No Feeling for the Future: Re-examining the Role of Emotions and Cognition in Forecasting

Thomas Y. Li
University of St. Gallen

Abstract

Our study set out to build on Pham et al.'s (2012) "emotional oracle effect," which claimed that one's trust-in-feelings (TIF) enhances forecasting accuracy by unlocking tacit knowledge. Inspired by this interesting finding, we aimed to explore how one's TIF interacts with crystallized intelligence (CI) and fluid intelligence (FI) to strengthen our understanding of the interplay between emotions and cognition when making predictions. However, a surprising discovery reshaped our investigation: the foundational TIF mechanism itself failed to replicate. Across two robust studies ($N = 226$; confirmation study $N = 432$), TIF showed negligible and inconsistent direct effect on prediction accuracy, challenging the theory's core premise. While marginally significant, results suggest TIF might improve accuracy for individuals with high CI; these effects were limited and contingent on specific contexts. Our findings undermine the generalizability of the emotional oracle effect and highlight the fragility of its theoretical foundation. The results present a cautionary tale about building on and extending theories without first thoroughly validating core mechanisms.

Keywords:

Forecasting, trust-in-feelings, TIF, cognition, intelligence, emotional oracle effect, dual-process theory, replication, strategic decision-making, predictions, superforecasters

Introduction

Throughout recorded history, humans have sought to foresee the future - from ancient Babylonians and Greeks using astrology and divination (e.g., Raphals, 2013) to modern applications in finance, politics, sports and even the weather. Although absolute certainty remains elusive, forecasting provides valuable insights that guide decision-making. Today's methods range from statistical models and scenario planning to the Delphi method and forecasting tournaments, drawing on both data-driven approaches and wisdom-of-crowds techniques (e.g., Amer et al., 2013; Linstone & Turoff, 1975; Tetlock et al., 2014).

In past decade, researchers have used forecasting tournaments to measure systematically individual abilities to forecast future events. Tetlock et al. (2014) conducted a multi-year forecasting tournament [1] and found that some individuals consistently outperformed others. Subsequent research by Mellers, Stone, Atanasov, et al. (2015) revealed significant differences between average and top-performing forecasting in areas such as probabilistic reasoning, and the ability to de-bias.

While Mellers, Stone, Atanasov, et al.'s (2015) findings provide valuable insight into cognitive capabilities that distinguish high-performing forecasters, their focus is largely confined to rational thought processes, such as logic and reasoning. This is unsurprising, given the long-standing perspective that views emotions as threats to rationality (e.g., de Sousa, 1979). This artificial dichotomy between emotion and cognition has shaped much of the decision-making literature, including the domain of forecasting, with traditional frameworks prioritizing rationality over intuition. However, emerging evidence from cognitive neuroscience and behavioral decision research challenge this view, suggesting that emotions are integral to navigating uncertainty and risk (Bechara & Damasio, 2005; Naqvi et al., 2006; Loewenstein et al., 2001).

Interestingly, some research (e.g., Pham et al., 2012; Seo & Barrett, 2007) suggests that emotional factors may also play a critical role in enhancing prediction accuracy. While

[1] Forecast tournaments invite participants to predict answers to a common set of questions over a defined period. These tournaments provide opportunities to evaluate top performers, extract wisdom from the crowd, and refine prediction methodologies (Tetlock & Gardner, 2015).

both cognitive and emotional approaches to prediction have shown promise independently, their intersection remains largely unexplored. This gap between the established logic-driven forecasting methods, and emerging evidence supporting emotion-based approaches provide an interesting opportunity to examine how these seemingly distinct mechanisms might work together to boost predictive accuracy.

Drawing on dual-process theory (Evans, 2008; Kahneman, 2011) – which distinguishes between Type 1 (intuitive, experiential) and Type 2 (analytical, deliberative) processes – this paper explores how cognitive reasoning and emotional intuition might work in concert to enhance forecasting accuracy. In doing so, we bridge two schools of thought: the cognition-focused view championed by Mellers, Stone, Atanasov, et al. (2015), which emphasizes intelligence and rationality to improve forecasts, and the emotions-driven perspective proposed by Pham et al. (2012), which positions trust-in-one's-feelings as a lever for accurate prediction. By examining each framework independently and then exploring their potential synergies, we aim to illuminate how these distinct modes of processing can combine to yield more robust forecasting outcomes.

Theory and Hypothesis Development

A study conducted by Pham et al. (2012) found that *trust* in one's feelings (TIF) can help improve the accuracy of predictions. The authors, through a series of experiments, showed that when individuals had a high level of trust in their feelings, they tend to make more accurate predictions across a wide range of forecasting tasks, such as predicting the outcome of a political nomination, the performance of the stock market, and even the weather. TIF[2] operates as a simple manipulation that increases or decreases reliance on one's gut feeling or what "feels right" (Avnet et al., 2012, p. 727; see also Avnet & Pham, 2005). This effect was observed if individuals have sufficient knowledge about the topic being predicted, and the situation is capable of being predicted (i.e., not an entirely random outcome).

[2]Pham et al. (2012, p. 463) manipulated TIF using a simple question to inquire about either two (high TIF condition) or ten (low TIF condition) situations in which "you trusted your feelings to make a judgment or a decision, and it was the right thing to do". People in the high TIF condition should easily identify two relevant situations where they were correct to trust their feelings. Thus, participants would infer that relying on their feelings would be useful. The opposite would be true for the low TIF condition because it is difficult to easily recall 10 such situations.

This phenomenon, referred to as the "emotional oracle effect", suggests that feelings can serve as informational signals that summarize what individuals learn about their environment – often outside awareness (Pham et al., 2012; Bargh & Chartrand, 1999; see also Clore & Parrott, 1994; Greifeneder et al., 2011) – and thus provide a "privileged window into all we tacitly know about the environment [around us]" (Pham et al., 2012, p. 468). Phrased differently, Pham et al. (2012) argue that TIF works through bridging conscious reasoning with accumulated tacit knowledge. Earlier work by Bechara et al. (1997) has similarly argued that humans can predict which options will lead to better outcomes before the conscious mind has identified a strategy on how to do so. While Pham et al. (2012) provided initial evidence for TIF's impact on prediction accuracy, the lack of published (or even publicly available) replications raises critical questions about its generalizability (e.g., Nosek et al., 2022), underlying mechanisms, and potential boundary conditions of the findings. Addressing this gap is the starting point of our investigation. As such, we first seek to replicate the effect of TIF through our first hypothesis.

Hypothesis 1. High TIF increases prediction accuracy

Pham et al. (2012) implicitly suggest that existing crystallized intelligence (CI) – knowledge that has been gained from education and experience and the ability to apply that knowledge in relevant contexts (Horn & Cattell, 1966, p. 254; see also Hunt, 2000) – may be key to unlocking the potential of the TIF; thereby, leading to higher prediction accuracy. Research has shown that CI serves as a cognitive foundation for both explicit and tacit knowledge, and tacit knowledge, the implicit and intuitive insights gained through experience, represents a related form of CI that is particularly relevant for tasks requiring intuitive judgment (Polanyi, 1966; Sternberg et al., 1995). As such, CI not only serves as a repository of explicit knowledge, but also underpins tacit, experience-driven insights, which are critical for intuitive judgments. TIF may then act as a bridge to access this tacit knowledge, particularly in high CI individuals.

Analyzing a group of particularly good forecasters, so-called superforecasters, representing the top 2% of all forecasters (Mellers et al., 2017), Mellers, Stone, Murray, et al. (2015) and Mellers, Stone, Atanasov, et al. (2015) found that these superforecasters scored not only higher on CI, but also on fluid intelligence (FI). FI is defined as the "use

of deliberate and controlled mental operations to solve novel problems" (McGrew, 2009; p. 5) independently of previously acquired knowledge (Jaeggi et al., 2008); FI was shown to have a stronger correlation than CI to better forecasting results.

Juxtaposing the ideas of Mellers, Stone, Murray, et al. (2015) and Mellers, Stone, Atanasov, et al. (2015) with Pham et al. (2012), a set of intriguing research questions appear. Is Pham et al.'s (2012) interpretation of their study results, that individuals with higher trust-in-their feelings have better forecasting performance only because they have more access to their CI, correct? To explicitly test Pham et al.'s (2012) conjecture, we formulate our second hypothesis.

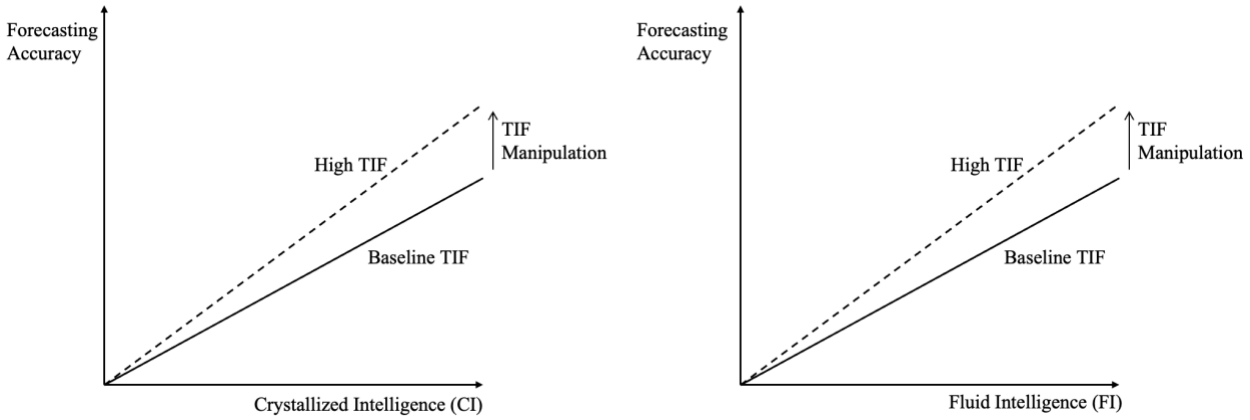
> *Hypothesis 2. High TIF positively moderates the relationship between CI and prediction accuracy, such that individuals with higher CI and high TIF make more accurate predictions*

Mellers, Stone, Murray, et al.'s (2015) and Mellers, Stone, Atanasov, et al.'s (2015) findings that the top 2% of all forecasters have higher CI suggest that as CI increases, the accuracy of the forecast improves as well; thus, indicating a linear relationship between CI and forecasting accuracy. The work from Pham et al. (2012) suggests that an increased reliance on feelings can facilitate access to knowledge, which in turn, can improve forecasting accuracy. Therefore, the implication is that the higher one's CI, the greater effect one's reliance on feelings would have, as individuals with higher CI can access a larger, existing knowledge base. This better access to knowledge should lead to higher forecasting accuracy (see Figure 1).

Given that Mellers, Stone, Murray, et al.'s (2015) and Mellers, Stone, Atanasov, et al.'s (2015) findings suggest that superforecasters not only have higher levels of CI but also higher levels of FI, is it possible that Pham et al. (2012, p. 468) interpretation that higher TIF provides "a privileged window into all we tacitly know about the environment [around us]" can also translate to an ability to access FI? Could higher trust-in-feelings facilitate access to one's FI, paralleling its hypothesized role in accessing CI? To put our inquiry formally, we propose a third hypothesis.

Hypothesis 3. High TIF positively moderates the relationship between FI and prediction accuracy, such that individuals with higher FI and high TIF make more accurate predictions

Figure 1:
Expected Impact of TIF on Intelligence and Forecasting Accuracy Relationship



Parallel to the reasoning offered around CI, an increased reliance on feelings may give individuals with higher FI a greater boost to their forecasting accuracy because they can access more of their innate FI (see Figure 1).

If our findings show that TIF neither moderates the relationship between CI and forecasting accuracy (consistent with Pham et al.'s [2012] conjecture), nor moderates the relationship between FI and forecasting accuracy, then the reason why higher TIF leads to better prediction performance would remain unsolved. While some potential alternative explanations have been proposed by Pham et al. (2012) (e.g., psi-like precognition and premonition [Bem, 2011]), the authors have refuted the theories' roles in explaining the current phenomenon. As such, our findings would require a new start to identifying the explanatory mechanism for how higher reliance on feelings leads to higher forecasting accuracy as Pham et al.'s (2012) arguments that TIF enhances access to one's knowledge would not be confirmed.

As part of this research, we want to replicate the findings of Mellers, Stone, Murray, et al. (2015) and Mellers, Stone, Atanasov, et al. (2015) that the top 2% of all forecasters have higher CI suggests and demonstrate that both higher CI and higher FI contribute to better forecasting accuracy (B. Mellers, personal communication, May 11, 2022). While the authors have shown that CI and FI, independently, are vital in supporting accurate predictions, their treatment of these cognitive capabilities as separate predictors may oversimplify CI and FI's actual utility in forecasting tasks. CI and FI likely work in tandem during prediction – CI provides the accumulated knowledge and experience necessary for a particular prediction domain, while FI enables individuals to process new information and problem-solve during the prediction tasks (Ziegler et al., 2012).

The potential synergy between accumulated knowledge, as well as reasoning and problem-solving ability could be particularly important in complex prediction tasks (e.g., Beier & Ackerman, 2005). However, the nature of this interaction has not been explicitly examined; as such, little is known about how these two types of intelligence complement each other, and what the implications for prediction accuracy are. Is it possible that having higher levels of both CI and FI leads to further combined positive effects on prediction accuracy? We devise our fourth hypothesis.

Hypothesis 4. Individuals with both high CI and high FI make more accurate predictions compared to those with either high CI only or high FI only

Should our earlier findings show that TIF positively moderates both CI's and FI's relationship with forecasting accuracy (H2, H3), and that there are combined effects of CI and FI on forecasting accuracy (H4), examining their three-way interaction becomes theoretically imperative. This interaction could reveal how one's trust-in-feelings might optimize the expression of both types of intelligence simultaneously. Given that superforecasters are characterized by both high CI and FI (Mellers, Stone, Atanasov, et al., 2015; Mellers, Stone, Murray, et al., 2015), examining how TIF moderates this combination could provide insights into whether emotional factors can enhance the cognitive advantages of superforecasters. Moreover, this three-way interaction could help uncover whether TIF serves as a general facilitator of cognitive resources or operates through more specific pathways. Would high CI and FI, which Mellers, Stone, Atanasov,

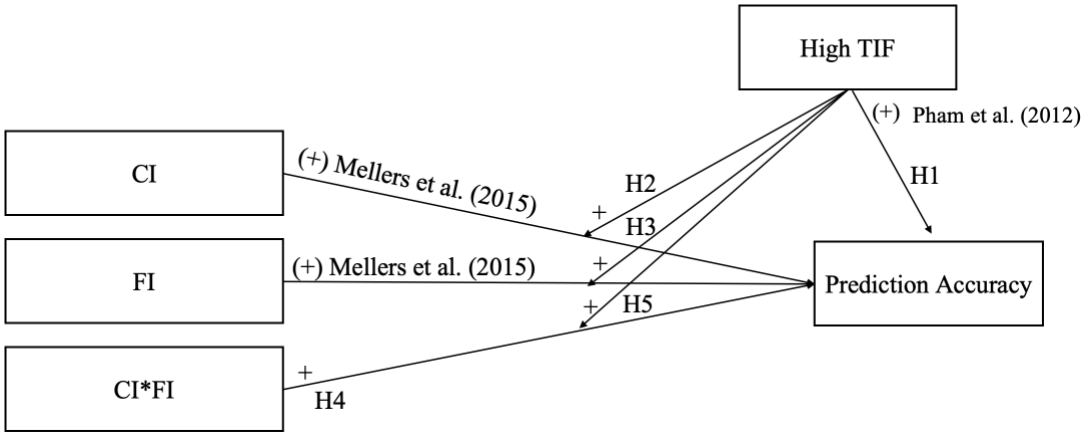
et al., (2015) and Mellers, Stone, Murray, et al. (2015) found to be characteristics of superforecasters, combined with a higher reliance on feelings, have a further additive or multiplicative effect on prediction accuracy?

Hypothesis 5. High TIF positively moderates the relationship between intelligence (CI and FI) and prediction performance, such that individuals with both higher CI and higher FI, as well as high TIF will make more accurate prediction

The results of this study are expected to pique the curiosity of several research communities. Pham et al. (2012) have called for a further investigation into the mechanism that underlies the TIF manipulations. Our research directly addresses this call: if we show that existing knowledge, measured as CI, and prediction accuracy are moderated by high TIF, we would have confirmed their speculation that high TIF provides a "privileged window into all we tacitly know about the environment" (Pham et al., 2012, p. 468). Conversely, an absence of TIF effects on CI would challenge their fundamental assumptions about how TIF operates, necessitating theoretical revisions. Furthermore, should FI emerge as the key factor driving TIF effects, it would shift attention toward problem solving ability and logical reasoning as a critical mechanism, requiring theoretical refinement and further investigation of the TIF framework.

Mellers, Stone, Atanasov, et al. (2015) and Mellers, Stone, Murray, et al. (2015) focused on identifying cognitive reasons why some people are better forecasters than others. Pham et al.'s (2012) demonstration that emotional factors (in their case trust in one's feelings) contribute to the accuracy of forecasts comes as a surprise to a cognitive approach. Therefore, replicating Pham et al. (2012) findings would provide Mellers, Stone, Atanasov, et al. (2015) and Mellers, Stone, Murray, et al. (2015) with more confidence that emotional factors should be introduced into the study of superforecasters. Given that Mellers, Stone, Atanasov, et al. (2015) and Mellers, Stone, Murray, et al. (2015) have not analyzed whether CI and FI have an additive or multiplicative effect on prediction accuracy, they would also welcome an investigation of the interaction between CI and FI. For a summary of our hypotheses, please refer to Figure 2.

Figure 2: *Overview of Hypotheses*



Methods

Overview of Study Design

A random sample of 241[3] online Prolific participants were randomly assigned to a high trust-in-feelings condition or control condition [4] between-subjects design (see Appendix A for a sample overview). The random allocation of participants was undertaken by the survey platform, Prolific. Prolific was chosen as the most suitable platform, given its participant pre-screening and verification features, as well as overall data quality (e.g., Albert & Smilek, 2023; Douglas et al., 2023). Upon closer inspection, 15 (~6%) of the participants were disqualified from the respondent pool due to incomplete responses, or clear signs of inaccuracy (detailed in a later section), resulting in a final sample size of 226 respondents. Participants were told that they would take part in multiple separate studies (tasks) and would receive a payment of 10 GBP (~12.75 USD) for their time. Participants were pre-screened to ensure eligibility for the prediction questions, following the process used by Pham et al. (2012) (see Appendix D). Data collection occurred through Prolific's sequential randomization system, which automatically assigned participants to conditions as they entered the study.

Most of the study tasks closely replicated Pham et al. (2012). First, participants in the high TIF group underwent a procedure to manipulate their reliance on feelings limited

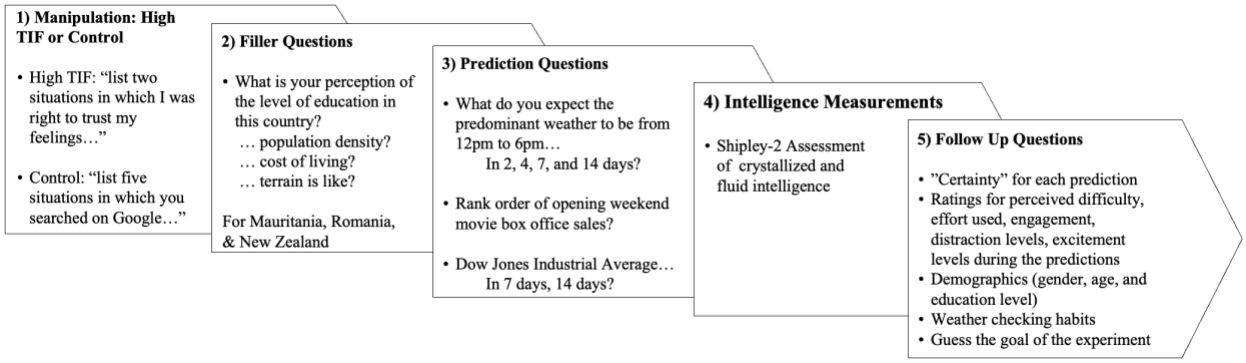
[3] 253 pre-screened participants joined the survey, 12 dropped out mid-way
[4] As the low TIF was not different from the control group in Pham et al. (2012), a low TIF group was not created

to a maximum of seven minutes (i.e., name two situations in which they were correct to trust in their feelings to make a judgment or decision). The control group was asked to complete a different task (i.e., name 5 most recent Google searches), which maintained the participants' baseline TIF levels. Following the first task, participants completed a neutral filler task that replicated the sequence and content used by Pham et al. (2012). The filler task (e.g., What is your perception of level of education for New Zealand?) is expected to reduce demand effects by masking the connection between the TIF manipulation and the subsequent prediction tasks. After the completion of the filler task, the main (third) task of the study was administered. The participants were asked to make predictions about three future scenarios, of which the accuracies of these predictions were the dependent variables in our study.

Next, for the fourth task, the participants were directed to complete a set of standardized assessments to measure both CI and FI (more details follow). The intelligence assessments were intentionally administered after the prediction tasks for two key reasons. First, this sequencing allowed for a clean replication of Pham et al.'s (2012) methodology before introducing additional measures. Second, while CI and FI are stable traits unlikely to be affected by the experimental procedure, exposure to intelligence tests prior to predictions could influence participants' emotional states or confidence levels, potentially contaminating the TIF manipulation.

At the completion of the fourth task, participants were asked a series of debriefing questions, including the level of task involvement (e.g., "I put effort into this task") (Pham et al., 2012, p. 465) as a confounding check, and a question to guess the goal of the study for an experiential demand check. Moreover, the participants were asked to rate their confidence in the predictions on a 1 (very unsure about the [result]) to 5 (almost certain about the [result]), as a manipulation check for the TIF lever. Additional questions regarding demographics were also asked. The time each participant takes to make every prediction was also recorded to ensure that the TIF manipulation does not manipulate effort. At the end of the experiment, participants were thanked and paid for their time (see Figure 3 for an overview of the study design, and Appendix B for screenshots of the survey administered during the study).

Figure 3: *Overview of Study Design*



Measures

Independent Variables

CI and FI. The CI and FI of participants were the independent variables. During the third task, aptitude tests for both CI and FI were completed by the participants. Drawing from prior work, (e.g., Husin et al., 2013; Kaya et al., 2012; Mellers, Stone, Atanasov, et al., 2015), the Shipley-2 assessment (Shipley et al., 2009) was given to assess CI and FI. The Shipley-2 Vocabulary sub-assessment, a test that requires a respondent to read a target word and choose, from four options, the word closest in meaning to the target word, explicitly measured CI. The raw score for CI was the Vocabulary sub-assessment score out of 43 points[5] based on the standard Shipley-2 assessment, and three tailored vocabulary groups for the movie prediction question. The Shipley-2 Abstraction assessment, containing nonverbal block patterns measured a participant's reasoning ability, namely FI. The raw score for FI was the Abstraction sub-assessment score out of 25 points. For an extensive list of the test questions, see Appendix B, Sections 24 to 27.

Trust-in-Feelings (TIF) Manipulation. Participants were first given a brief overview of the difference between using feelings compared to facts and logical reason when making decisions. The participants were then asked to describe two "situations in which you trusted your feelings to make a judgment or decision, and it was the right thing to do". It is important to note that participants were explicitly asked to list situations that

[5]We added three additional questions (based on Logline, Montage, Diegesis) as a proxy to measure the domain knowledge of movie buffs/enthusiasts. This added three points to the standard 40-point Shipley-2 CI Assessment.

had indeed happened. Participants were given seven minutes to complete the task. Please see Appendix C.1 and Appendix C.2, respectively, for detailed instructions for the two conditions: high TIF, and no TIF (control).

Dependent Variables

The method for measuring prediction accuracy was adopted from Pham et al. (2012) to evaluate the performance of the participants in the study. We compared the predictions made by each participant with the actual outcomes of the events (e.g., predicted weather in 2 days vs. actual weather in 2 days). The correct response was deemed to be highly uncertain at the time the questions are given to the participant. Each prediction of every participant was coded for their accuracy as described in detail in the next few paragraphs.

Local Weather Forecasting (Prediction 1). Study participants were tasked with forecasting the predominant weather conditions in their local area (identified by zip code) for several different time horizons in the future: 2 days, 4 days, 7 days, and 14 days, specifically focusing on the weather between 12pm and 6pm[6] for each respective day. For each prediction, participants selected from seven weather categories: a) sunny/fine/clear, b) partially sunny with some clouds, c) cloudy/overcast, d) haze/fog e) thunderstorms, f) rain, and g) snow. To evaluate forecast accuracy, we compared participants' predictions with actual meteorological conditions that occurred at each participant's specific location during the designated timeframe of 12pm to 6pm. A binary scoring system was implemented, whereby accurate predictions received a score of "1", while mismatches between predicted and actual conditions were marked as "0". We then conducted comparative analyses between the experimental and control groups for each prediction across the different points in time. Participants were asked a supplementary question regarding their weather-checking behavior for their respective local region to check for balance distribution of this potential confound across conditions.

Opening Weekend Box Office (Prediction 2). The second prediction question examined the participants' ability to forecast the relative opening weekend box office

[6] Pham et al. (2012) relied on the "predominant" daily weather as their source of truth; however, the authors could not specify exact source or platform used for their data. We chose the 12pm to 6pm time window given availability of data

results for three movies scheduled for the April 14, 2023 weekend in the domestic US market: Nefarious, Mafia Mamma, and Sweetwater. All participants were provided simple descriptions to guide their prediction; these included a) the movie's artwork (i.e., advertising poster), b) a summary of the plot (from IMDB), c) the genre (e.g., thriller), and d) the MPAA rating (e.g., G). Participants were instructed to arrange these three films in descending order based on their anticipated opening weekend sales. If the rank order of all three movies were accurate, the prediction was labeled as correct (binary variable of "1"), while any deviation from the correct order, including partially accurate rankings were labeled as incorrect (binary variable of "0"). Movies were chosen with the same relative equal pre-release popularity to avoid a potential clear crowd favorite (e.g., Superman 4).

Dow Jones Industrial Average Index (Prediction 3). Prior to the study, all participants underwent initial screening (Appendix D) through a basic financial knowledge test that assessed their understanding of key economic concepts (e.g., GDP and price-to-earnings ratio). The study was conducted during the first half of 2023, a period marked by substantial economic uncertainty and market volatility (IMF, 2023). Participants were asked to predict the Dow Jones Industrial Average Index (DJI) close values at two future points: 7 days and 14 days from the study date. To ensure all participants had access to relevant information, they were provided with background material about the stock market and the Dow Jones, following the approach used by Pham et al. (2012) (Appendix B, Section 23). The accuracy of the predictions was measured by calculating the absolute difference between the participants' predicted DJI values and the actual closing values on each target date (i.e., lower the difference is better).

Sample Size and Power Analysis

To ascertain the minimum number of participants required for the experiment, we calculated the ideal sample size for the three types of analyses we wanted to do.

ANOVA

Using a standard setting for behavioral sciences (statistical power = .80, medium effect size $f = .25$, $\alpha = .05$, $df = 2$), G*Power 3.1 (Faul et al., 2007) indicated the ideal sample size to be 128. This method was used to compute the direct effects of TIF, for prediction questions with continuous outcomes (i.e., DJI prediction).

Chi-Square Test

G*Power 3.1 analysis indicated a minimum total sample size of 88 for a standard statistical setting (statistical power = .80, medium effect size $w = 0.3$, $\alpha = .05$, $df = 3$). This method was used to compute the direct effects of TIF for prediction questions with binary outcomes (i.e., weather and opening weekend movie box office).

Regression

For logistic (weather and opening weekend box office) and hierarchical (DJI) regression analyses, G*Power 3.1 indicated a minimum total sample size requirement of 134 for a standard statistical setting (statistical power = 0.8, medium effect size of $f^2 = .10$, $\alpha = .05$, predictors = 5). The regression methods were used to compute the precise effects of our IVs on the DVs. Given our control variables (age, gender, and education), the minimum total sample size increases to 159. As described previously, domain knowledge, a boundary condition of Pham et al.'s (2012) study, was controlled for in a pre-screening questionnaire given to participants.

While demographic variables (age, gender, education) were collected, they were not included as controls in our analyses for several reasons. First, our randomization check confirmed no meaningful differences in demographic characteristics between TIF and control conditions, except for minor imbalances in certain education categories (e.g., 9 vs. 1 PhDs, and 8 vs. 1 trade school degrees). However, these differences were both small in absolute numbers and theoretically inconsequential given previous research. Second, this approach maintains methodological consistency with Pham et al.'s (2012) original study. Third, and most importantly, previous research on forecasting accuracy found these demographic variables, including education level, to be non-significant predictors of prediction performance (Mellers, Stone, Atanasov, et al., 2015), suggesting their inclusion as controls would unnecessarily complicate our models without theoretical or empirical justification. Our final sample size of 226 exceeds all minimum sample size requirements across analyses, ensuring adequate statistical power for direct effects.

Data Analysis

Data preparation

Before conducting our main analyses, we implemented a comprehensive four-step data cleaning process. First, we reviewed response completeness and excluded any submission with less than 99% survey completion rate. This threshold accounts for cases where participants may have completed the entire survey but forgot to click the final "finish" button before returning to Prolific for their completion code. Second, all CI and FI responses for the first five questions were manually checked. Given the extreme easiness of the first five questions, no participants should fail any of the five questions. This helped to ensure that participants were attentive during the intelligence tests. Third, we identified and removed problematic responses, including those with unreasonable completion times (such as surveys taking over 60 minutes) and highly unrealistic predictions (e.g., predictions of 50,000 for the DJI, which is virtually impossible). Lastly, we excluded participants who provided incomplete or truncated responses to the TIF manipulation. This final criterion was based on two key considerations: a) participants who couldn't complete the manipulation within the seven minutes were unlikely to achieve the desired high TIF state, and b) participants exceeding the seven-minute limit might experience increasing anxiety about potential payment rejection (from Prolific) due to an incomplete response; thereby, influencing their emotional state throughout the remainder of the study.

Data Analysis Overview

We implemented a multi-method approach to rigorously test our hypotheses. For outcomes with binary variables (weather and movie prediction), our approach combined chi-square tests – the method used to Pham et al. (2012) to analyze direct effects of TIF – with logistic regression analyses, to better isolate the impact of every IV. When analyzing continuous outcome variables (DJI predictions), we employed a combination of ANOVA (replicating Pham et al.) and hierarchical regression methods – again, to better isolate the impact of every IV. This dual analytical strategy serves multiple purposes: while chi-square tests and ANOVAs maintain methodological consistency necessary for direct replication, the regression analyses enable examination of more complex patterns while controlling for multiple predictors simultaneously. The regression approach is particularly crucial given

the theoretical relationships between CI and FI (Schneider & McGrew, 2018), allowing us to account for their potential interdependencies.

Direct effects of TIF (H1, replication of Pham). Replicating Pham et al.'s (2012) methodology, we conducted chi-square tests of independence to examine the relationship between trust-in-feelings (TIF vs. control [no TIF]) and prediction accuracy (correct vs. incorrect) across key domains: weather predictions at varying time horizons (2, 4, 7, and 14 days in the future), and opening weekend movie box office predictions. For DJI predictions at both 7-day and 14-day horizons, we employed one-way ANOVA[7]s to examine TIF's direct effect, consistent with Pham et al.'s (2012) approach for continuous outcome variables. Our direct replication of methods allows us to accurately measure the efficacy of TIF using the same tests as Pham et al. (2012). Moreover, when using chi-square tests, we assessed cross-tabulation frequencies and percentages, Pearson's chi-square statistics (χ^2), effect size through Cramer's *V*, as well as associated significant levels. For ANOVAs, we assessed the *F*-statistic, Levene's test for homogeneity of variance, the Eta-squared (η^2).

Two-way and Three-Way Interactions (H2 to H5). We employed both logistic and linear regression analyses, both hierarchical in nature, to investigate the direct effects and interactions among intelligence, trust-in-feelings, and their combinations (i.e., CI, FI, TIF, CI*FI, CI*FI*TIF). For binary outcomes (weather and movie box office predictions), we utilized binary logistic regression, an approach particularly suited for examining dichotomous outcomes with multiple predictors and their interactions (Hosmer et al., 2013, p. 35, 73; Peng et al., 2002). For continuous outcomes (DJI predictions in 7 or 14-days), we employed linear regression.

We utilized a blocking strategy in both regression approaches. Block 1 entered main effects (TIF, CI, FI)[8] to establish baseline relationships between individual predictors and outcomes. Block 2 introduced two-way interactions (TIF*CI, TIF*FI, CI*FI) to test our core theoretical propositions about how emotional and cognitive factors might work

[7] ANOVA replaced ANCOVA as pre-screening only allowed those with basic economic knowledge to participate, rather than controlling for it through statistical means

[8] All CI and FI values used during both logistic and linear regressions were mean-centered

together. Finally, Block 3 added the three-way interaction (TIF*CI*FI) to examine whether TIF moderates the joint expression of different forms of intelligence. This sequential approach (e.g., Goktan & Miles, 2011; Ismail, 2005) allows us to isolate specific theoretical mechanisms while controlling for lower-order effect (Hayes, 2017) and explore whether more complex interactions explain additional variances beyond simpler effects, providing a more nuanced test of our theoretical framework.

For the logistic regression, we assessed model improvements through Nagelkerke R^2, evaluated individual predictors using Wald statistics and p-values, as well as examined effect sizes, if any, via odds ratio (*Exp(B)*) with 95% confidence intervals. For the hierarchical regression, we examined model improvements through R^2 and ΔR^2 changes across blocks, assessed overall model fit using F-statistics, and evaluated individual predictors through coefficients (B) with corresponding significance levels.

Following best practices (e.g., Aiken, 1991; Dawson, 2014), we employed mean centering to reduce nonessential collinearity between interaction terms and their constituent variables. Some collinearity was anticipated as measures of FI and CI are typically positively correlated. (e.g., Cunningham et al., 1975; Stamenković & Holyoak, 2018). We measured multicollinearity through variance inflation factors (VIF) (Craney & Surles, 2002; Hayes, 2017, p. 305). After mean-centring, VIF values were expected to remain below the recommended threshold of 3.0, (Hair et al., 2019), addressing any initially high pre-centred values.

Confirmation Study: TIF Direct Effects

Given the surprising and inconclusive results of our main study regarding TIF's direct effect, we conducted a high-powered confirmation study (N = 432) to further investigate the efficacy of the TIF manipulation and its corresponding theoretical foundations. This confirmation study aimed to replicate Pham et al.' (2012) opening weekend box office[9] prediction task, which showed one of the highest effect sizes in their original work ($\chi^2 = .57$).

[9] Movies (released on Feb 23, 2024): Ordinary Angels, Demon Slayer: Kimetsu no Yaiba - To the Hashira Training, and Drive-Away Dolls.

Participants and Design

A total of 432 participants were recruited through Prolific. Participants were randomly assigned to either a high TIF condition ($N = 213$) or a control condition ($N = 219$) in a between-subjects design.

Procedure

The procedure followed the same protocol as the main study. We specifically selected a single prediction task to maintain even closer alignment with Pham et al.'s (2012) original methodology, where each experiment focused on one prediction question. However, we enhanced the experimental design by adding one key question at the study's conclusion: participants were asked about their movie preferences regarding the three movies they had evaluated. This question was placed after the prediction task to avoid contaminating participants' predictions by first priming them to consider their personal preferences.

Results

Descriptives

Our final sample ($N = 226$) consisted predominantly of male participants (65.93%), with females (31.42%) and other gender identities (2.65%) comprising the remainder. The majority of participants were between 26-35 years old (37.17%), followed by 36-45 (19.47%), 56+ (17.26%), 46-55 (15.04%), and 18-25 (11.06%). Regarding education, most participants held a bachelor's degree (52.21%), followed by High School diploma (20.80%), master's degree (17.70%), Ph.D. (4.42%), and trade school qualifications (3.98%).

Independent samples t-tests confirmed no significant differences between TIF and control conditions across numerous measures during the prediction process: confidence, difficulty, effort used (e.g., "I put effort into this task") (Pham et al., 2012, p. 465), engagement, distraction levels, and excitement levels (all ps $> .05$). Demographics were largely balanced across conditions for gender and age, with only minor, inconsequential differences in certain education categories (9 vs. 1 PhDs and 8 vs. 1 trade school degrees in control vs. TIF conditions, respectively). Just as important, no differences were observed in weather-checking behavior prior to the study (i.e., checking weather 2 days in the future) or prediction completion times between the two conditions.

TIF Direct Effects, Replication of Pham et al. (2012) (H1)

We conducted a chi-square test of independence to examine the relationship between TIF and prediction accuracy across the multiple domains. All chi-square analyses met the minimum expected cell frequency assumption. For weather predictions, contrary to expectations, we found no significant relationship between TIF and prediction accuracy across any of the time horizons: 2-day forecasts ($\chi^2(1) = 0.003$, $p = .957$, $V = .004$), 4-day forecasts ($\chi^2(1) = 2.806$, $p = .094$, $V = .112$), 7-day forecasts ($\chi^2(1) = 0.001$, $p = .977$, $V = .002$), as well as 14-day forecasts ($\chi^2(1) = 0.011$, $p = .915$, $V = .007$).

The prediction accuracy rates between high TIF and control conditions were largely similar across all weather conditions, with only 4-day forecast showing a marginally significant difference (28.3% accuracy in high TIF vs. 18.8% in control). The results did not replicate Pham et al.'s (2012) results that high TIF would yield more accurate 2-day weather predictions, suggesting that the relationship between TIF and weather prediction may be more tenuous than previously thought (see Table 1).

For opening weekend movie box office predictions, results indicated only a marginally significant relationship between TIF and prediction accuracy ($\chi^2(1) = 2.801$, $p = .094$, $V = .110$), not fully replicating Pham et al. (2012). Participants in the high TIF condition showed higher accuracy rates (20.8%) compared to the control condition (12.5%).

For DJI predictions, we conducted one-way ANOVAs to determine the effect of TIF on prediction accuracy at both 7-day and 14-day horizons. Levene's test indicated equal variances for both 7-day ($F(1, 224) = 0.190$, $p = .664$), and 14-day predictions ($F(1, 224) = 0.315$, $p = .575$). The results showed no significant effect of TIF on prediction accuracy for either the 7-day horizon ($F(1, 224) = 0.012$, $p = 0.912$, $\eta^2 < .001$) or the 14-day horizon ($F(1, 224) = 0.043$, $p = .836$, $\eta^2 < .001$). Again, we could not replicate Pham et al. (2012) who found a significant relationship between high TIF and the accuracy of 7-day DJI predictions. Given violations of normality in the DJI prediction differences (Shapiro-Wilk: $p < .001$), we conducted supplementary non-parametric analyses. A Mann-Whitney U test confirmed the absence of TIF effects ($U = 6243.00$, $p = .811$), supporting the robustness of our ANOVA findings.

Table 1: *Comparison of Results (TIF Direct Effects) vs. Pham et al. (2012)*

Prediction	Replication?	Context	Sample Size	Prediction Horizon	High TIF (Accurate)	Control (Accurate)	Method	Results	Effect Size	Pham et al.'s (2012) Result	Effect Size
1	Yes	Weather	226	2 days	23.60%	23.30%	Chi-Square	**Not significant, $\chi^2 = 0.003$**	0.004	**Significant**	0.48
2	Yes, with extended horizon	Weather	222	4 days	28.30%	18.80%	Chi-Square	Marginally Significant, $\chi^2 = 2.806$	0.112	-	-
3	Yes, with extended horizon	Weather	223	7 days	19.80%	19.70%	Chi-Square	Not significant, $\chi^2 = 0.001$	0.002	-	-
4	Yes, with extended horizon	Weather	223	14 days	16.00%	15.50%	Chi-Square	Not significant, $\chi^2 = 0.011$	0.007	Not significant	-
5	Yes	Movie box office	222	3 days	20.80%	12.50%	Chi-Square	**Marginally Significant, $\chi^2 = 2.801$**	0.110	**Significant**	0.57
8	Yes	Movie (confirmation study)	432	3 days	14.20%	19.20%	Chi-Square	**Not significant, $\chi^2 = 2.017$**	0.068	**Significant**	0.57
6	Yes	Dow Jones Index	226	7 days	826 points	802 points	ANOVA	**Not significant, $F = 0.012$**	0.000	**Significant**	0.36
7	Yes, with extended horizon	Dow Jones Index	226	14 days	1233 points	1183 points	ANOVA	Not Significant, $F = 0.043$	0.000	-	-

Note: bolded results indicate discrepancy between Pham et al. (2012) and our results

Participants in the high TIF condition for 7-day predictions (M = 826.13, SD = 2063.87) did not differ significantly in their prediction accuracy from those in the control group (M = 801.88, SD = 1170.25). The results are similar for 14-day predictions, high TIF participants (M = 1233.11, SD = 2191.00), showed no significant differences from the control group (M = 1182.62, SD = 1423.09). While these differences did not reach statistical significance, it is worth noting that the control condition demonstrated both lower mean differences and smaller standard deviations across both time horizons, suggesting more consistent and accurate predictions compared to the high TIF group.

Using comparable statistical analyses to Pham et al. (2012) – chi-square tests for categorical outcomes (movie and weather predictions), and ANOVA for continuous outcomes (DJI predictions) – our study was unable to replicate the robust emotional oracle effect reported in their paper. Of the three main predictions that demonstrated significant effects in Pham et al. (2012), only one showed marginal significance in our study (movie box office, p = .094), while the remaining two showed no significant differences between high TIF and control conditions. Of the marginally significant results, the effect size was substantially lower than that of Pham et al.'s (2012) study. For a direct comparison of Pham et al. (2012)'s studies and those of ours (see Table 1 above).

Exploration of Two-Way, and Three-Way Interactions (H2-H5)

To further investigate the relationship between TIF, CI, FI, and prediction accuracy, we conducted a binary logistic regression analysis. The model was built in three steps: first entering the main effects (TIF, CI, FI), then two-way interactions (TIF*CI, TIF*FI, CI*FI), and finally the three-way interaction (TIF*CI*FI). Across all regression models reported below, VIF values remained below 3.0 after mean-centering, indicating no multicollinearity concerns.

Weather Predictions

For 2-day weather predictions, block 1 showed no significant direct effects of TIF (*Exp(B)* = 1.017, p = .957), CI (*Exp(B)* = 0.988, p = .805), or FI (*Exp(B)* = 1.018, p = .684). Block 2 revealed a marginally significant TIF*CI interaction (*Exp(B)* = 1.226, p = .059), while the other two-way interactions, TIF*FI (*Exp(B)* = 0.984, p = .860) and CI*FI (*Exp(B)* = 0.977, p = .105), remained non-significant. The addition of the three-way interaction,

TIF*CI*FI, in Block 3 did not improve the model fit (*Exp(B)* = 0.963, *p* = .202), though the TIF*CI interaction remained marginally significant (*Exp(B)* = 1.209, *p* = .087). The final model explained only 6.1% of the variance (R^2 = .061).

For 4-day weather predictions, Block 1 showed a marginally significant direct effect of TIF (*Exp(B)* = 1.739, *p* = .086), while CI (*Exp(B)* = .962, *p* = .437) and FI (*Exp(B)* = .963, *p* = .390) showed no significant effects. The TIF effect remained marginally significant in both Block 2 (*Exp(B)* = 1.831, *p* = .067) and Block 3 (*Exp(B)* = 1.811, *p* = .079). No significant two-way interactions emerged in Block 2 (all *p*'s > .496), and the addition of the TIF*CI*FI three-way interaction (*Exp(B)* = 1.004, *p* = .879) did not improve the model fit. In Block 3, the model only explained 4.1% of the variance (R^2 = .041). None of the other variables, two-way or three-way interactions were marginally significant or significant.

For 7-day weather predictions, Block 1 showed a marginally significant negative direct effect of FI (*Exp(B)* = .922, *p* = .085). In Block 2, the FI effect strengthened to be statistically significant (*Exp(B)* = 0.831, *p* = .014) and a negative significant CI*FI interaction emerged (*Exp(B)* = .960, *p* = .011). In Block 3, the addition of the TIF*CI*FI interaction (*Exp(B)* = .997, *p* = .931) did not improve model fit, while the FI effect remained significant (*Exp(B)* = .832, *p* = .016) and the CI*FI interaction became marginally significant (*Exp(B)* = .961, *p* = .063). The final model explained 9.8% of the variance (R^2 = .098). None of the other variables, two-way or three-way interactions were marginally significant or significant.

For 14-day weather predictions, no significant effects emerged across any of the blocks. Neither main effects (all *p*'s > .669 in Block 1), two-way interactions (all *p*'s > .458 in Block 2), nor the three-way interaction (*Exp(B)* = .983, *p* = .552 in Block 3) reached significance. The final model only explained 1.2% of the variance (R^2 = .012).

Movie Box Office Predictions

The setup of the binary logistic regression is identical to that of the weather predictions, where the model was built in three steps. Results in Block 1 showed no significant effects of TIF (*Exp*(B) = 1.805, *p* = .108), CI (*Exp*(B) = 1.034, *p* = .580), or FI (*Exp*(B) = 1.053, *p* = .316). In Block 2, TIF showed marginal significance (*Exp*(B) = 1.887, *p* = .093), and a marginally significant CI*FI interaction emerged (*Exp*(B) = 1.021, *p* = .090),

while other two-way interactions, TIF*FI (*Exp*(B) = .950, *p* = .636) and TIF*CI (*Exp*(B) = .970, *p* = .812), were non-significant. In Block 3, TIF maintained marginal significance (*Exp*(B) = 2.045, *p* = .085), while all other effects, including the CI*FI interaction (*Exp*(B) = 1.027, *p* = .126), and the TIF*CI*FI three-way interaction (*Exp*(B) = 0.989, *p* = .633), were non-significant. The final model explained only 6.3% of the variance (R^2 = .063).

Dow Jones Index Predictions

For 7-day DJI predictions, we conducted a hierarchical linear regression analysis to further test our hypotheses. Block 1 showed a marginally significant effect of CI (*B* = -68.582, *p* = .051), while TIF (*B* = 42.545, *p* = .846) and FI (*B* = -14.645, *p* = .629) showed no significant effects. In block 2, both TIF*CI (*B* = -120.423, *p* = .087), as well as CI*FI (*B* = -14.627, *p* = .059) approached marginally significance, while TIF*FI remained non-significant (*B* = 65.735, *p* = .276), and CI becomes non-significant. In Block 3, only TIF*CI remained marginally significant (*B* = -132.811, *p* = .068), while all other effects, including TIF*CI*FI three-way interaction (*B* = -10.490, *p* = .499), were non-significant. The final model only explained 2.2% of the variance (adjusted R^2 = .022).

For 14-day DJI predictions, Block 1 showed no significant effects for TIF (*B* = 70.881, *p* = .770), CI (*B* = -59.628, *p* = .125), or FI (*B* = -33.102, *p* = .324). In Block 2, TIF*CI (*B* = -135.460, *p* = .082) and CI*FI (*B* = -16.110, *p* = .060) approached marginal significance, while TIF*FI (*B* = 55.182, *p* = .408) was non-significant. In Block 3, only TIF*CI maintained marginal significance (*B* = -155.747, *p* = .053), while all other predictors, including the TIF*CI*FI three-way interaction (*B* = -17.178, *p* = .317), were non-significant. The final model explained only 2.2% of the variance (adjusted R^2 = .022). For an overview of all Block 3 findings, see Appendix E.

Confirmation Study Results

Initial analysis of the full sample showed no significant differences in prediction accuracy between the TIF group (*N* = 213; 14.16% correct), and the control group (*N* = 219; 19.24% correct; $x^2(1) = 2.017$). Cramer's post-hoc test confirmed this non-significance (*V* = .07, *p* = .17). Notably, the direction of the full sample result ran counter to the predicted TIF advantage, with the control group achieving higher accuracy (19.24%) than the TIF

group (14.16%), though this difference did not reach significance. The confirmation study's results failed to replicate the emotion oracle effect reported by Pham et al. (2012).

However, when analyzing only participants who expressed no preference for any of the three movies, a different pattern emerged. Amongst these indifferent participants, the TIF $group_{no\ pref}$ (N = 77, 21.42% correct) significantly outperformed the control $group_{no\ pref}$ (N = 95; 8.42% correct; $x^2(1) = 6.39$, $p < .05$; $V = .18$, $p < .05$).

Discussion

Our study aimed to explore the underlying mechanisms of the emotional oracle effect (Pham et al., 2012) by examining how trust-in-feelings (TIF) facilitates access to predictive information. Building on Pham et al.'s (2012, p. 468) arguments that feelings provide "a privileged window into all we tacitly know about our environment", we hypothesized that this 'tacit knowledge' could be represented by CI – knowledge gained through education and experience. Additionally, we explored whether fluid intelligence (FI) – the ability to solve novel problems independently of acquired knowledge – might also become more accessible when individuals trust their feelings. Our study had two main intentions. First, we aimed to replicate Pham et al.'s (2012) findings for the direct effect of TIF on prediction accuracy. Second, we sought to extend Pham et al.'s (2012) findings by examining the interplay between TIF, CI, and FI in prediction accuracy across the multiple domains. Our results present an intriguing contrast to Pham et al.'s (2012) findings.

Replication of Pham et al. (2012)

Unexpectedly, before we could fully examine the underlying accessibility mechanisms of the TIF lever, we encountered a more fundamental challenge: we were unable to fully replicate the foundational emotional oracle effect. Across multiple prediction domains – weather, movie box office, and stock market predictions – the support for the direct effect of TIF on prediction accuracy was either weak or notably absent. This was both surprising and unexpected, given that our theoretical framework was built directly on Pham et al.'s (2012) compelling explanation of how TIF, alone, can increase prediction accuracy across multiple domains. Pham et al. (2012) claimed that the emotional oracle effect was not only limited to socially-determined outcomes (i.e., elections or American Idol, or perhaps even the stock market if we assume humans do not make purely rational

decisions) but extended to diverse prediction domains including acts of nature (weather) – suggesting that TIF provided broad access to predictive information across various domains. Yet, across these same types of predictions in our study, most notably weather and DJI, we were largely unable to replicate these effects, even if we consider marginally significant effects.

We did observe some marginal effects that warrant mentioning. For movie box office predictions, TIF showed marginal significance across multiple blocks of analysis (Block 2: *Exp*(B) = 1.887, $p = .093$, Block 3: *Exp*(B) = 2.045, $p = .085$), suggesting that trusting one's feelings might still play some role in helping to predict future outcomes of opening weekend movies. Similarly, for 4-day weather predictions, a time horizon not tested in Pham et al.'s (2012) study, we found marginal effects of TIF across all blocks (Block 3: *Exp(B)* = 1.811, $p = .079$). However, the results are inconclusive and warrant further exploration.

Confirmation Study: TIF Direct Effect (H1)

The confirmation study reinforces concerns about the robustness of the emotional oracle effect while introducing an additional puzzle. In the full sample ($N = 432$), we again failed to replicate Pham et al.'s (2012) findings, despite designing the study to provide favourable conditions for detection – a larger sample, a single prediction domain that showed strong effects in their previous work, and virtually identical methodological alignment. Notably, a significant TIF effect emerged only among participants who expressed no preference for any of the three movies (21.42% vs 8.42% group accuracy, $p < .05$).

This finding, combined with our main study results, forces us to reconsider the theoretical foundation proposed for the emotional oracle effect. Pham et al. (2012) had investigated two potential explanations for their findings: a social attunement hypothesis, where TIF might enhance attunement to collective behaviour, and a privileged-window hypothesis, where feelings provide access to tacitly learned information about the environment. Their results – not replicated by any other author – seemed to support the privileged-window hypothesis, as the effect extended beyond socially determined outcomes to include natural phenomena. Our results challenge both explanations of TIF. If trust-in-

feelings operates through social attunement, we would expect strong effects in movie predictions, where collective behaviour plays a clear role. If TIF operates as a privileged-window to tacit knowledge, we would expect consistent effects for weather predictions, at least the 2-day prediction window, rather than the isolated marginal effect at 4 days that we observed.

Furthermore, our confirmation study highlights a tension with the privileged-window account. If TIF enhances predictions by providing access to accumulated knowledge, one might expect participants with stronger preferences, who presumably possess more domain-relevant knowledge and opinions, to benefit most from trusting their feelings. Instead, TIF appeared effective only among those expressing no preference for movies, suggesting that pre-existing preferences may override or contaminate feeling-based signals that TIF is theorized to activate. This insight raises the possibility that TIF's utility, where it exists, may depend on the absence of strong prior attitudes rather than the presence of deep domain knowledge. Nonetheless, both our main and confirmation study provide no substantial support for the two hypotheses from Pham et al. (2012) and creates more questions than it answers.

TIF and CI Interaction (H2)

The TIF*CI interactions showed a consistent pattern of marginal significance across both 2-day weather predictions and DJI forecasts, with the pattern being most notable in DJI predictions (7-day: B = -132.811, p = .068; 14-day: B = -155.747, p = .053). The negative coefficient of the DJI results indicates that as CI increases, TIF leads to smaller differences between predicted and actual DJI values – in other words, better prediction accuracy. A similar pattern emerged for 2-day weather predictions, where TIF*CI interaction showed marginal significant (Block 2: *Exp(B)* = 1.226, p = .059; Block 3: *Exp(B)* = 1.209; p = .087). Both effects point in the same direction, where TIF combined with higher CI was associated with better prediction accuracy; however, the DJI pattern was more pronounced.

While these findings are tentative given their marginal significance, they align with our theoretical framework: TIF may enhance access to stored knowledge (CI), providing some support for Pham et al.'s (2012) proposition about feelings providing a 'privileged

window' into tacit knowledge. These TIF*CI interactions, while modest, are particularly interesting given our failure to replicate the direct TIF effect, suggesting that TIF might only improve prediction accuracy when supported by vast amounts of domain knowledge.

This pattern of findings suggests the possibility of a more nuanced understanding of the emotional oracle effect. Rather than TIF alone providing predictive power, feelings might serve as a mechanism to access and utilize stored knowledge. This interpretation would align with research on expert intuition (Kahneman & Klein, 2009), which suggests that intuitive judgments are most accurate when based on extensive domain experience in a predictable environment. The marginal significance of this interaction persisting across different prediction horizons (7-day and 14-day) hints at a potentially robust pattern, even as the prediction task becomes more challenging with longer time horizons. While these findings require further validation, their consistency across time horizons is suggestive.

CI and FI Interaction (H4)

Our analysis of CI*FI interactions across different domains revealed a mix pattern of results. For 7-day weather predictions, we found a significant negative interaction in Block 2 (*Exp(B)* = .960, p = .011) that remained marginally significant in our final model (Block 3: *Exp(B)* = .961, p = .063). This pattern, coupled with the previously discussed detrimental effect of FI alone, suggests that cognitive sophistication – whether through FI alone, or in combination with CI – might impair weather predictions at the 7-day horizon. While knowledge of some historical weather patterns (CI) might be beneficial in isolation, its combination with high analytical capabilities (FI) appears to compound the overanalyses for what is essentially a chaotic system (Lorenz, 1963).

We also observed some suggestive patterns in our intermediate analyses (Block 2) regarding CI*FI interactions in financial and entertainment domains (DJI 7-day: B = -14.627, p = .059; DJI 14-day: B = -16.110, p = .060; Movie: *Exp(B)* = 1.021, p = .090). However, these effects did not persist in our final models, suggesting the need for cautious interpretation. This synergistic effect makes theoretical sense for both financial markets and social outcomes, where success likely requires accumulated knowledge over time (CI) and the ability to process information and recognize emerging patterns (FI). For example, during DJI predictions, accurate forecasting might benefit from combining fundamental

market knowledge (subconsciously) accumulated over time, insights provided by the prompt, as well as the ability for quick analysis and deriving judgments. While it is tempting to theorize that CI and FI might work synergistically in these domains, where success would benefit from both accumulated knowledge and analytical ability, more research would be needed to establish such relationships conclusively.

These preliminary findings raise interesting questions about how different types of intelligence might function across prediction domains. While we found a significant negative interaction between CI and FI for weather predictions at the 7-day horizon, the relationship between cognitive capabilities in other domains remain unclear. Further research is needed to better understand when and why combinations of CI and FI might help or hinder prediction accuracy.

Three-Way Interaction: TIF*CI*FI (H5)

The absence of significant or marginally significant three-way interactions (TIF*CI*FI) across all domains (weather 2, 4, 7 and 14 days; movie box office, and DJI predictions) presents an important theoretical insight. While we found some suggestive evidence for two-way interactions between TIF and CI, as well as between CI and FI in certain domains, the consistent lack of three-way interactions suggests that TIF might not meaningfully alter how CI and FI work together to affect prediction accuracy. This pattern does not support our initial hypothesis that TIF might serve as a mechanism to optimize the combined expression of CI and FI during prediction tasks.

This result warrants consideration alongside our earlier findings regarding two-way interactions. While our results hint at the possibility that TIF might enhance access to CI in some contexts (as suggested by the marginally significant TIF*CI interactions), and that CI and FI might show domain-specific patterns, these processes appear to operate independently rather than synergistically. The consistent absence of three-way interactions might suggest that the emotional oracle effect, if it exists, could operate through simpler mechanisms than we initially theorized (i.e., through relatively straightforward enhancement of access to stored knowledge, rather than through more complex modulation of how different types of intelligence interact). If true, the simpler mechanism would align with the dual-process theories of decision making, which distinguish fast, automatic,

affective processes from slower, deliberative processes and propose that their relationship may be parallel, rather than in intricate combination (Evans, 2008). However, given the marginal nature of many of our two-way interactions, further research is needed to establish these relationships more conclusively.

Additional Findings: Direct Effects of CI and FI

While not part of our formal hypotheses, we observed some noteworthy effects for CI and FI. Most robustly, FI showed a significant negative effect in 7-day weather predictions that remained stable through our final model (Block 3: *Exp(B)* = .831, $p = .014$). This finding presents an interesting contrast to Mellers, Stone, Atanasov, et al. (2015), which found FI to be positively associated with forecasting accuracy. In our weather predictions, higher levels of FI appear to impair prediction accuracy – at least at the 7-day range. This seems highly plausible, given that higher abstract reasoning abilities might lead to over-analysis of inherently complex, and sometimes random, meteorological patterns (e.g., Lorenz, 1963, Palmer, 2017). The finding aligns with research on decision-making under uncertainty (Artinger et al., 2015; Gigerenzer & Gaissmaier, 2011), which suggests that simpler, more intuitive approaches often outperform complex analytical strategies when dealing with highly dynamic systems.

CI, meanwhile, showed mixed and inconsistent effects – a marginally significant negative effect for 2-day weather predictions in Block 2 (*Exp(B)* = 0.887, $p = .093$), and a marginally significant positive effect for 7-day DJI predictions in Block 1 ($B = -68.582$, $p = .051$). However, neither of these CI effects persisted in our final models. These divergent findings from Mellers, Stone, Atanasov, et al. (2015) and Mellers, Stone, Murray, et al. (2015) might be explained by fundamental differences in prediction formats: while their research examined prediction tournaments where forecasters could update their predictions multiple times over extended periods, our study focused on single-point predictions. These patterns suggest that cognitive capabilities might function differently not only across prediction domains, but also across prediction formats, with FI potentially being detrimental in one-time predictions of chaotic systems, while CI's role remains less clear.

Conclusion

Our research had two key objectives: to replicate and extend Pham et al.'s (2012) emotional oracle effect, and to examine how emotional and cognitive processes might work together in prediction tasks. However, our results present a significant challenge to the original findings. Across multiple studies, including a high-powered confirmation study (N = 432), we were unable to replicate the direct effects of trust-in-feelings (TIF) on prediction accuracy. The failure to replicate is particularly noteworthy given our methodological approach – we employed two studies with larger sample sizes, included manipulation checks, and conducted a targeted confirmation study focused on the prediction that showed a strong effect in Pham et al.'s (2012) study. The fact that the emotional oracle effect did not emerge, even under these favourable conditions, raises fundamental questions about the robustness of TIF as a predictive mechanism and its purported ability to work across diverse prediction domains as suggested by Pham et al. (2012).

What emerged instead was a more complex but tentative picture, with two key findings. First, we found consistent marginally significant interactions suggesting TIF might enhance prediction accuracy when support by sufficient CI, at least across two of the three prediction domains. This pattern suggests that if TIF does provide some "privileged window" into tacit knowledge, as proposed by Pham et al. (2012), this window may only be valuable when there is substantial domain knowledge to access. However, we must be cautious in drawing even these conclusions, as they represent circumstantial evidence built on our inability to replicate the foundational effect. Second, in examining cognitive capabilities, we found mixed patterns. Most robustly, we found a significant negative effect of FI on weather predictions – our only significant result in Block 3, and only for the 7-day prediction domain – suggesting that higher levels of FI might impair accuracy when predicting events of natural phenomena. CI showed some marginally significant effects that varied by domain, but these effects did not persist in our final models. The attenuation of CI's direct effects in later blocks likely reflects shared variance with interaction terms, suggesting that CI's influence on prediction accuracy may be conditional on other factors rather than operating independently. These patterns, particularly when contrasted with Mellers, Stone, Atanasov, et al.'s (2015) findings from prediction tournaments, suggest that

cognitive capabilities might function differently in single-point predictions versus extended forecasting tasks.

Ultimately, our findings underscore the critical importance of replicating foundational claims as a prerequisite for extending theoretical frameworks (Open Science Collaboration, 2015); only with rigorously validated theories can we hope to build toward a deeper and more nuanced understanding of how cognition and emotion interact in forecasting.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.

Albert, D. A., & Smilek, D. (2023). Comparing attentional disengagement between Prolific and MTurk samples. *Scientific Reports, 13*(1), 20574.

Amer, M., Daim, T. U., & Jetter, A. (2013). A review of scenario planning. *Futures, 46*, 23–40.

Artinger, F., Petersen, M., Gigerenzer, G., & Weibler, J. (2015). Heuristics as adaptive decision strategies in management. *Journal of Organizational Behavior, 36*(S1), S33-S52.

Avnet, T., & Pham, M. T. (2005). Should I trust my feelings or not? The metacognition of affect as information in judgment. *Advances in Consumer Research, 32*, 38-39.

Avnet, T., Pham, M. T., & Stephen, A. T. (2012). Consumers' trust in feelings as information. *Journal of Consumer Research, 39*(4), 720-735.

Bargh, J. A., & Chartrand, T. L. (1999). The unbearable automaticity of being. *American Psychologist, 54*(7), 462–479.

Bechara, A., & Damasio, A. R. (2005). The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior, 52*(2), 336-372.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science, 275*(5304), 1293-1295.

Beier, M. E., & Ackerman, P. L. (2005). Age, ability, and the role of prior knowledge on the acquisition of new domain knowledge: Promising results in a real-world learning environment. *Psychology and Aging, 20*(2), 341-355.

Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology, 100*(3), 407–425.

Clore, G. L., & Parrott, W. G. (1994). Cognitive feelings and metacognitive judgments. *European Journal of Social Psychology, 24*(1), 101–115.

Clore, G. L., Schwarz, N., & Conway, M. (1994). Affective causes and consequences of social information processing. In R. S. Wyer, Jr. & T. K. Srull (Eds.), *Handbook of social cognition: Basic processes; Applications* (2nd ed., pp. 323-417). Lawrence Erlbaum Associates.

Craney, T. A., & Surles, J. G. (2002). Model-dependent variance inflation factor cutoff values. *Quality Engineering, 14*(3), 391–403.

Cunningham, W. R., Clayton, V., & Overton, W. (1975). Fluid and crystallized intelligence in young adulthood and old age. *Journal of Gerontology, 30*(1), 53–55.

Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. *Journal of Business and Psychology, 29*(1), 1–19.

de Sousa, R. (1979). The rationality of emotions. *Dialogue: Canadian Philosophical Review, 18*(1), 41–63.

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLOS ONE, 18*(3), e0279720.

Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology, 59*, 255-278.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Gaissmaier, W., & Marewski, J. N. (2011). Forecasting elections with mere recognition from small, lousy samples: A comparison of collective recognition, wisdom of crowds, and representative polls. *Judgment and Decision Making, 6*(1), 73-88.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology, 62*(1), 451-482.

Goktan, A. B., & Miles, G. (2011). Innovation speed and radicalness: Are they inversely related? *Management Decision, 49*(4), 533–547.

Greifeneder, R., Bless, H., & Pham, M. T. (2011). When do people rely on affective and cognitive feelings in judgment? A review. *Personality and Social Psychology Review, 15*(2), 107-141.

Hair, J. F., Risher, J. J., Sarstedt, M., & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review, 31*(1), 2–24.

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). Guilford Press.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*(5), 253–270.

Hosmer, D. W., Jr., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). John Wiley & Sons.

Hunt, E. (2000). Let's hear it for crystallized intelligence. *Learning and Individual Differences, 12*(1), 123–129.

Husin, W. N. I. W., Santos, A., Ramos, H. M., & Nordin, M. S. (2013). The place of emotional intelligence in the 'intelligence' taxonomy: Crystallized intelligence or fluid intelligence factor? *Procedia - Social and Behavioral Sciences, 97*, 214–223.

International Monetary Fund. (2023). *World economic outlook: A rocky recovery* (April 2023). https://www.imf.org/en/publications/weo/issues/2023/04/11/world-economic-outlook-april-2023

Ismail, M. (2005). Creative climate and learning organization factors: Their contribution towards innovation. *Leadership & Organization Development Journal, 26*(8), 639–654.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, 105*(19), 6829–6833.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist, 64*(6), 515–526.

Kaya, F., Delen, E., & Bulut, O. (2012). Test review: Shipley-2 manual. *Journal of Psychoeducational Assessment, 30*(6), 593–597.

Linstone, H. A., & Turoff, M. (Eds.). (1975). *The Delphi method: Techniques and applications*. Addison-Wesley.

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin, 127*(2), 267-286.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences, 20*(2), 130–141.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1–10.

Mellers, B., Baker, J. D., Chen, E., Mandel, D. R., & Tetlock, P. E. (2017). How generalizable is good judgment? A multi-task, multi-benchmark study. *Judgment and Decision Making, 12*(4), 369–381.

Mellers, B., Stone, E., Atanasov, P., Rohrbaugh, N., Metz, S. E., Ungar, L., Bishop, M. M., Horowitz, M., Merkle, E., & Tetlock, P. (2015). The psychology of intelligence analysis: Drivers of prediction accuracy in world politics. *Journal of Experimental Psychology: Applied, 21*(1), 1–14.

Mellers, B., Stone, E., Murray, T., Minster, A., Rohrbaugh, N., Bishop, M., Chen, E., Baker, J., Hou, Y., Horowitz, M., Ungar, L., & Tetlock, P. (2015). Identifying and cultivating superforecasters as a method of improving probabilistic predictions. *Perspectives on Psychological Science, 10*(3), 267–281.

Naqvi, N., Shiv, B., & Bechara, A. (2006). The role of emotion in decision making: A cognitive neuroscience perspective. *Current Directions in Psychological Science, 15*(5), 260–264.

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., ... & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology, 73*, 719-748.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716.

Palmer, T. N. (2017). The primacy of doubt: Evolution of numerical weather prediction from determinism to probability. *Journal of Advances in Modeling Earth Systems, 9*(2), 730-734.

Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research, 96*(1), 3–14.

Pham, M. T., Lee, L., & Stephen, A. T. (2012). Feeling the future: The emotional oracle effect. *Journal of Consumer Research, 39*(3), 461-477.

Polanyi, M. (1966). The logic of tacit inference. *Philosophy, 41*(155), 1–18.

Raphals, L. (2013). *Divination and prediction in early China and ancient Greece*. Cambridge University Press.

Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & M. P. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73–163). Guilford Press.

Seo, M.-G., & Barrett, L. F. (2007). Being emotional during decision making—Good or bad? An empirical investigation. *Academy of Management Journal, 50*(4), 923-940.

Shipley, W. C., Gruber, C. P., Martin, T. A., & Klein, A. M. (2009). *Shipley-2: Manual*. Western Psychological Services.

Stamenković, D., & Holyoak, K. J. (2018). Individual differences in both fluid and crystalized intelligence predict metaphor comprehension. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society* (pp. 2561–2566). Cognitive Science Society.

Sternberg, R. J., Wagner, R. K., Williams, W. M., & Horvath, J. A. (1995). Testing common sense. *American Psychologist, 50*(11), 912–927.

Tetlock, P. E., & Gardner, D. (2015). *Superforecasting: The art and science of prediction*. Crown Publishers/Random House.

Tetlock, P. E., Mellers, B. A., Rohrbaugh, N., & Chen, E. (2014). Forecasting tournaments: Tools for increasing transparency and improving the quality of debate. *Current Directions in Psychological Science, 23*(4), 290-295.

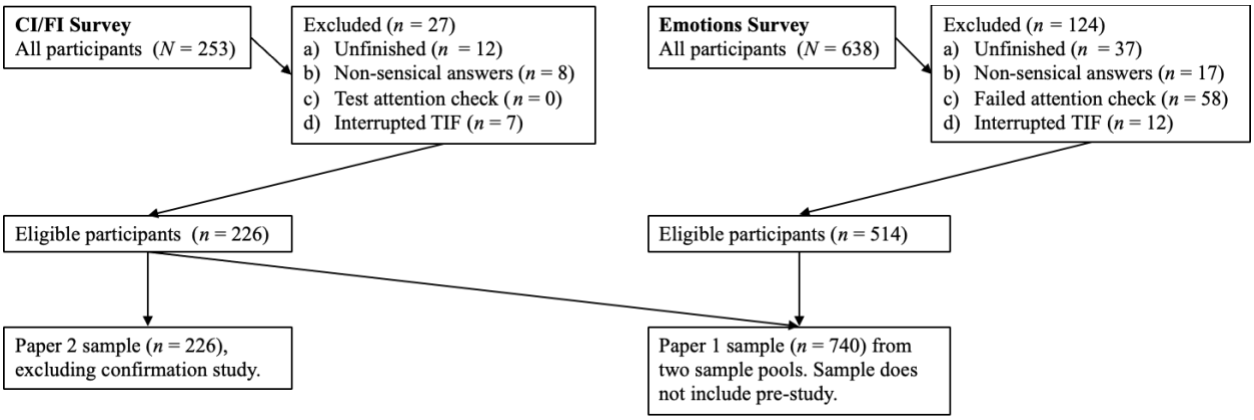
Ziegler, M., Danay, E., Heene, M., Asendorpf, J., & Bühner, M. (2012). Openness, fluid intelligence, and crystallized intelligence: Toward an integrative model. *Journal of Research in Personality, 46*(2), 173-18

Appendix

Appendix A

Sample Composition

(*Paper 2 is current paper*)



Appendix B

Screenshots of Administered Participant Surveys

Note. The section numbers may not be sequential as they are part of a larger survey.

Section 1

Thank you for participating in our research!

Please **DO NOT use a mobile phone** to participate in this study. You **CANNOT** use any external sources (e.g., Google) during the study. Please make sure you are answering this survey in a quiet environment. You must have access to sound, either through headphones or loud speakers.

General Information
The study is conducted only for scientific purposes, and takes approximately 25 minutes to complete. To receive your compensation, you must answer all questions provided. Kindly note that there are no right or wrong answers - all that we ask is for you to give the study your full attention. During the study, you may view films that may be distressing for you.

All collected data will be confidential. Participation in this study is anonymous. Results are published only in an aggregated format, and will not contain identifiable information to any specific participant. Please note that participation in this study is entirely voluntary, and you may discontinue at any time.

Study Information
The goal of the study is test your mental flexibility and memory. You will be asked to complete a range of different tasks. Should you have any questions, please do not hesitate to contact me through Prolific.

Do you agree to proceed with the study?

I agree

I do not agree

Section 2

What is your Prolific ID?

Please note that this response should auto-fill with the correct ID

Section 10

For the second task, please read the instructions below carefully.

There are two basic ways of making decisions. Sometimes people make decisions based on facts and logical reasons. Sometimes, they make decisions based on their feelings, that is, “does it feel good or bad?”, “does it feel right or wrong?”.

For this task, we are interested in how much confidence individuals have in their feelings when making decisions.

Instructions:

We would like you to briefly describe different situations in which you trusted your feelings to make a judgment or a decision and it was the right thing to do.

In other words, describe specific situations that really happened to you in which you felt confident in making a judgment or a decision based on how you felt, and later found out that you were right in using your feelings for this judgment or decision. The situations you list could involve any type of decision, provided that they were real situations in which you felt pretty certain that you should go by what your feelings were telling you to do, and it turned out that your feelings were right. For example, one participant in a previous session of this study mentioned the following:

“I had a very good feeling toward my new roommate from the moment I met him. Although objectively we were quite different, my feelings told me that we would be good friends. My feelings were correct! We became best friends after only a week. Trusting how I felt about my new roommate was definitely the right thing to do."

In the next **seven minutes**, please list **TWO (2)** similar situations that happened to you in which you felt quite confident in your feelings, and you were right in trusting them to make your decision (three or four sentences per situation will be enough).

Please list below TWO (2) situations in which you were right in trusting your feelings...

Section 11

You will be automatically advanced to the next page after seven (7) minutes.

1st situation in which I was right in trusting my feelings:

2nd situation in which I was right in trusting my feelings:

Section 12

We would like to assess people's ***perceptions*** of various countries that are not particularly well-known, but have an interest in expanding their tourism sectors. In this questionnaire, we will ask you about your general perceptions (mental images) about three relatively less-known countries. There are no right or wrong answers. We are only interested in your own perceptions of these countries. This information will be used to assess how these countries could better market themselves. Please answer on the scales provided below.

Country 1: Mauritania

What is your perception of the level of education in this country?

Far below average	Below average	Average	Above average	Far above average

What is your perception of the population density in this country?

Very dense	Somewhat dense	Neither dense nor sparse	Somewhat sparse	Very sparse

What is your perception of the cost of living in this country?

Very low	Somewhat low	Neither high nor low	Somewhat high	Very high

What is your perception of what the terrain is like in this country?

Mostly mountainous	Somewhat mountainous	Neither mountainous nor flat	Somewhat flat	Mostly flat

Section 13

Country 2: Romania

What is your perception of the level of education in this country?

Far below average	Below average	Average	Above average	Far above average

What is your perception of the population density in this country?

Very dense	Somewhat dense	Neither dense nor sparse	Somewhat sparse	Very sparse

What is your perception of the cost of living in this country?

Very low	Somewhat low	Neither high nor low	Somewhat high	Very high

What is your perception of what the terrain is like in this country?

Mostly mountainous	Somewhat mountainous	Neither mountainous nor flat	Somewhat flat	Mostly flat

Section 14

Country 3: New Zealand

What is your perception of the level of education in this country?

Far below average	Below average	Average	Above average	Far above average

What is your perception of the population density in this country?

Very dense	Somewhat dense	Neither dense nor sparse	Somewhat sparse	Very sparse

What is your perception of the cost of living in this country?

Very low	Somewhat low	Neither high nor low	Somewhat high	Very high

What is your perception of what the terrain is like in this country?

Mostly mountainous	Somewhat mountainous	Neither mountainous nor flat	Somewhat flat	Mostly flat

Section 15

For the fourth task, we would like to understand your perspectives about potential events in the future.

Section 16

What zip code are you currently residing in? (e.g., 10010)

What do you expect the **predominant** weather to be from **12pm to 6pm** at your zip code in **2 days** (Sunday, April 16)?

Sunny/fine/clear	Partially sunny with some clouds	Cloudy/overcast	Haze/fog	Thunderstorm	Rain	Snow

Section 17

What do you expect the **predominant** weather to be from **12pm to 6pm** at your zip code in **4 days** (Tuesday, April 18)?

Sunny/fine/clear	Partially sunny with some clouds	Cloudy/overcast	Haze/fog	Thunderstorm	Rain	Snow

Section 18

What do you expect the **predominant** weather to be from **12pm to 6pm** at your zip code in **7 days** (Friday, April 21)?

Sunny/fine/clear	Partially sunny with some clouds	Cloudy/overcast	Haze/fog	Thunderstorm	Rain	Snow

Section 19

What do you expect the **predominant** weather to be from **12pm to 6pm** at your zip code in **14 days** (Friday, April 28)?

Sunny/fine/clear	Partially sunny with some clouds	Cloudy/overcast	Haze/fog	Thunderstorm	Rain	Snow

Section 20

In the movie industry, the "opening weekend box office sales" is the amount of money generated during the first weekend (Friday to Sunday) when a movie is released in theatres.

On the next page, you will be provided basic information about three upcoming movies this weekend. Based on the information, please rank, in order of **highest to lowest**, what you expect the ranking of the weekend box office sales to be domestically (i.e., only within United States).

Section 21

Name: Mafia Mamma
MPAA: R
Main Cast: Toni Collette, Monica Bellucci, Sophia Nomvete, Alessandro Bressanello
Genre: Action Comedy Crime
Brief Synopsis: An American mom inherits her grandfather's mafia empire in Italy. Guided by the firm's consigliere, she hilariously defies everyone's expectations as the new head of the family business.



Name: Nefarius
MPAA: R
Main Cast: Sean Patrick Flanery, Jordan Belfi, James Healy Jr., Eric Hanson
Genre: Horror Thriller
Brief Synopsis: On the day of his scheduled execution, a convicted serial killer gets a psychiatric evaluation during which he claims he is a demon, and further claims that before their time is over, the psychiatrist will commit three murders of his own.



Name: Sweetwater
MPAA: PG-13
Main Cast: Cary Elwes, Eric Roberts, Richard Dreyfuss, Jeremy Piven
Genre: Biography Drama Sport
Brief Synopsis: Hall of Famer Nat "Sweetwater" Clifton makes history as the first African American to sign an NBA contract, forever changing how the game of basketball is played.



Section 22

Name: Nefarius
MPAA: R
Main Cast: Sean Patrick Flanery, Jordan Belfi, James Healy Jr., Eric Hanson
Genre: Horror Thriller
Brief Synopsis: On the day of his scheduled execution, a convicted serial killer gets a psychiatric evaluation during which he claims he is a demon, and further claims that before their time is over, the psychiatrist will commit three murders of his own.



Name: Mafia Mamma
MPAA: R
Main Cast: Toni Collette, Monica Bellucci, Sophia Nomvete, Alessandro Bressanello
Genre: Action Comedy Crime
Brief Synopsis: An American mom inherits her grandfather's mafia empire in Italy. Guided by the firm's consigliere, she hilariously defies everyone's expectations as the new head of the family business.



Name: Sweetwater
MPAA: PG-13
Main Cast: Cary Elwes, Eric Roberts, Richard Dreyfuss, Jeremy Piven
Genre: Biography Drama Sport
Brief Synopsis: Hall of Famer Nat "Sweetwater" Clifton makes history as the first African American to sign an NBA contract, forever changing how the game of basketball is played.



For the three movies, what do you expect the ranking of the opening weekend box office sales to be domestically?

To rank the movies, please drag the name of the movie to the corresponding box on the right.
(**Rank #1** = **highest** sales out of the three, **Rank #2** = **second highest** sales out of the three, and **Rank #3** = **lowest** sales out of the three).

Items
Nefarius
Sweetwater
Mafia Mamma

Rank #1

Rank #2

Rank #3

Section 23

The “Dow Jones” Index (officially known as the Dow Jones Industrial Average Index), is one of the most famous and most widely followed stock market indices in the world. This index is a weighted-average of the stock prices of thirty of the largest publicly traded companies in the United States, such as Boeing, Disney, Microsoft and Goldman Sachs. The stock prices of these companies reflect the value that investors collectively perceive in these companies: their profitability, growth potential, etc. The level of the Dow Jones is thus widely regarded as a summary indicator of the performance of the stock market, investors' expectations of future performance of the stock market, and the strength and growth potential of the US economy.

Here are some indicative values of the Dow Jones Index (DJI) over the past years:

- On February 27, 1991, the day the first Gulf War ended and the economy was in a recession, the DJI closed at 2,889
- On January 14, 2000, at the height of the “Dot-Com Bubble”, the DJI closed at 11,723
- On March 13, 2000, just after the “Dot-Com Bubble” burst, the DJI closed at 9,947
- On May 6, 2010, after a computer-driven trading glitch, the DJI closed at 10,520
- On Jan 15, 2020, after a multi-year bull run and lessening trade disputes between China and US, the DJI closed at 29,030
- On March 12, 2020, after the US President declared a public health emergency (COVID), the DJI closed at 21,200
- On Nov 24, 2020, the awaited transition to the Biden administration began, the DJI closed at 30,046
- On Sept 13, 2022, after an inflation report of 8.3% year-over-year, DJI dropped 1,276 points and closed at 31,105
- On April 12, 2023 - this Wednesday, two days ago, the DJI closed at 33,647

Where do you predict the Dow Jones index, at closing, will be on the following days in the future? Please enter numbers only (i.e, without ",").

One week from today **(Friday, April 21)**

Two weeks from today **(Friday, April 28)**

Section 24

For the next task, please complete the following questions to the best of your ability. No external aids are allowed.

Section 25

Select the word that has the same meaning as the one written in capital letters.

Example: LARGE ... a) red b) big c) silent d) wet
Answer: b) big

TALK

- draw
- eat
- speak
- sleep

COUCH

- pin
- eraser
- sofa
- glass

REMEMBER

- swim
- recall
- number
- plan

PARDON

- forgive
- pound
- divide
- crash

HIDEOUS

- silvery
- tilted
- young
- dreadful

MASSIVE

- bright
- large
- speedy
- low

PROBABLE

- likely
- portable
- friendly
- comprehensive

IMPOSTER

- conductor
- officer
- book
- pretender

FASCINATE

- welcome
- fix
- stir
- enchant

HILARITY

- laughter
- speed
- grace
- malice

EVIDENT

- green
- obvious
- skeptical
- afraid

IGNORANT

- red
- sharp
- uninformed
- precise

NARRATE

- yield
- buy
- associate
- tell

CAPTION

- drum
- ballast
- heading
- ape

HAUL

- respond
- twist
- pull
- realize

INDICATE

- defy
- excite
- signify
- bicker

SOLEMN

- serious
- satisfying
- rough
- tremendous

FACILITATE

- turn
- help
- strip
- bewilder

FORTIFY

- submerge
- strengthen
- vent
- deaden

AMULET

- charm
- orphan
- dingo
- pond

MERIT

- deserve
- distrust
- fight
- separate

STERILE

- barren
- illegal
- helpless
- tart

RENOWN

- length
- head
- fame
- loyalty

CORDIAL

- swift
- muddy
- leafy
- affable

SQUANDER

- tease
- belittle
- slice
- waste

PRISTINE

- vain
- sound
- unspoiled
- level

SERRATED

- dried
- notched
- armed
- blunt

INNOCUOUS

- powerful
- pure
- medicinal
- harmless

PLAGIARIZE

- maintain
- intend
- revoke
- pilfer

JOCOSE

- humorous
- paltry
- fervid
- plain

ORIFICE

- brush
- hole
- building
- lute

LOGLINE

- coverage
- nefarious
- summary
- slides

RUE

- deal
- lament
- dominate
- cure

MOLLIFY

- mitigate
- direct
- pertain
- abuse

INEXORABLE

- untidy
- inviolable
- relentless
- sparse

DIEGESIS

- equivalent
- methodological
- disruption
- narrative

DIVEST

- dispossess
- intrude
- rally
- pledge

QUERULOUS

- maniacal
- curious
- devout
- complaining

MONTAGE

- solidarity
- oblique
- melange
- origin

ABET

- waken
- ensue
- incite
- placate

DESUETUDE

- disuse
- remonstrance
- corruption
- inanity

PEREGRINATE

- contemplate
- mince
- solidify
- traverse

QUOTIDIAN

- travesty
- everyday
- calculation
- promise

Section 26

Please complete the following questions to the best of your ability. You may use a writing utensil and blank pieces of paper. No other aids (e.g., mobile phone, internet, calculator, etc.) are allowed. This section will take approximately 20-30 minutes.

Section 27

Fill in the missing letter, number, or word to complete each sequence. Input only one character in each blank space (_). Please turn on caps lock for the duration of this section.

Example: big little high low cold _ _ _ (answer: HOT)

1, 2, 3, 4, 5, _

white black short long down _ _

AB BC CD D_

10 1 9 2 8 _

A Z B Y C X D _

oh ho rat tar mood _ _ _ _

escape scape cape _ _ _

mist is wasp as pint in tone _ _

NE/SW SE/NW E/W N/_

12321 23432 34543 456_ _

knit in spud up both to stay _ _

57326 73265 32657 26573 _ _ _ _ _

Scotland landscape scapegoat _ _ _ _ee

3 7 5 11 2 _

tam tan rib rid rat raw hip _ _ _

G V J T M R _ _

surgeon 1234567 snore 17635 rogue _ _ _ _ _

4 11 7 20 3 _

two w four r one o three _

thicken 10 founder 4 nectarine _

3124 82 73 154 46 13_

K W M S P P T N _ _

pole post mail carton box fight film picture depict money _ _ _ _ _ _ alter

9 6 12 7 3 _

trauma tuna flight fit wife _a_ _ glossy gravity

Section 28

How certain are you about each of your predictions?

	Very unsure	Somewhat unsure	Neither sure nor unsure	Somewhat certain	Almost certain
Prediction 1a: Weather in 2 days	○	○	○	○	○
Prediction 1b: Weather in 4 days	○	○	○	○	○
Prediction 1c: Weather in 7 days	○	○	○	○	○
Prediction 1d: Weather in 14 days	○	○	○	○	○
Prediction 2: Weekend movie box office	○	○	○	○	○
Prediction 3a: Dow Jones Index in 7 days	○	○	○	○	○
Prediction 3b: Dow Jones Index in 14 days	○	○	○	○	○

Section 29

To what extent do you agree with the following statements?

	Strongly disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I found, on average, the prediction tasks difficult	○	○	○	○	○
I took the prediction tasks seriously	○	○	○	○	○
I was distracted during the prediction tasks	○	○	○	○	○
I was excited to make predictions	○	○	○	○	○
I put effort into the prediction tasks	○	○	○	○	○

Section 30

Have you checked what the weather is for this Sunday (2 days from now)?

Yes

No

Section 34

Having completed all tasks, what would you say is the overall goal of this study?

Section 35

What gender do you identify as?

- Male
- Female
- Non-binary / third gender
- Prefer not to say

What is your age?

- 18-25
- 26-35
- 36-45
- 46-55
- 56+

What is the highest degree or level of education you have completed?

- High school
- Bachelor's degree
- Master's degree
- Ph.D. or higher
- Trade school
- Prefer not to say

Appendix C.1

High Trust-in-feelings (TIF) Manipulation Instructions

For the first task, please read the instructions below carefully.

There are two basic ways of making decisions. Sometimes people make decisions based on facts and logical reasons. Sometimes, they make decisions based on their feelings, that is, “does it feel good or bad?”, “does it feel right or wrong?”.

For this task, we are interested in how much confidence individuals have in their feelings when making decisions.

Instructions:

We would like you to briefly describe different situations in which you trusted your feelings to make a judgment or a decision and it was the right thing to do.

In other words, describe specific situations that really happened to you in which you felt confident in making a judgment or a decision based on how you felt, and later found out that you were right in using your feelings for this judgment or decision. The situations you list could involve any type of decision, provided that they were real situations in which you felt pretty certain that you should go by what your feelings were telling you to do, and it turned out that your feelings were right. For example, one participant in a previous session of this study mentioned the following:

“I had a very good feeling toward my new roommate from the moment I met him. Although objectively we were quite different, my feelings told me that we would be good friends. My feelings were correct! We became best friends after only a week. Trusting how I felt about my new roommate was definitely the right thing to do."

In the next **seven minutes**, please list **TWO (2)** similar situations that happened to you in which you felt quite confident in your feelings, and you were right in trusting them to make your decision (three or four sentences per situation will be enough).

Please list below TWO (2) situations in which you were right in trusting your feelings...

Appendix C.2

No Trust-in-feelings (TIF) instructions

For the first task, please read the instructions below carefully.

The purpose of this task is to examine how online search technologies are used to facilitate people's lives. We would like you to briefly describe different situations in which you had to do a search on "Google" to find information that you needed and managed to find the information you were looking for.

In other words, describe specific situations that really happened to you in which you needed information about something and were able to use Google to search for this information. The situations you list could involve any type of situation, provided that they were real situations in which you needed some information and did a Google search to find it.

For example, one participant in a previous session of this study mentioned the following:

"I needed to find a good Thai restaurant in my neighborhood because I had just moved to the area. I used Google to search for Thai restaurants near my home and found a number of places, which I then checked out further."

In the next **seven minutes**, please list **FIVE (5)** similar situations that happened to you in which you needed information and you did a Google search to find this information (three or four sentences per situation will be enough).

Please list FIVE (5) situations in which you searched on Google…

Appendix D

Pre-Screening Questions for Participants

1. **Gross Domestic Product, or GDP, _____** (please select the most correct answer)
    a. Measures national income and output, including consumption, investment, and government spending.
    b. Measures national income and output, including consumption, and government spending.
    c. Measures national income and output, including consumption, investment, government spending, and net exports (i.e., exports minus imports).
    d. Measures national income and output, including consumption, government spending and inflation

2. **What is a P/E ratio?**
    a. The percentage of profits that a company reinvests in its operations
    b. The amount of debt a company has compared to its equity
    c. The percentage of the market capitalization represented by the company's net income
    d. The ratio of a company's stock price to its earnings per share (EPS)

3. **What is the typical relationship between changes in interest rates and stock market levels?** (please circle the most correct answer)
    a. Positive: if interest rates go up, stock prices also go up
    b. Negative: if interest rates go up, stock prices go down
    c. Same: Stock prices and interest rates move in the same direction, whether up or down
    d. There is no relationship

4. **Have you checked the weather in your area (zip code) for any time within the next two weeks?**

    Yes _ No _

Appendix E

Block 3 Results for Weather, Movie and DJI Predictions (Logistic and Linear Regression)

			95% C.I. for OR				
Predictor Variable	***B***	**OR (Exp(B))**	**Lower Bound**	**Upper Bound**	***p***	**Sig**	**Nagelkerke R^2**
2-Day Weather							0.061
TIF	0.096	1.101	0.568	2.135	0.775		
CI	-0.115	0.892	0.776	1.025	0.106		
FI	0.034	1.035	0.907	1.180	0.612		
TIF*FI	-0.009	0.991	0.823	1.192	0.922		
TIF*CI	0.190	1.209	0.973	1.502	0.087	†	
CI*FI	-0.009	0.991	0.959	1.024	0.596		
TIF*CI*FI	-0.038	0.963	0.908	1.020	0.202		
4-Day Weather							0.041
TIF	0.594	1.811	0.933	3.517	0.079	†	
CI	-0.090	0.914	0.786	1.063	0.243		
FI	-0.063	0.939	0.818	1.079	0.375		
TIF*FI	0.033	1.034	0.865	1.236	0.714		
TIF*CI	0.074	1.077	0.873	1.330	0.488		
CI*FI	-0.009	0.991	0.958	1.025	0.607		
TIF*CI*FI	0.004	1.004	0.958	1.051	0.879		
7-Day Weather							0.098
TIF	0.218	1.244	0.607	2.548	0.551		
CI	-0.104	0.901	0.766	1.060	0.210		
FI	-0.184	0.832	0.716	0.967	0.016	*	
TIF*FI	0.162	1.176	0.963	1.435	0.112		
TIF*CI	0.053	1.055	0.831	1.339	0.662		
CI*FI	-0.040	0.961	0.921	1.002	0.063	†	
TIF*CI*FI	-0.003	0.997	0.936	1.063	0.931		
14-Day Weather							0.012
TIF	0.117	1.124	0.533	2.372	0.759		
CI	-0.046	0.955	0.813	1.121	0.571		
FI	-0.027	0.973	0.841	1.126	0.715		
TIF*FI	0.033	1.034	0.845	1.265	0.747		
TIF*CI	0.033	1.033	0.812	1.316	0.791		
CI*FI	-0.003	0.997	0.962	1.033	0.855		
TIF*CI*FI	-0.017	0.983	0.928	1.041	0.552		
Movie Box Office							0.063
TIF	0.715	2.045	0.906	4.613	0.085	†	
CI	0.068	1.070	0.881	1.301	0.494		
FI	0.080	1.083	0.914	1.284	0.357		
TIF*FI	-0.051	0.950	0.767	1.176	0.637		
TIF*CI	-0.040	0.960	0.746	1.237	0.755		
CI*FI	0.027	1.027	0.993	1.063	0.126		
TIF*CI*FI	-0.011	0.989	0.943	1.036	0.633		

Note. Effects significant or marginally significant in earlier blocks: CI for 2-Day Weather (Block 2: p = .093†); CI*FI for 7-Day Weather (Block 2: p = .011*); CI*FI for Movie (Block 2: p = .090†). †$p < .10$, *$p < .05$, **$p < .01$.

Predictor Variable		B	β (Stand.)	95% C.I. for B		*p*	Sig	VIF
				Lower Bound	Upper Bound			
7-Day DJI								
	CI	-25.060	-0.051	-119.986	69.867	0.603		2.184
	FI	-47.572	-0.111	-132.028	36.883	0.268		2.310
	TIF	102.297	0.031	-348.015	552.608	0.655		1.108
	TIF*FI	64.648	0.108	-54.099	183.395	0.284		2.310
	TIF*CI	-132.811	-0.184	-275.721	10.099	0.068	†	2.333
	CI*FI	-9.877	-0.088	-30.438	10.684	0.345		1.991
	TIF*CI*FI	-10.490	-0.067	-41.044	20.065	0.499		2.229
14-Day DJI								
	CI	-8.952	-0.016	-113.833	95.930	0.867		2.184
	FI	-59.878	-0.127	-153.191	33.434	0.207		2.310
	TIF	162.426	0.045	-335.108	659.960	0.521		1.108
	TIF*FI	53.403	0.080	-77.796	184.603	0.423		2.310
	TIF*CI	-155.747	-0.196	-313.643	2.150	0.053		2.333
	CI*FI	-8.331	-0.067	-31.048	14.386	0.471		1.991
	TIF*CI*FI	-17.178	-0.099	-50.936	16.581	0.317		2.229

Note. Block 3 hierarchical linear regression results. R^2 (Adjusted): 7-Day = .022; 14-Day = .022. Effects significant or marginally significant in earlier blocks: CI for 7-Day DJI (Block 1: $p = .051^{\dagger}$); CI*FI for 7-Day DJI (Block 2: $p = .059^{\dagger}$); CI*FI for 14-Day DJI (Block 2: $p = .060^{\dagger}$). $^{\dagger}p < .10$, $*p < .05$, $**p < .01$.