate cortex function. *Neuron, 79*(2), 217-240. https://doi.org/10.1016/j.neuron.2013.07.007

Sheth, B. R., Sandkühler, S., & Bhattacharya, J. (2009). Posterior beta and anterior gamma oscillations predict cognitive insight. *Journal of Cognitive Neuroscience, 21*(7), 1269-1279. https://doi.org/10.1162/jocn.2009.2106935

Stuyck, H., Aben, B., Cleeremans, A., & Van den Bussche, E. (2021). The AHA! moment: Is insight a different form of problem solving? *Consciousness and Cognition, 90*, 103055. https://doi.org/10.1016/j.concog.2020.103055

Tyler, C. P., Geldhof, G. J., Black, K. L., & Bowers, E. P. (2019). Critical reflection and positive youth development among white and Black adolescents: Is understanding inequality connected to thriving? *Journal of Youth and Adolescence, 49*(4), 757-771. https://doi.org/10.1007/s10964-019-01092-1

Van der Groen, O., & Wenderoth, N. (2016). Transcranial random noise stimulation of visual cortex: Stochastic resonance enhances central mechanisms of perception. *The Journal of Neuroscience,*

36(19), 5289-5298. https://doi.org/10.1523/jneurosci.4519-15.2016

Wadhera, T., & Kakkar, D. (2020). Modeling risk perception using independent and social learning: Application to individuals with autism spectrum disorder. *The Journal of Mathematical Sociology, 45*(4), 223-245. https://doi.org/10.1080/0022250x.2020.1774877

Wang, Y., Cheung, H., Yee, L. T., & Tse, C. (2020). Feedback-related negativity (FRN) and theta oscillations: Different feedback signals for non-conform and conform decisions. *Biological Psychology*, *153*, 107880. https://doi.org/10.1016/j.biopsycho.2020.107880

Wild, B., & Treue, S. (2021). Primate extrastriate cortical area MST: A gateway between sensation and cognition. *Journal of Neurophysiology, 125*(5), 1851-1882. https://doi.org/10.1152/jn.00384.202036

Würtz, F., Kube, T., Woud, M. L., Margraf, J., & Blackwell, S. E. (2024). Reduced belief updating in the context of depressive symptoms: An investigation of the associations with interpretation biases and self-evaluation. *Cognitive Therapy and Research*, *48*(2), 225-241. https://doi.org/10.1007/s10608-023-10454-w

Xing, Q., Lu, Z., & Hu, J. (2019). The effect of working memory updating ability on spatial insight problem solving: Evidence from behavior and eye movement studies. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.00927

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., & Wager, T. D. (2011). Large-scale automated synthesis of human functional neuroimaging data. *Nature Methods, 8*(8), 665-670. https://doi.org/10.1038/nmeth.1635

Yigiter, M. S., Boduroglu, E. (2024) Item Response Theory assumptions: A comprehensive review of studies with document analysis. International Journal of Educational Studies and Policy, 5(2). https://doi.org/10.5281/zenodo.14016086

Zagaria, A., Ballesio, A., Vacca, M., & Lombardo, C. (2023). Repetitive negative thinking as a central node between Psychopathological domains: A network analysis. *International Journal of Cognitive Therapy*, *16*(2), 143-160. https://doi.org/10.1007/s41811-023-00162-4

Appendices

The following Github repository contains the complete codebase of this academic project with: a) all Dutch and translated stimuli that were used in the experiment b) the full code of the Psychopy experiment c) the raw data of the experiment d) the full code of the analysis of the raw data e) all figures that were used in the main text, supplementary materials and more: see link2 to codebase.

Supplementary Table 1

Table with all translated classical insight items that were used in the experiment

Task ID	Classical Insight Problem
1	"A window cleaner falls from a 12-meter ladder onto a concrete surface, but is not injured. How is that possible?"
2	"What has cities without houses, forests without trees, and rivers without water?"
3	"What can be broken without ever being touched or seen?"
4	"What occurs once in a minute, twice in a moment, but never in a thousand years?"
5	"What travels around the world but stays in one corner?"
6	"A man keeps reading while he is in complete darkness. How is this possible?"
7	"How can someone walk on the surface of a lake without sinking and without using any aids?"
8	"Ruben is taking part in a running race on Friday. He runs faster than Marit, who trains with Him every Monday, Wednesday, and Friday. Despite her busy training schedule, Marit has never run faster than Ruben. Him is a world record holder and, by coincidence, also Marit's coach. He is slower than the coach, but faster than Ruben. Arrange the FOUR runners from fastest to slowest using '>' to separate them:"
9	"An antique coin dealer received an offer to purchase a beautiful bronze coin. The coin featured an emperor's head on one side and the date "544 BC" on the other. The dealer examined the coin, but instead of buying it, he called the police to arrest the man. Why did the dealer suspect that the coin was fake?"
10	"Using only a 7-minute hourglass and an 11-minute hourglass, how can you time exactly 15 minutes to cook an egg?"
11	"What can go up and come down without ever moving?"
12	"I am not alive, but I grow; I have no lungs, but I breathe; I have no mouth, yet water kills me. What am I?"

Note. Out of 56 identified classical insight tasks, 44 were excluded based on several criteria: unsuitable accuracy levels, excessive response times, digital formatting issues, potential prior exposure, and non-verbal content. This selection procedure resulted in a refined set of 12 verbal insight tasks appropriate for this study's framework. The correct answers to the classical insight tasks can be provided by the author upon request.

Supplementary Table 2

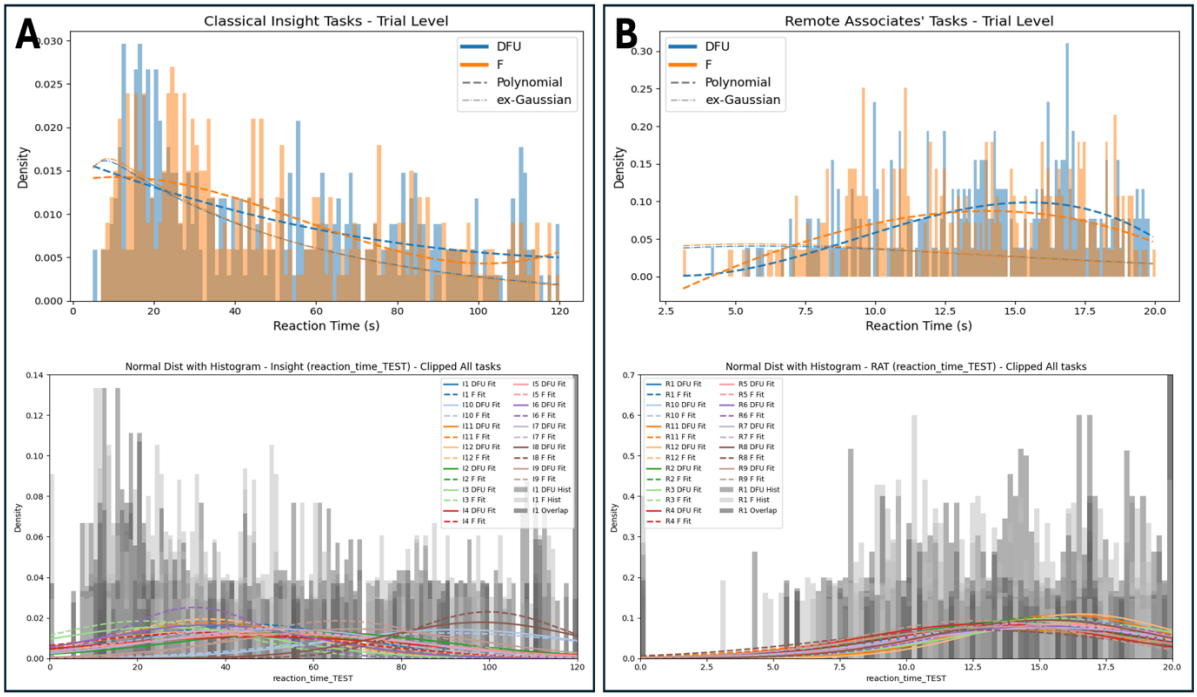
Table with all Dutch remote associates' items that were used in the experiment

Task ID	Cue Words
1	["Bar", "jurk", "glas"]
2	["Kaas", "land", "huis"]
3	["Vlokken", "ketting", "pet"]
4	["Val", "meloen", "lelie"]
5	["Vis", "mijn", "geel"]
6	["Achter", "kruk", "mat"]
7	["Worm", "kast", "legger"]
8	["Water", "schoorsteen", "lucht"]
9	["Trommel", "beleg", "mes"]
10	["Hond", "druk", "band"]
11	["Controle", "plaats", "gewicht"]
12	["Goot", "kool", "bak"]

Note. Another approach to estimating convergent creativity is to use remote associates' items. A Dutch 22-item version was developed by Chermahini et al. (2012). Twelve items were selected by excluding the 10 most difficult, based on Item Response Theory. Each item requires identifying a single target word that matches three seemingly unrelated cue words. Correct answers can be provided by the author upon request.

Supplementary figure 1

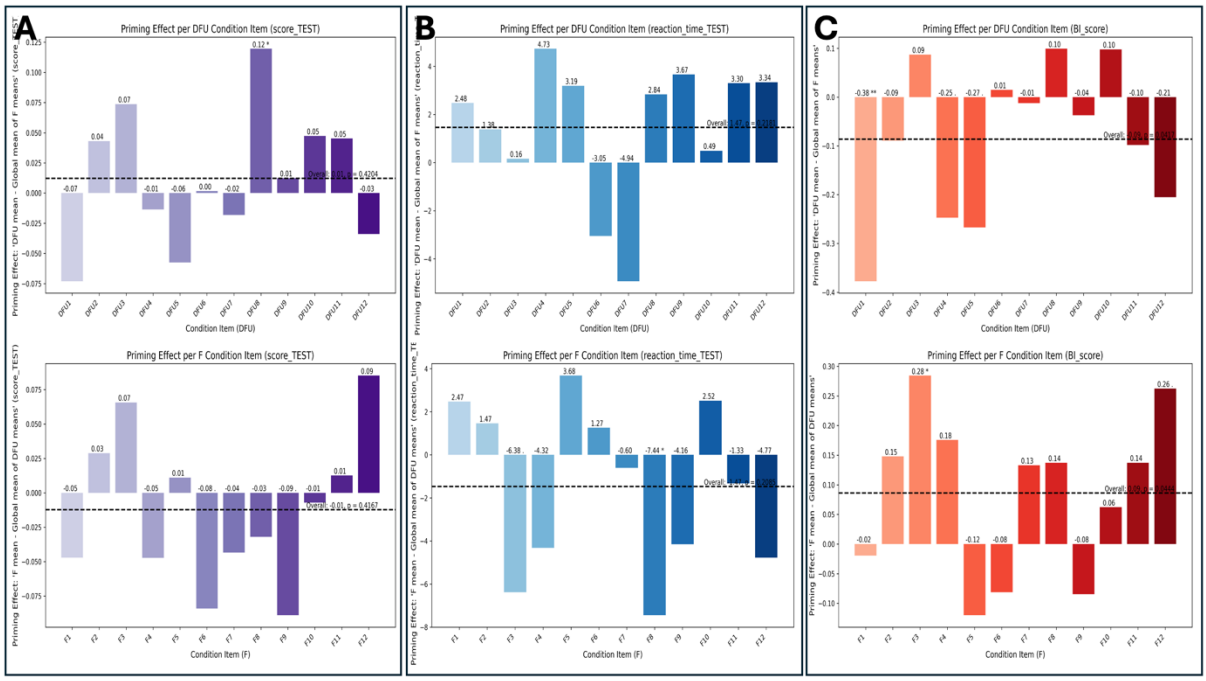
Results of reaction time distributions fitted with polynomial, ex-Gaussian and normal models



Note. A) For the classical insight tasks, a higher density for the DFU condition is seen for higher reaction times compared to the F condition (upper plot). When fitting the preprocessed reaction time data with 2-parameter Gaussian models separately for experimental conditions and task IDs, the substantial between-tasks performance can be illustrated (lower plot). B) A noticeably similar pattern is present for the RAT data where higher densities are found for the DFU condition for higher reaction times compared to the F condition.

Supplementary figure 2

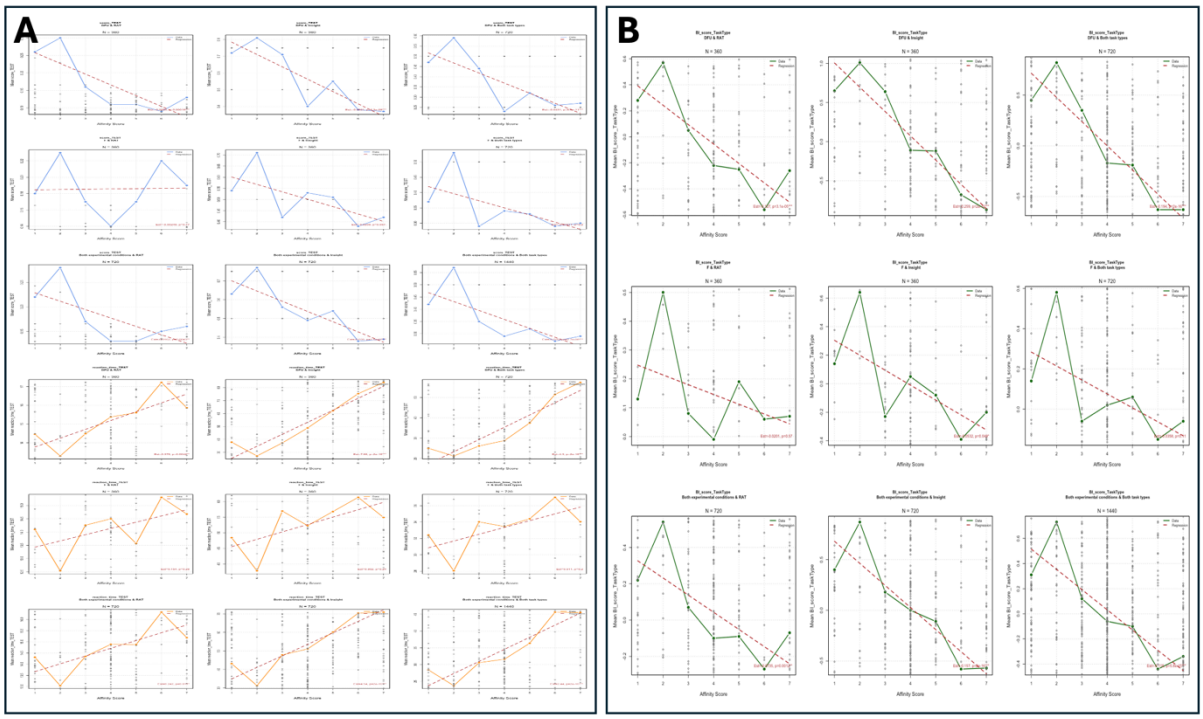
Item-level analysis illustrates variation in priming effect across different condition IDs



Note. The between-conditionIDs variation is shown in the upper and lower plots for the DFU condition and F condition, respectively. Each value was computed by taking the mean of the respective condition ID values and subtracting it by the global mean of all 12 condition ID values from the other experimental condition. An unpaired Welch's t-test was run for each separate higher-order priming effect across all three performance measures. A) For both experimental conditions, the respective mean accuracies non-significantly moved towards the direction that did not align with the hypothesis. Only 10 out of 24 items showed a pattern consistent with the hypothesis. B) For the reaction times, the predicted pattern occurred for both experimental conditions. However, two DFU items and five F items trended towards the opposite direction. C) For the BI scores, most condition IDs gravitated towards the predicted directions. On average, this was the case for the condition IDs in both experimental conditions.

Supplementary figure 3

Post-hoc analysis with mixed-effects models with non-aggregated data for all three measures



Note. Using maximal random effects structure, the statistical association between the non-aggregated affinity scores and insight performance was estimated. In each model, self-perceived verbal intelligence was included as a fixed effect. Rows vary in experimental condition (1: DFU, 2: F, 3: both); columns vary in task type (1: RAT, 2: classical, 3: both). A) For each unit increase in affinity towards unfalsifiability, the accuracy (upper blue plots) decreased by 2.68% (B=0.0268; p=1.38e-08). For each unit increase, the reaction time (bottom orange plots) increased by 2.43 seconds (B=2.427; p=2e-16). B) The overall insight performance decreased by approximately one tenth of a standard deviation for each unit increase in DFU affinity (B=-0.1146; p=3.14e-15).

Supplementary figure 4

Results of post-hoc factor analysis to evaluate latent disposition estimation techniques

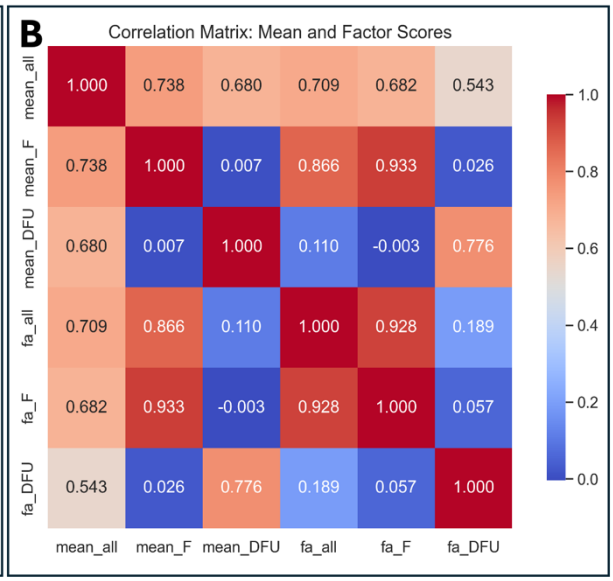
A *DFU items: PropVar = 10%*

Item ID	MR1	h2
DFU1	0.31	9.8e-02
DFU2	0.41	1.7e-01
DFU3	0.08	6.9e-03
DFU4	0.00	2.5e-06
DFU5	-0.04	1.5e-03
DFU6	0.09	8.6e-03
DFU7	-0.22	4.7e-02
DFU8	0.53	2.8e-01
DFU9	0.38	1.5e-01
DFU10	0.16	2.6e-02
DFU11	0.49	2.4e-01
DFU12	0.46	2.1e-01

F items: PropVar = 14%

Item ID	MR1	h2
F1	0.34	0.116
F2	0.47	0.223
F3	0.17	0.031
F4	0.58	0.338
F5	0.26	0.068
F6	0.52	0.275
F7	0.29	0.086
F8	0.32	0.101
F9	0.41	0.166
F10	-0.13	0.018
F11	0.10	0.010
F12	0.50	0.246

B



Note. A) Factor loadings and communalities were computed three times with one latent factor. The separate results for the 12 DFU items (left table) and 12 F items (right table) are displayed. The proportion of explained variance was considerably low for both DFU items and F items: 10% and 14%, respectively. B) A pairwise correlation matrix was constructed by computing all correlations between the three factor analysis regression scores and the three mean scores for each participant. Although a strong correlation between the two latent disposition estimation methods was found ($0.709 > r > 0.933$) for all three datasets (i.e., 12 DFU items, 12F items, and 24 DFU+F items), future research could benefit from using better condition items to derive a single latent disposition score for each participant.