Can Explicit Epistemic Goals Reduce Primacy Bias in Political Impression Formation?

Experimental Study

Bachelor's Thesis

submitted in partial fulfillment of the requirements

for the degree of Bachelor of Science in Psychology

Supervisor:

Dr. Johannes Ziegler

LUDWIG MAXIMILIAN UNIVERSITY OF MUNICH

Submitted by Konstantinos-Nestor Germanos

On July 15th, 2025

Matriculation Number: 12415361 E-mail: nestor.germanos@campus.lmu.de

Abstract

How do voters form impressions of political candidates in an era of information overload? This study examines whether explicitly induced epistemic goals, hypothesized to prompt individuals to value accuracy over appeal, can reduce primacy bias in impression formation. Building on cognitive-ecological models and sampling paradigms, we conducted an online experiment (N = 103) in which participants evaluated nine political candidates by sequentially sampling trait descriptions under either hedonic or epistemic goal framing. Four preregistered hypotheses tested: (1) whether smaller sample sizes yield more extreme judgments as a marker of primacy bias; (2) whether epistemic goals attenuate this effect; (3) whether trait diagnosticity moderates the extremity-sampling relationship; and (4) whether general political trust enhances epistemically motivated processing. Contrary to expectations, epistemic framing did not significantly reduce extremity or increase sampling depth. While diagnosticity and sample size interacted as predicted, this relationship was not moderated by motivation. Exploratory analyses further indicated that larger samples did not enhance, and sometimes even impaired judgment accuracy. These findings underscore the difficulty of altering sampling behavior through subtle motivational cues in cognitively demanding, politically relevant tasks. Primacy effects and biased impressions may thus be more robust in political contexts than assumed, with implications for democratic decision-making and future impression formation research.

Table of contents

"Uncertainty is the only certainty there is, and knowing how to live with insecurity is the only security."
– John Allen Paulos, *A Mathematician Plays the Stock Market* (2003)

Introduction

Knowledge – meaning internalized and contextualized information – is often the difference between life and death, as the opening scene of Stanley Kubrick's *2001: A Space Odyssey* (1968) effectively depicts. In it, two proto-human tribes contest a scarce waterhole. When the member of one tribe acknowledges that a bone can function as a weapon, the balance of power shifts, leading to the demise of the opposing group. But the acquisition of information can sometimes prove equally fatal to ignorance: spotting a bear in the woods, one is sensibly inclined to respond by calmly retreating rather than by inquiring into the animal's intentions.

Over millennia, human evolution has refined our heuristic mechanisms of identifying relevant information and optimizing not just for impression and judgement accuracy, but also time, memory and access constraints (Gigerenzer & Goldstein, 1996; Simon, 1978). As such, we rely on probabilistic cues, socially embedded schemata and simplified mental representations to address the practical necessity of limiting our exposure to information (Brunswik, 1956; Cottam et al., 2022; Fiedler, 2000). Heuristics are thus generally not regarded as cognitive or evolutionary failures, but rather as ecologically adaptive responses under utility constraints, taking into account all the considerations relevant to a particular person and context (Dhami et al., 2004, 2019; Gigerenzer & Goldstein, 1996; Poulsen & Sonntag, 2020; Simon, 1978).

Nevertheless, these same heuristics also render us vulnerable to systematic distortions. Initial impressions often outweigh careful reasoning (Caplin, 2011, 2017; Goodin & Spiekermann, 2018), with early cues disproportionately shaping downstream evaluations and generating stronger, more polarized impressions. Assuming no significant preconceptions, meaning a neutral prior, this can be primarily attributed to the statistically skewed nature of small samples causing the formation of unrepresentative impressions (Denrell, 2005; Prager & Fiedler, 2021; Juslin & Olsson, 1997). Individuals may also sample selectively to avoid cognitive dissonance (Ditto et al., 2009; Hannon, 2022; Taber & Lodge, 2006), leading to biased impressions even when the truth is equally accessible (Arigapudi et al., 2023), or stop their research when coherence is felt, and not when the truth is reached (Salant & Cherry, 2020).

Though many contexts have remained the same, our daily life has drastically changed in recent times, along with the judgements we are called to make. We live in an age of unparalleled information overload that degrades decision quality (Peng et al., 2021), while also reducing productivity, and increasing anxiety and burnout (Arnold et al., 2023). Through globalization and the topological interconnectedness of today's world, social decisions have become especially complex. One of the most consequential decisions we are confronted with is which leaders to elect into power. There are empirical parallels between systematic distortions observed in impression formation research and voting patterns in political science, hinting towards the idea that our heuristics are stretched to their limits by the overwhelmingly rich contemporary sampling environment (Biella & Hütter, 2024; Druckman et al., 2013; Iyengar et al., 2009; Lau & Redlawsk, 1997; Prager et al., 2018; Ziegler & Fiedler, 2024). Such parallels justify the question of whether our intuitions still serve us well, or whether we must rethink social choice models to ensure more accurate, epistemically justified collective decisions (Giavazzi 2023; Goodin & Spiekermann, 2018). This thesis examines how motivational states shape sampling and impression formation in political candidate evaluation.

Impression formation research deals with the transformation of trait samples into coherent judgments. Asch (1946) envisaged a configural process in which impressions emerge through holistic, interdependent interpretation, and central traits disproportionately shape how subsequent traits are perceived. In contrast, Anderson's (1965) Information Integration Theory and Carroll's (1982) three-stage model of valuation, integration, and response assumed additive and multiplicative mechanisms of interplay. Modern formal approaches, including latent-variable models (Winsberg & De Soete, 1993), QRE frameworks (McKelvey & Palfrey, 1998) and quantum cognitive models (Busemeyer et al., 2022), further underscore how trait-based impressions are probabilistic and dynamically constructed. Contemporary research broadly acknowledges that impression formation involves both configural and additive aspects, integrated under bounded cognitive conditions (Biella & Hütter, 2024; Ziegler & Fiedler, 2024).

Motivation fundamentally shapes how information is processed and how deeply individuals choose to sample traits. Under hedonic or directional goals, new or conflicting information often feels aversive, prompting premature truncation of the sampling process and amplifying the weight of early or salient cues (Caplin, 2011, 2017; Taber & Lodge, 2006). This mirrors findings from selective exposure and motivated reasoning research, where individuals disengage from counter-attitudinal or uncertainty-inducing evidence (Lau et al., 2008). Models of rational inattention and coherence-based stopping further demonstrate that subjective stopping thresholds shift systematically with motivational states, perceived sampling costs, and affective context (Anderson, 1981; Arigapudi et al., 2023; Mechtenberg & Tyran, 2019; Poulsen & Sonntag, 2020). Under time pressure or strong hedonic motives, sampling truncates more quickly, judgments become more extreme and confident, and potential inconsistencies in the evidence are more readily ignored (Chen & Krajbich, 2018). Under epistemic goals, individuals value accuracy, actively seek out additional information, and sustain a more balanced sampling process, reducing primacy effects and judgment bias (Biella & Hütter, 2024).

The impact of a sampled trait on impression formation depends not just on its content or position in the sequence, but also on its diagnosticity, the perceived informativeness of the trait for reducing uncertainty about a target's latent attributes (Prager & Fiedler, 2021; Skowronski & Carlston, 1987). Highly diagnostic traits are often extreme or negative, standing out against baseline expectations and bolstering confidence by narrowing possible interpretations (Reeder & Brewer, 1979; Unkelbach et al., 2008). Thus, high diagnosticity reduces perceived ambiguity, increases latent trait separation and judgment consistency (Regenwetter et al., 2007; Winsberg & De Soete, 1993), and narrows the cognitive state space (Busemeyer et al., 2022). Diagnosticity is often amplified when cues align with identity or affective orientation (Ditto et al., 2009; Hannon, 2022; Taber & Lodge, 2006; Jung & Mittal, 2020;). Sparse samples can produce exaggerated certainty and base-rate neglect (Carroll, 1982; Juslin & Olsson, 1997), while salient signals may be systematically overweighted as a result (Salant & Cherry, 2020).

After decades of research exploring motivated reasoning, impression formation, and bounded rationality, no framework has yet unified these strands into a coherent account of how sampling shapes our political impressions (Prager & Fiedler, 2021; Ziegler & Fiedler, 2024). While most models of political decision-making focus on voter choice, the more foundational question may lie upstream, in how impressions of political candidates are formed in the first place. Judgments of competence, morality, and trustworthiness guide our voting intentions, yet these assessments emerge from limited, often biased sampling paths that rely on heuristics shaped more by evolutionary pressures than democratic deliberation (Cottam et al., 2022). Our tendency to satisfice (Simon, 1978) and favor early information (Denrell, 2005) is often adaptive when speed is critical, but in contexts demanding accuracy, truncation driven by primacy bias can undermine judgment quality. Recent research findings suggest that allowing autonomous decisions over sample size highlights how this primacy bias, although functional under time pressure, can be disadvantageous when accuracy should take precedent (Biella & Hütter, 2024).

In political science, voters' preference for news from sources sharing their views (Iyengar et al., 2009) resembles the "more-is-more" effect (Ullrich et al., 2013) in impression formation, where "… participants accumulate more information related to targets of whom they acquired a positive impression." (Biella & Hütter, 2024). Similarly, polarization intensifies partisanship, reduces the impact of substantive information, and boosts confidence in weaker opinions (Druckman et al., 2013), reflecting how diagnostic early cues cut sampling short and small, homogeneous samples strengthen judgments (Ziegler & Fiedler, 2024). Aversion to incongruent information also appears in the enhanced diagnosticity of negative traits enabling earlier truncation and exaggerated estimations (Prager et al., 2018). Thus, it is not far-fetched to claim that sampling characteristics might be inadvertently linked to the finding that only "an average of a little over three-quarters of the voters in the past nine presidential elections in the United States have, according to at least one defensible criterion, voted correctly – that is, in accordance with what their fully informed preferences should be" (Lau & Redlawsk, 1997).

The epistemic legitimacy of democracy as a system of preference aggregation relies on the correctness of individual judgements (Condorcet, 1785; List & Goodin, 2001). While social polarization and political bias emerge from many different sources, a crucial predictor of misaligned impressions and inaccurate predispositions about persons and social groups seems to be the systematic sampling error inherent in our limited searches for the truth (Biella & Hütter, 2024; Ziegler & Fiedler, 2024). Recent research suggests that introducing an epistemic goal can mitigate the impact of strong first impressions and the associated biases, effectively diminishing their influence on final judgments (Biella & Hütter, 2024). At the same time, political literature stands behind the view that voters default to non-epistemic, hedonic-like goals in collective, political decision making contexts (Ditto et al., 2009; Taber & Lodge, 2006).

Thus, two main questions guide this work. Initially, whether the apparent parallels between mechanisms behind biased impressions in social psychology and political science truly hold, allowing us to conceptually transfer insights about sampling artifacts as at least partial, but significant drivers of political misperceptions. Finally, whether introducing epistemic goals can meaningfully reduce primacy effects and improve the accuracy of political impression formation. In line with Fiedler's (2017) call for research grounded in rigorous theorizing, this thesis integrates methodological insights from Ziegler and Fiedler (2024) and Biella and Hütter (2024). It transfers their findings and generalizes their paradigm to a political context by combining free sampling termination (Prager et al., 2018) with source-switching between strategic partners (Biella & Hütter, 2024), and investigates whether epistemic goal framing modulates sampling behavior and downstream judgments. Sampling depth, impression strength, and confidence are all examined in a paradigm simulating electoral decision making.

In an experimental study, four confirmatory hypotheses are tested: the first serves as a sanity check, replicating previous findings by evaluating whether the negative relationship between sample size and judgement strength, linked to premature truncation and primacy bias, also appears in a political context (H1). This aims to ground the study in established effects, ensuring mechanisms observed in prior research generalize to political impression formation and our specific operationalization of the paradigm. The second, most central hypothesis, posits that inducing epistemic goals significantly reduces primacy effects, increasing sampling depth and improving judgement accuracy compared to hedonic goals (H2). This tests whether motivational framing can counteract biases in sequential trait sampling. The third hypothesis highlights the role of trait diagnosticity in explaining the relationship between information sampled and impression strength. We expect highly diagnostic traits to negatively moderate this relationship across the experiment by producing confident impressions and reducing the perceived necessity for further sampling, particularly in regards to the hedonic condition (H3).

The fourth and final confirmatory hypothesis is dedicated to examining general political trust as a between-subjects predictor of average voting likelihood and judgment confidence. Building on prior findings linking political efficacy and trust to increased engagement and discernment in political contexts (OECD, 2021) we expect higher political trust to positively moderate sampling depth, attenuating primacy biases, more so in the epistemic condition (H4). Lastly, we will explore whether participants' judgment accuracy, measured as actuarial deviation from sampled and population trait valence, varies systematically with sampling depth and motivational condition, offering further insight into factors shaping political impressions.

Methods

Participants

A total of 120 participants were recruited via the online research platform Prolific (www.prolific.com). Inclusion criteria required participants to be fluent in German, and aged over 17 years. Monetary compensation approximated a rate in line with Prolific's fair-pay guidelines. The final sample included 103 participants after excluding 17 based on pre-registered criteria, including sampling less than 35% of available traits, non-differentiated behavior, as well as judgement latencies under 500ms, which was examined using a mixture of algorithmic data marking, manual data inspection and trait-level cleaning. Of these, 87 identified as female, and 33 as male. The mean age was 33.5 years (SD = 11.1, range = 19-75). An a priori power analysis determined that this sample size would be sufficient to detect a conservative effect (Cohen's $d = 0.2$) with $\alpha = 0.05$ and power > 0.80 in our design. All participants provided informed consent and were thoroughly debriefed at the end of the study.

Design & Procedure

The study used a mixed design with the between-subjects factor being goal condition (epistemic vs. hedonic) and within-subject sampling behavior being measured across nine political candidates. Participants were randomly assigned to one of the two conditions before sampling began. Trait order and candidate display were also randomized for each participant.

After a welcome screen and informed consent form (per Art. 13 EU-DSGVO), participants completed a short demographic questionnaire. They then received instructions specific to their condition: either to imagine voting personally and choose the candidate who appeals to them most for the hedonic condition, or to put themselves in the position of a political journalist aiming to estimate each candidate's popularity as accurately as possible for the epistemic condition. An attention check followed and had to be answered correctly before continuing.

The main task displayed a 3×3 grid of candidate portraits. Participants selected one candidate at a time and sampled up to eight traits per person by clicking a "next" button to reveal each trait. Sampling was limited to a global maximum of 45 traits. Once a candidate was exited, returning was not permitted. Portraits were used solely for orientation and explicitly framed as irrelevant to the evaluation. A trait subpool was randomly assigned to each candidate portrait.

After completing the sampling phase, participants rated each candidate on voting likelihood and judgment confidence using continuous slider scales labeled with four verbal anchors. Participants were asked to rate 14 items pertaining to their subjective political efficacy drawn from the Perceived Political Self-Efficacy (P-PSE) and Political Efficacy Kurzskala (PEKS) questionnaires. Finally, participants were offered the chance to leave a comment about their experience, and were presented with a debriefing explaining primacy bias and motivational framing. The study was programmed in JavaScript and hosted on a secure university server.

Operationalization

The primary independent variable was goal condition, randomly manipulated between subjects. It was operationalized through the condition-specific instructions presented to the participants before sampling, designed to guide participants to either sample in their default, hedonic mode of cognition, or to put themselves in an explicitly epistemic sampling state of mind respectively. The wording of the instructions was inspired by the motivational framing used in Biella and Hütter (2024) to distinguish between intuitive and accuracy-oriented information processing.

The main dependent variable was judgment strength, operationalized as participants' voting likelihood for each candidate on a continuous slider scale. Judgment confidence, measured analogously, and sampling depth, the number of traits drawn per candidate, served as secondary dependent variables. All three variables were adapted from Ziegler & Fiedler (2024) and reformulated to fit the political evaluation context. The square root of sample size ($\sqrt{n}$) was used in modeling, as prior work has shown it better captures the non-linear relationship between self-truncated sampling and impression extremity (Prager et al., 2018; Prager & Fiedler, 2021).

Trait diagnosticity served as a continuous, trial-level moderator. While not manipulated experimentally, it was inferred post hoc from trait valence and extremity, in line with the density model framework (Unkelbach et al., 2008). The trait subsets were calibrated ex ante for each participant to approximate one of five base rate levels, inspired by Prager et al. (2018).

Political trust was included as a continuous covariate to explore its moderating effects on sampling behavior and confidence and was measured using a combined 14-item scale from the Perceived Political Self-Efficacy (P-PSE, $\Omega = 0.91$) and Political Efficacy Kurzskala (PEKS, $\Omega \geq 0.84$) instruments, based on their respective manuals. Items assessed participants' perceived capacity to influence politics and their trust in politicians. Responses were given on a 5-point Likert scale measuring agreement, with higher values indicating greater political trust.

Stimulus Material

Trait descriptions were developed using a hybrid procedure combining validated trait labels from Prager et al. (2018) with sentence generation via large language models (LLMs), followed by human filtering and optimization. Politically connotative traits were manually excluded. For each remaining trait, three sentence variants were generated using OpenAI's GPT4.5 API.

Sentences were then valence-rated in two rounds via GPT to increase reliability, removing those with inconsistent or mismatched polarity. The final pool included 73 reliably rated sentences (two per trait), selected for valence accuracy and minimal redundancy. To validate alignment between trait labels and sentence-level affective tone, Pearson correlations were computed between trait-level and sentence-level ratings, yielding a correspondence of $r = .90$.

Trait subsets were assembled using a Python-based optimization script that randomly generated 10 million possible 9×8 trait allocations. The final subsets were selected to minimize trait overlap between candidates while ensuring controlled positivity base rates (0.20, 0.33, 0.50, 0.67, and 0.80). These valence ratios were assigned across candidates to reflect realistic variability in how political figures are portrayed and were modelled as a continuous predictor.

Candidate portraits were generated via the ChatGPT DALL·E interface using prompts inspired by Oosterhof & Todorov (2008), emphasizing neutral, non-political appearance and consistency in likeability, competence, and morality. Images were manually reviewed and adjusted for consistency and realism. To validate perceptual balance, an independent GPT-4o session run without memory, rated each face across the three target dimensions. Visual inspection showed no systematic outliers or confounds. A final check using blind GPT-4o evaluation confirmed no resemblance to real politicians. Participants were explicitly told that the portraits were AI-generated and irrelevant for evaluation, serving only as mnemonic aids.

Effect Size Estimation

Selecting a realistic and defensible effect size was complicated by differences between our paradigm and prior studies. Although both Biella & Hütter (2024) and Ziegler & Fiedler (2024) address truncation and motivational framing, we based our estimation on Biella & Hütter (2024), whose design more closely matches ours, featuring multiple sources at once. However, our paradigm differs in three key ways: (1) participants evaluated political candidates rather than trust-game partners, (2) they were not primed with a future behavioral outcome, and (3) we used sentence-based trait descriptions instead of binary cues. These differences likely increase cognitive variability and reduce diagnostic clarity, attenuating expected effects.

Biella & Hütter reported a large effect (Cohen's $f = 0.60–0.75$), but due to our added stimulus complexity and lower outcome salience, we selected a more conservative estimate. Following Ziegler & Fiedler (2024), who report $\eta^2 = .06–.09$ (approx. Cohen's $f = [0.25; 0.31]$), we anchored our power analysis on Cohen's $f = 0.20$, balancing sensitivity with ecological realism.

Power Analyses

We used simulation-based methods in R to determine the required sample size for our main hypothesis (H2), focusing on the two-way interaction between sample size and goal condition. The model included fixed effects for sample size, condition, and diagnosticity, plus random intercepts for participants. Simulations with 1,000 iterations showed that with 100 participants, power to detect the two-way interaction reliably exceeded 80%, with the entire 95% CI above this threshold. This supports our planned sample size. For the three-way interaction (sample size × condition × diagnosticity; H3), power was insufficient: with 100 participants, estimates ranged from 40–60%, suggesting exploratory status and increased risk of a Type II error. Given the simplicity of H1 and the similarity of H4 to H2, no further analysis was deemed necessary.

Statistical Analyses

To test our hypotheses, a combination of linear regressions, mixed-effects models, correlation analyses, and visualizations was employed. Judgment strength was modeled as a function of the square root of sample size ($\sqrt{n}$) using both fixed- and random-effects approaches, including random slopes where appropriate (H1). To examine differences in sampling behavior and truncation as a function of motivational framing, we estimated models with interaction terms between $\sqrt{n}$ and goal condition (H2). Trait diagnosticity, operationalized through candidate-level valence base rates, was entered as a continuous moderator to assess its influence on sampling behavior and impression formation across conditions (H3). Finally, political trust was included as a between-subjects covariate to assess its moderating effect on confidence and sampling depth (H4). In all analyses, participants were modeled as random intercepts, and key results were visualized using smoothed regression plots. Supplementary models examined judgment deviation from sampled and population-level trait means to provide additional context. Competing model specifications were compared using likelihood-ratio tests and AIC, while model fit was furthermore evaluated through residual diagnostics and visual inspection.

Open Science

All model specifications were grounded in theory and validated through diagnostic checks, residual inspection, and model comparison using AIC and likelihood-ratio tests. Assumptions such as normality and homoscedasticity were visually inspected, and model complexity was justified by theoretical relevance. Confirmatory and exploratory analyses were clearly distinguished, with confirmatory results aligned to preregistered hypotheses. Analyses were conducted in R using reproducible, well-commented scripts, and a simulation-based power analysis with clearly documented assumptions supported the sample size justification. All data, scripts, and preregistration materials are available via persistent links provided in the Appendix.

Results

Hypothesis 1: Relationship Between Sample Size and Judgment Extremity

To test Hypothesis 1, a linear mixed-effects model was fitted to predict participants' average judgment extremity (|rating − 50|) based on the square root of the sample size (√n). The model included random intercepts and slopes for √n at the participant level. Descriptive analysis showed that participants in the hedonic condition had a mean judgment extremity of M = 21.73 (SD = 13.81). The effect of sample size on judgment extremity was tendentially positive but not statistically significant, b = 1.95, SE = 1.02, t(63) = 1.91, p = .061, while the marginal R^2 was .016, indicating that the fixed effect explained just a small portion of variance in extremity.

To explore potential non-linear trends, two alternative models were estimated: a linear mixed-effects model with a natural spline transformation of √n, and a Generalized Additive Model (GAM). In the spline model, the first component was non-significant (b = 7.09, SE = 5.20, p = .17), while the second component reached minor statistical significance (b = 6.43, SE = 3.18, p = .044), suggesting a possible non-linear association between √n and judgment extremity. The GAM model also included a smooth term for √n but showed no evidence of a significant effect, F(1.21, 1.39) = 0.24, p = .625. Explained variance remained similarly low in these two.

Model comparison using the Akaike Information Criterion (AIC) favored the spline model (AIC = 2299.35) over the linear model (AIC = 2303.44) and the GAM (AIC = 2315.12), although differences were modest and did not indicate substantial improvements in model fit. Taken together, these results offer weak evidence for a positive relationship between sample size and judgment extremity under hedonic motivation, and hint towards a more complex, non-linear relationship. The low explained variance values combined with the wide skew of extremity in each sample size (Figures 1.1 & 1.2) indicate that other factors such as diagnosticity need to be accounted for to better understand how sample size affects extremity.

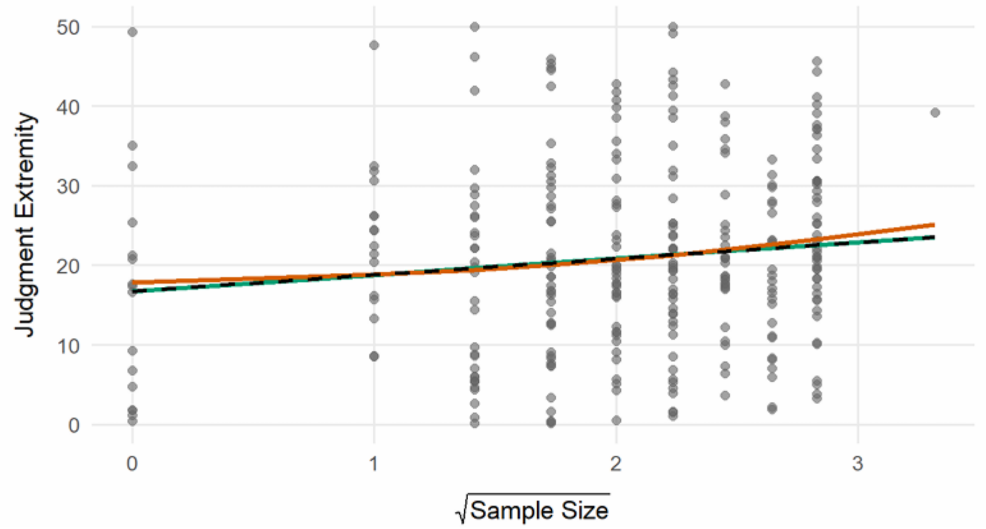

Figure 1.1

Predicted values from linear, spline, and GAM for judgment extremity as a function of √(n) in the hedonic condition. The linear model (dashed black) and GAM (green) suggest a weak positive trend. In contrast, the spline model (orange) indicates a slight non-monotonic U-shaped pattern, with lower extremity at intermediate sample sizes and increase at higher values.

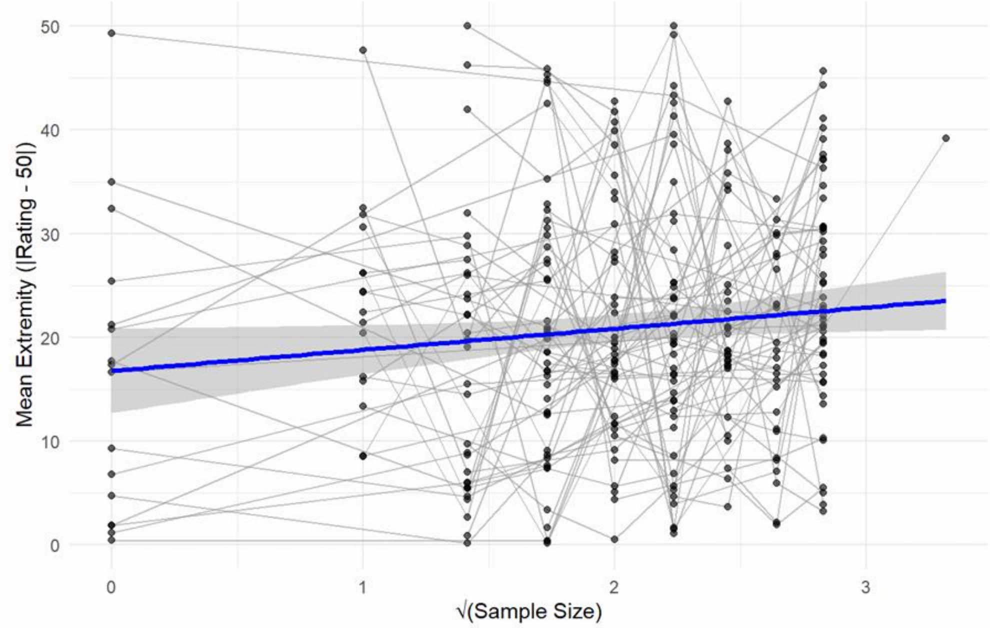

Figure 1.2

Spaghetti plot of judgment extremity across √n in the hedonic condition.Each line represents a participant's trajectory. While individual variability was high, the smoothed trend (blue line with 95% CI) indicates a weak positive association, consistent with the linear model results.

Hypothesis 2: Effect of Epistemic Goals on Sampling Depth and Judgment Strength

Descriptive analyses showed that participants in the epistemic condition sampled less traits ($M = 4.07$, $SD = 2.21$) than those in the hedonic condition ($M = 4.29$, $SD = 2.51$) and formed slightly less extreme judgments ($M = 21.26$, $SD = 14.58$), seemingly contradicting our theoretical assumptions. A linear mixed-effects model with judgment extremity as the outcome and $\sqrt{n}$, sampling condition, and their interaction as fixed effects with random slopes and intercepts for $\sqrt{n}$ revealed no significant interaction, $b = -0.10$, $SE = 1.52$, $t(64.6) = -0.06$, $p = .950$. The model explained similarly little variance (marginal $R^2 = .010$; conditional $R^2 = .166$).

Simple slopes indicated a positive association between $\sqrt{n}$ and extremity in both conditions, but only the hedonic slope reached significance ($b = 1.92$, $p = .04$; epistemic: $b = 2.02$, $p = .10$), contrary to expectations. Neither the spline model ($df = 2$) nor the GAM with condition-specific smooths showed significant interactions. The spline model showed the lowest AIC (7460.0), slightly outperforming the linear (7466.5) and GAM (7513.7) models – practically negligible. Predicted trajectories (Figure 2.1) showed a general increase in extremity with sample size in both conditions, slightly steeper under hedonic framing. Weak non-linearities emerged in the spline (Figure 2.2) and GAM fits but with no evidence of flattening in the epistemic condition.

An interesting observation that should however not be overinterpreted, is that the largest and furthermore statistically significant difference in judgment extremity in favour of our hypotheses emerges when participants rated candidates from whom they had sampled no traits. Also significant is the same difference when participants sampled the most amount of traits.

Model diagnostics indicated no major violations. Residuals were approximately homoscedastic and normally distributed, though some curvature and mild collinearity, mostly in the interaction term, were observed. A few influential points were present but did not bias estimates. In sum, these results offer no statistically significant evidence that motivation moderates this relation.

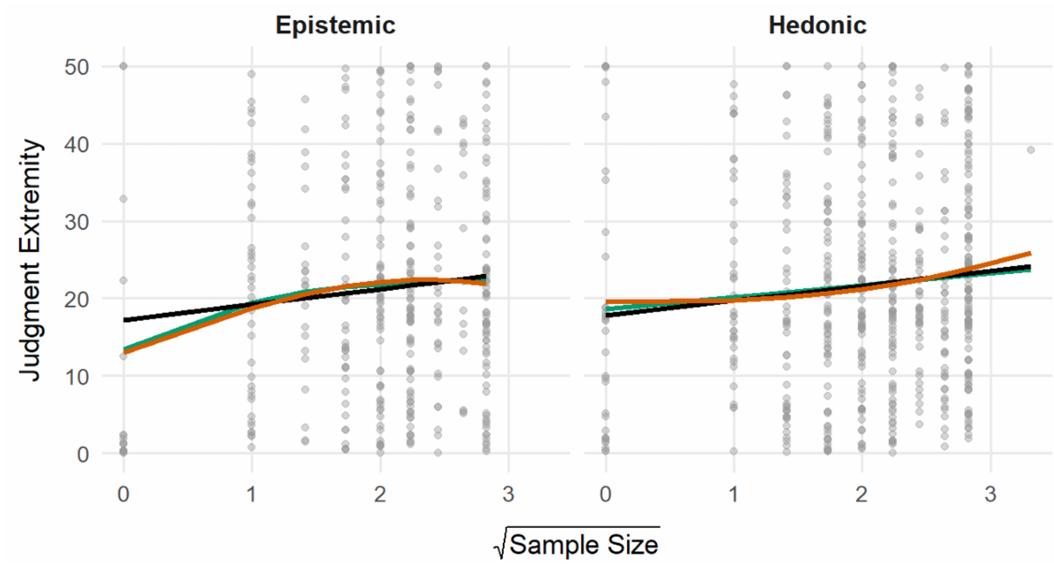

Figure 2.1

Predicted judgment extremity as a function of √n, separated by motivational condition. Data points represent individual ratings, with model-predicted lines overlaid for each condition. The linear model (black, dashed), spline model (orange), and Generalized Additive Model (green) show largely parallel trends across epistemic and hedonic conditions. While all models suggest increasing extremity with larger samples, this is more pronounced in the hedonic condition.

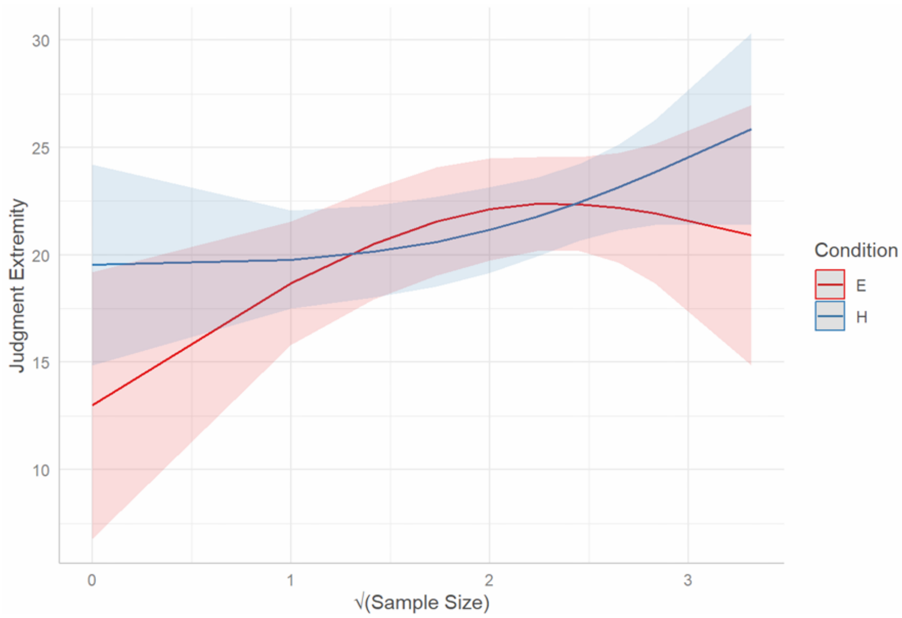

Figure 2.2

Predicted judgment extremity by √n and condition. Lines represent model estimates with 95% confidence intervals for epistemic (red) and hedonic (blue) sampling. The epistemic condition starts significantly lower, the lines then shortly converge, only to diverge at higher sample sizes.

Hypothesis 3: Moderation by Diagnosticity (Candidate Positivity)

To test Hypothesis 3, a linear mixed-effects model predicted candidate evaluations from the square root of sample size ($\sqrt{n}$), motivational condition, candidate positivity, and all their interactions. The model included random intercepts for participants. The three-way interaction did not reach statistical significance ($b = –16.30$, $SE = 10.12$, $t(910.5) = –1.61$, $p = .108$), though the effect trended in the hypothesized direction. The two-way interactions involving condition were statistically insignificant ($ps > .07$). The significant interaction between $\sqrt{n}$ and candidate positivity ($b = 38.19$, $SE = 7.80$, $t = 4.90$, $p < .001$), suggested that evaluation extremity increased more steeply with sampling depth for highly polarizing candidates. The model explained a moderate proportion of variance ($R^2 = .214$; conditional $R^2 = .279$. Visual inspection showed more extreme ratings to emerged as sample size increased, especially for polarizing candidates. However, this effect was consistent across motivational conditions.

To explore non-linear relationships, two additional models were estimated: a natural spline mixed-effects model with two degrees of freedom and a Generalized Additive Model (GAM) with smooth terms for $\sqrt{n}$ and candidate positivity. Neither model revealed significant interaction effects. Model comparison based on the Akaike Information Criterion (AIC) favored the spline model (AIC = 8334.2) over the linear model (AIC = 8345.1) and the GAM (AIC = 8574.3), though the differences did not indicate a substantial improvement in model fit.

Visual inspection of model-predicted trajectories revealed an overall increase in evaluative judgments with larger sample sizes across both conditions and diagnosticity levels. However, these patterns remained parallel across predictor levels, and no systematic interaction, flattening, or amplification, emerged. Descriptive trends did not support the notion that epistemic motivation attenuates the influence of candidate-level positivity in deeper sampling.

Model diagnostics revealed modest violations, including mild non-linearity, heteroskedasticity, and right-skewed residuals. While some influential cases were flagged, none exerted strong leverage. Most notably, multicollinearity was severe, with VIFs > 10 for all predictors involving candidate positivity, suggesting inflated standard errors that may have obscured interaction effects. Its effects will be further discussed down the line. Taken together, the results provide no statistical support for the hypothesis that diagnosticity moderates the effect of sample size and motivational framing on evaluative judgments. Nonetheless, interpretability is constrained by multicollinearity, which substantially limits confidence in the null findings.

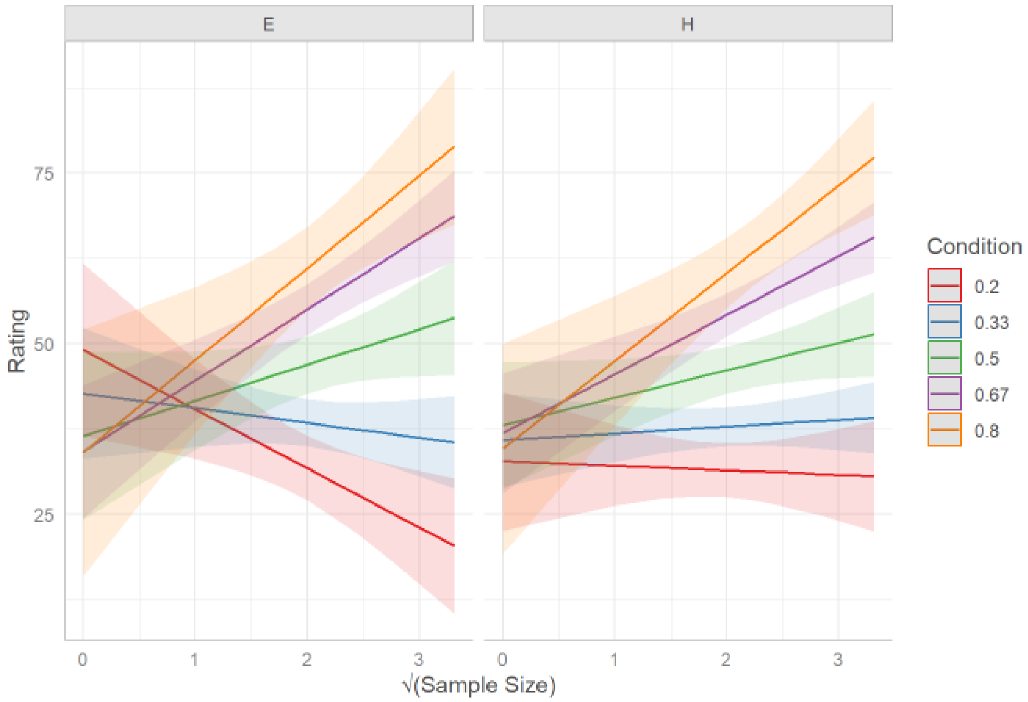

Figure 3

Predicted candidate ratings from the spline mixed-effects model (H3), plotted as a function of √n, separately by motivational condition and candidate positivity levels. Lines represent predicted values; shaded areas indicate 95% confidence intervals. More extreme trait profiles show stronger rating polarization with increasing sample size, whereas moderate profiles remain stable. The pattern is comparable across conditions, suggesting that the element of diagnosticity in both epistemic and hedonic contexts has an independent, unmoderated effect.

Hypothesis 4: Moderating Role of Political Trust (Efficacy)

To test whether political trust moderates the relationship between sample size, motivational condition, and confidence in candidate evaluations, we fitted a linear mixed-effects model including all two-way and three-way interactions between $\sqrt{n}$, condition, and internal efficacy, with random intercepts for participants. The three-way interaction did not reach statistical significance ($b = -0.63$, $SE = 1.66$, $t = -0.38$, $p = .707$), and none of the two-way interactions involving condition or efficacy were significant ($ps > .12$) – although all this was to be expected based on our power analyses. The model explained a modest proportion of variance (marginal $R^2 = .097$; conditional $R^2 = .375$). Confidence increased with sample size overall ($b = 7.60$, $SE = 1.32$, $t = 5.75$, $p < .001$), but this effect was not moderated by motivational or political trust.

Descriptively, however, an immediate reversal of the efficacy extremes is visible, and could serve to partially explain the change from converging to diverging lines in the hedonic condition (Figure 4.1). For small sample sizes, perceived efficacy seems to have played no role in judgement confidence in the hedonic condition. For larger sample sizes in the hedonic condition, the effect we originally postulated for the epistemic condition emerges: Those with higher perceived efficacy grow clearly more confident in their judgments. In the epistemic condition, those with the lowest efficacy were the most confident in their judgments of smaller samples, whereas efficacy ceased to play a role, the larger the sample sizes became. We will come to address these interesting, complex interactions, which become clearer in Figure 4.2.

Model diagnostics for all H4.1 models showed good fit. Residuals were approximately normal, with slight skew and minor heteroskedasticity. Influential points were flagged but did not unduly affect model estimates. Collinearity was acceptable in the linear and spline models (VIFs < 10), suggesting interpretable coefficients. A spline-based model (df = 2) was also estimated to test for non-linear moderation by efficacy. This model showed improved fit (AIC

= 7785.5 vs. 7831.6), and revealed a significant non-linear interaction between √n and internal efficacy (b = 17.51, SE = 7.64, p = .022) but no significant three-way interaction. A GAM with condition-specific smooths explained 5.79% deviance and was not favored by AIC (8059.3).

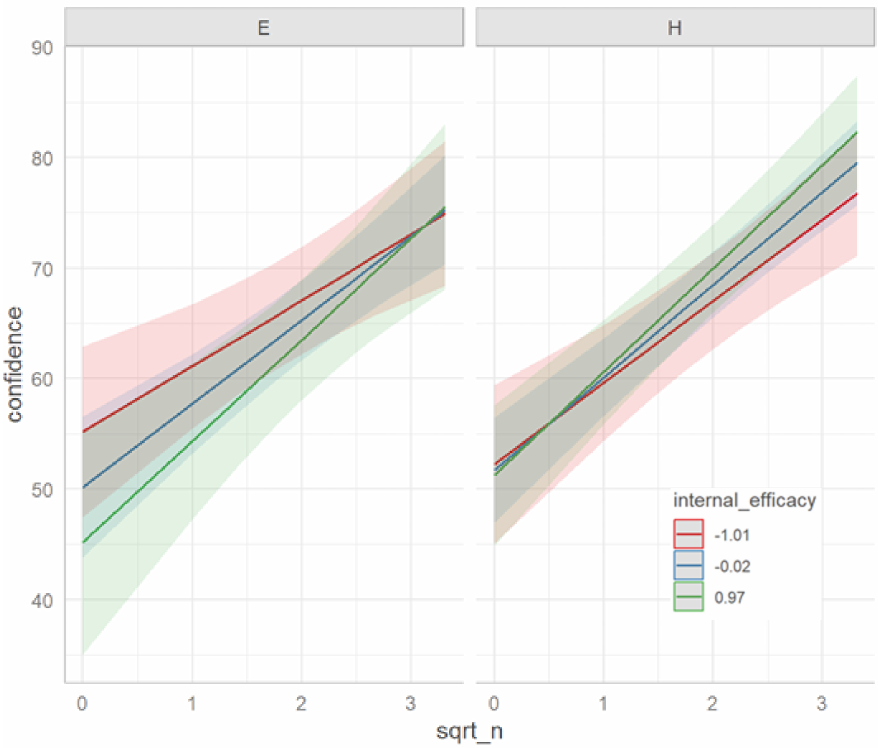

Figure 4.1: Predicted confidence as a function of √n, shown by motivational condition (epistemic vs. hedonic) and internal political efficacy (−1 SD, mean, +1 SD). Shaded areas indicate 95% confidence intervals. Confidence rose steeply with sample size in all groups, but only in the epistemic condition did trust amplify this.

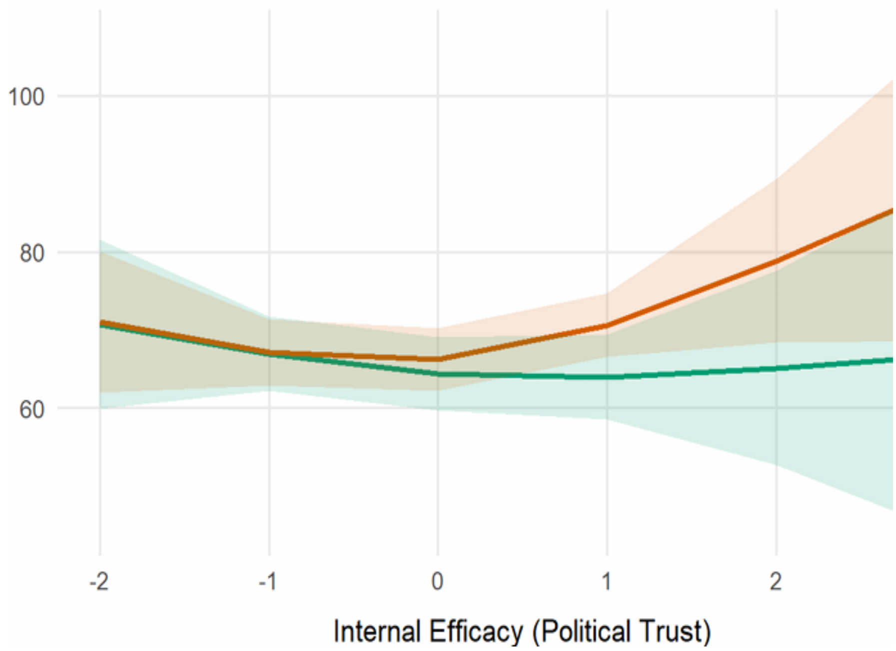

Figure 4.2: Predicted confidence as a function of internal efficacy across sample sizes. Shaded areas indicate 95% confidence intervals. In the epistemic condition (orange), higher trust increased confidence, whereas hedonic sampling (green) was largely unaffected by perceived efficacy.

The same model structure was applied to judgment extremity. The three-way interaction again did not approximate statistical significance ($b = 1.85$, $SE = 1.36$, $t = 1.36$, $p = .175$), nor did most lower-order interactions ($ps > .13$), except for a main effect of $\sqrt{n}$ ($b = 2.59$, $SE = 1.08$, $t = 2.39$, $p = .017$), indicating that extremity increased with sampling in both conditions equally. Descriptively, the same reversal in the extremes of perceived efficacy can be observed between the two conditions, as well as the pattern of divergence in the epistemic versus convergence in the hedonic condition for different levels of perceived efficacy (Figure 4.3 The alignment of patterns between the two hypotheses on efficacy further motivates a reflection of their meaning. The model explained little variance (marginal $R^2 = .020$; conditional $R^2 = .141$). Simple slopes suggested that for those with low political trust (–1 SD), sample size predicted more extreme judgments in the epistemic condition ($b = 3.08$, $p = .02$), but not in the hedonic condition ($b = 0.29$, $p = .80$). At mean levels, sample size predicted extremity in both conditions ($ps < .03$).

To explore non-linear patterns, we fitted a spline model ($df = 2$) and a GAM with condition-specific tensor interaction smooths. The spline model slightly improved fit (AIC = 7394.7) compared to the linear model (AIC = 7424.4), while the GAM performed worse (AIC = 7474.8) and explained <1% deviance. No significant non-linear interactions emerged. Model diagnostics revealed no major violations, although some heteroskedasticity and slight right-skew were observed. Importantly, VIFs for the spline model remained below 10, supporting reliable interpretation. A likelihood ratio test favored including random slopes for $\sqrt{n}$ ($\chi^2(2) = 9.75$, $p = .0076$). Altogether, results for H4.2 provide only suggestive evidence that political trust moderates the relationship between sample size, motivational condition, and judgment extremity. Although simple slopes suggested a possible crossover pattern, with sample size increasing extremity more under epistemic framing at low trust and more under hedonic framing at high trust, these differences were not statistically supported by the full interaction.

Hypothesis E – Exploratory analyses

To explore potential differences in evaluative accuracy between motivational conditions, two metrics of judgment error, known as actuarial measures, were analyzed: (1) the absolute deviation between participants' ratings and the mean valence of sampled traits (sample error), and (2) the deviation from the population mean valence of all traits in a candidate's pool (population error). For each metric, linear mixed-effects models were fitted with random intercepts for participants, predicting judgment error from √n, condition, and their interaction.

For sample error, the model yielded a marginal R^2 of .038 and a conditional R^2 of .088. A significant main effect of √n (b = 6.49, SE = 2.47, *p* = .009) indicated that increasing sample size actually increased the deviation from sampled traits. However, no significant interaction with condition (b = 3.71, *p* = .242), or main condition effect (b = –8.11, *p* = .251) was observed. Diagnostics showed only minor overall violations of linearity and slight collinearity inflation. While statistically insignificant, a trend towards better approximation of true valence at sample level can be visually observed (Figure E1), in line with previous findings on hedonic sampling. For population error, the pattern was similar (Figure E2). The model revealed a weaker but still significant main effect of √n (b = 4.11, SE = 1.85, *p* = .027), indicating greater deviation with more samples. The interaction with condition was non-significant (b = 3.06, *p* = .188), and no main effect of condition emerged (*p* = .2). Marginal R^2 was .032; conditional R^2 was .071. Model diagnostics were comparable to the sample error model and did not reveal critical issues. Descriptive inspection showed a similar trend to the sample error condition, which contradicts pervious research, as well as the implicit assumption that epistemic sampling is more accurate.

Taken together, these results suggest that drawing more information paradoxically increased the distance between final judgments and both the sampled and actual candidate profiles, contrary to statistical consensus such as that larger samples lead to more accurate impressions.

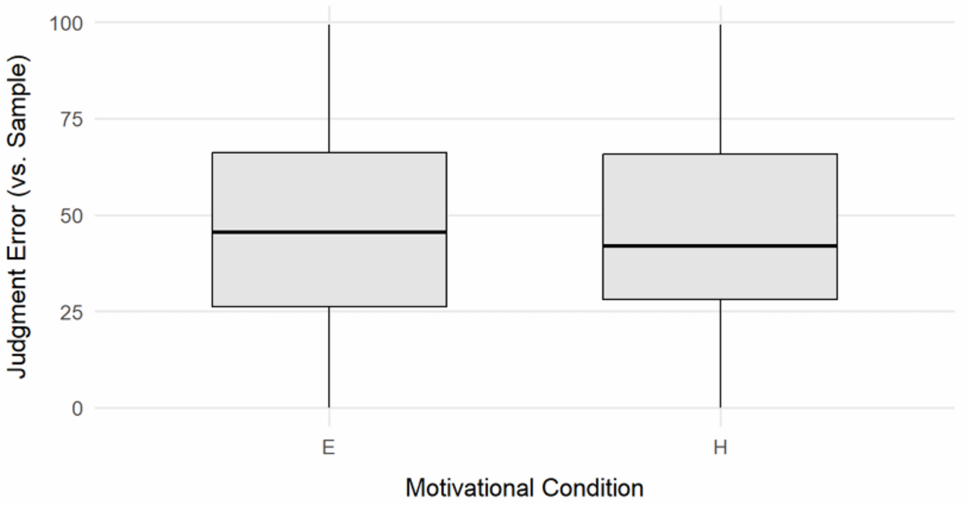

Figure E1

Judgment error relative to sample trait valence by motivational condition. Boxplots show the absolute deviation between participants' ratings and the average valence of all traits in the respective candidate's full pool. Although not statistically significant, a descriptively lower median error in the hedonic condition mirrors the pattern in population-level accuracy, tentatively challenging the assumption that epistemic goals yield more accurate impressions.

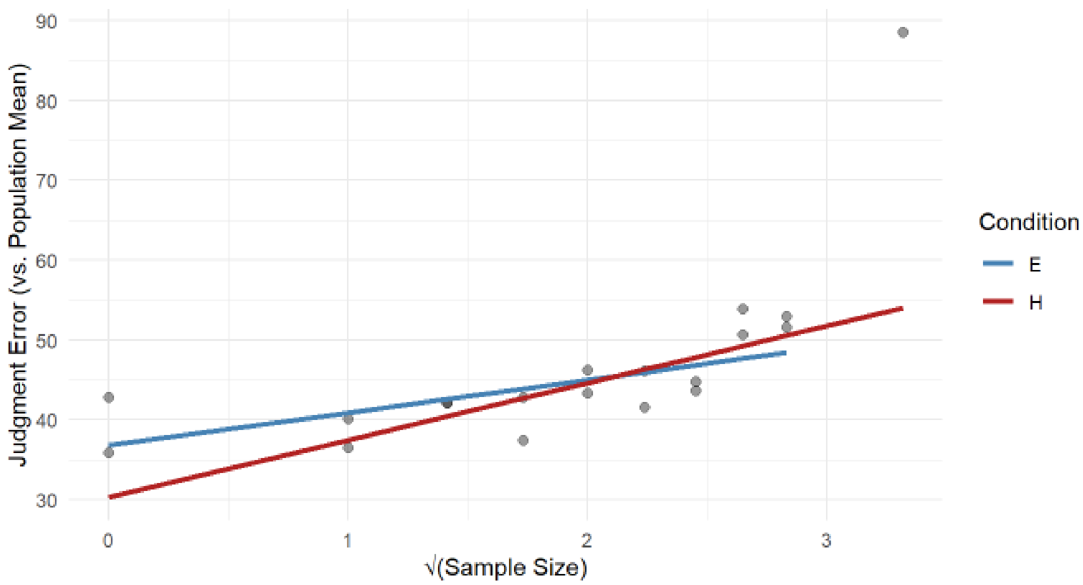

Figure E2

Predicted judgment error as a function of √n, by motivational condition. Lines represent model-based predictions from a linear mixed-effects model, and dots show aggregated participant means per condition and sample size. Contrary to normative expectations, judgment error increased with more sampled information in both this and the sample mean version of the graph.

“They may be the best that nature could do under the constraints within which organisms evolve, but the outcome may be far from ideal.”

– Noam Chomsky, *Powers and Prospects* (1996)

Discussion

This thesis examined whether explicitly induced epistemic goals can reduce primacy bias in political impression formation. Building on recent work by Biella and Hütter (2024) and Ziegler and Fiedler (2024), we adapted a sequential sampling paradigm to simulate a political decision-making context, asking whether motivational framing can influence sampling behavior, judgment strength, and confidence when evaluating political candidates. Our central hypothesis (H2) posited that epistemic motivation would increase sampling depth and reduce impression extremity compared to hedonic motivation, thereby attenuating primacy bias. To validate and extend prior work, three additional hypotheses were formulated: H1 tested the expected negative relationship between sample size and judgment extremity under hedonic framing; H3 explored whether trait diagnosticity moderates the link between sample size and evaluation extremity across conditions; and H4 examined whether participants’ general political trust moderates the effect of sampling and motivational framing on confidence and judgment extremity. Beyond these confirmatory analyses, we conducted exploratory tests of evaluative accuracy. Specifically, we examined whether judgments aligned more closely with the sampled traits or with the candidate's population-level trait profile, and whether this alignment varied as a function of motivational condition. These analyses served to assess whether epistemic goal framing yields more veridical, reality-congruent evaluations, representing a key question for understanding the quality of political impression formation.

Contrary to expectations, Hypothesis 1 was not statistically supported. We were not able to reproduce the significantly negative correlation between sample size and judgment extremity found in Ziegler & Fiedler (2024) as well as Biella & Hütter (2024), attributed to a primacy bias in sampling contexts where the hedonic motive of satisficing is salient. While descriptive trends suggested that larger sample sizes may be associated with greater judgment extremity, the effect did not reach conventional significance thresholds. A spline model indicated a weak U-shaped pattern, with extremity dipping slightly at intermediate sample sizes and increasing again at higher levels. This may reflect a cognitive mechanism described in Denrell's (2005) sampling model: while early sampling corrects extreme impressions, further sampling can reintroduce extremity when contradicting traits appear later. The U-shape is also faintly reminiscent of the Dunning-Kruger effect, where subjective certainty, akin to impression extremity, dips as people learn enough to recognize complexity. Another explanation may lie in task design. Unlike Ziegler & Fiedler (2024), participants in this study rated all candidates only at the end of the sampling phase, adopting Biella & Hütter's structure. Given the cognitive load of tracking nine candidates, this may very well have introduced substantial memory decay or information overload, effectively muting or distorting primacy effects – an issue that might explain other patterns of our results, too. Earlier-sampled traits may not have been adequately recalled during final judgments, flattening observed extremity differences. Lastly, this first hypothesis did not take into statistical consideration the pivotal role of diagnosticity. Without it, the relationship between sampling depth and judgement strength remains fuzzy, as was illustrated in the back-and forth between Norton et al. (2007, 2013) and Ullrich et al. (2013). In total, our H1 results do not mimic those of previous studies we based our operationalization on, hinting towards more than one manipulations in experimental design and thereby limiting, but in no way annuls the strength of conclusions to be drawn from the rest of our hypotheses.

Our central hypothesis (H2) anticipated that epistemic goal framing would increase sampling and reduce judgment extremity. However, no significant differences between conditions emerged, and trends slightly contradicted predictions, in that hedonic participants sampled marginally more traits. One explanation is that the motivational framing was too subtle to induce a robust epistemic mindset. Whereas our epistemic condition relied on a contextual goal induction via brief instructional framing, Biella and Hütter (2024) embedded epistemic motivation structurally into the task by rewarding those participants who formed a justified and accurate judgment about all targets. Another key factor may be unmodeled variance: In both conditions, participant behavior showed large heterogeneity in sampling and evaluation patterns. Despite controlling for known moderators, none of the models could capture the full extent of this variability. Future analyses may benefit from mixture modeling or latent profile analysis to identify subgroups with qualitatively distinct strategies. However, as it stands, our sample size remains too small to fit in a model with complex parameters such as latent classes.

Though not originally postulated, the largest difference observed between conditions appearing in the judgment of candidates, for whom participants did not manage to sample any traits, suggests that there is perhaps a positive effect of epistemic framing, in that it moderates people's confidence in states of complete ignorance – states that, in an ever more informative world, we necessarily often find ourselves in. When taking into account the results from our explorative analyses however, one quickly shies away from the conclusion that epistemic framing lead to more accurate judgments. This could possibly mean that, entirely contrary to our line of argumentation but entirely in line with findings from Prager & Fiedler (2021), epistemic framing is disadvantageous when evaluating small samples, overriding individuals' evolutionarily crafted urge to utilize early diagnostic traits by prioritizing statistical reliability. The better approximation of the population valence mean in the hedonic condition, may very well be attributed to precisely this factor, as we will shortly discuss in the exploratory section.

Hypothesis 3 investigated whether trait diagnosticity moderates the relationship between sample size and judgment extremity, based on the assumption that more diagnostic traits amplify the weight of early impressions and reduce the marginal value of additional sampling. Although the full three-way interaction between diagnosticity, sample size, and condition was not statistically significant, a robust two-way interaction emerged, in which candidate-level positivity strongly predicted impression extremity, especially as sample size increased. This suggests that more extreme candidates elicited increasingly polarized evaluations as more traits were sampled – an effect consistent with density-based models of impression formation (Unkelbach et al., 2008; Fiedler, 1996). From a theoretical standpoint, this partial support highlights the importance of content in sampling decisions: more diagnostic candidates may trigger confirmation-like sampling or lead to earlier commitment, thereby promoting inaccuracies. An interesting observation, however, pertains to the much steeper declines in the ratings of unpopular candidates with population base rates of 0.20 and 0.33 in the epistemic condition, possibly implying that, although epistemic samplers initially give the benefit of the doubt, they are harsher critics of negative samples, when ample evidence is presented to them.

The lack of a significant three-way interaction with motivational condition might reflect several limitations. First, the complex interaction terms in the model produced multicollinearity, which significantly diminished sensitivity to subtle effects, and second, these interaction hypotheses likely suffered from low statistical power. Future studies could address these issues by recruiting larger samples to detect interactions robustly, since different avenues of counteracting multicollinearity such as Residualization, Regularization and Principal Component Analyses (PCA) simultaneously would have reduced the interpretability of the original constructs. While our results stop short of a confirmation, they trend towards the well-established notion that diagnosticity is an important moderator of sample size and extremity.

Hypothesis 4 proposed that political trust would moderate the relationship between sample size, motivational framing, and confidence in candidate evaluations. While no significant three-way or two-way interactions involving trust and condition emerged, descriptive trends and non-linear analyses suggest more nuanced dynamics. Confidence increased overall with sample size, but the degree of this increase varied with both motivation and efficacy levels in descriptive plots. In the epistemic condition, high-trust individuals showed a steeper rise in confidence as they sampled more traits. Conversely, under hedonic motivation, trust effects emerged only at higher sample sizes, with confident impressions forming primarily among high-trust participants. These crossover trends were mirrored in judgment extremity. Notably, a spline model revealed a significant non-linear interaction between sample size and efficacy, suggesting that trust effects may emerge only at particular sampling depths. These findings challenge the idea that epistemic goals categorically foster better processing.

For individuals low in trust, epistemic cues may be met with skepticism or resistance, undermining deliberative intentions and triggering premature confidence or polarization. In contrast, high-trust individuals may be more receptive to either framing, using hedonic cues as a license for intuitive evaluation, while epistemic cues are seen as a call for thorough scrutiny. Rather than uniformly increasing rational engagement, trust may shape which motivational frames are interpreted as legitimate or credible. The absence of significant interactions cautions against definitive claims, but the descriptive reversals highlight a possible boundary condition for epistemic interventions, in that they may work best on audiences already inclined to trust the system they are asked to scrutinize. Future research should aim to experimentally manipulate trust and incorporate more ecologically valid trust measures, such as behavioral trust games or strategic belief elicitations. While Hypothesis 4 was not supported in the inferential sense, the observed non-linear moderation and consistent crossover patterns suggest that trust plays a complex, context-dependent role in how motivational goals shape impressions.

Finally, our exploratory findings run counter to the assumption that more information necessarily yields more accurate evaluations. One explanation may lie in the accumulation of cognitive noise: as participants process more traits, they may lose track of base rates or anchor on emotionally salient outliers, thereby drifting further from both the sampled average and the population truth (Fiedler, 2000; Carroll, 1982). Alternatively, epistemic participants motivated by fairness or comprehensiveness may have engaged in cognitive balancing, giving disproportionate weight to countervalent or late-sampled traits. While this strategy may promote internal coherence or perceived objectivity, it could paradoxically reduce alignment with the actual data, highlighting a possible trade-off between perceived epistemic diligence and numerical accuracy. Though contrary to our hypothesized direction, these findings, at least partially, correspond to those of Prager & Fiedler (2021) in regards to actuarial accuracy. They also found participants under the hedonically predisposed self-truncation condition to be more accurate in approximating the sample valence. However, this effect reversed in favour of the more epistemically inclined auto-truncation condition, contrary to our current findings, indicating that the conservative mentality of those participants to not read into initial diagnostic cues paid off in the long run, leading to more globally founded impressions. A very plausible explanation to this effect has to do with differences in paradigm operationalization. The much higher complexity of our traits combined with the fact that participants in our study had to wait until the very end to translate their impressions into candidate judgments, meant that any epistemic advantage might have been eroded by information overload and reduced data alignment. Ironically, we touched upon this very problem in the introduction to this thesis, but were seemingly unable to properly account for it in our own design, emphasizing the blindsight and lack of awareness surrounding this issue in our everyday lives. Further evidence supporting this line of explanation can be found in Figure E2. The epistemic condition does in fact lead to actuarily better judgments in the largest sample sizes, though complexity diminishes this effect.

Taken together, the findings suggest that while motivational framing may shape participants' intentions, it does not reliably translate into measurable differences in sampling behavior, judgment extremity, or confidence, at least not under the contextual framing used in this study. Sampling decisions appeared to be driven more by trait-level diagnosticity and individual differences than by the global motivational condition. This points to the limitations of subtle framing manipulations in influencing complex cognitive tasks, especially when participants are managing large amounts of information under memory constraints, in a far from natural setting. Several design features may have contributed to the null and mixed results. Most notably, the decision to rate all candidates only after sampling may have reduced the salience of early trait impressions and increased reliance on imperfect memory. Low variance in political trust and limited statistical power further constrained the detection of moderation effects. Furthermore, the much greater complexity of our traits compared to Biella & Hütter (2024) and even Ziegler & Fiedler (2024) not only exacerbated the delayed and compounded rating of candidates, but also added to the demandingness of the task in their own right. Future research could address these issues through improved operationalizations, stronger manipulations, and larger samples.

Finally, the unnaturally dense sampling context which participants have thus far been subjected to in impression formation research might prevent more organic traits such as the ones conceived in this study to be processed adequately. This raises the question of whether an insightful union of social psychology and political science is even possible within the confounds of short-form online experiments, or whether we need to bring social psychology out to the real world, conducting real-time field experiments on political impression formation. Despite these limitations, the present study extends sequential sampling paradigms into more ecologically valid political contexts and highlights the challenges of linking motivational states to information-processing behavior. Understanding the sampling dynamics examined in this empirical study remains a key step toward improving the quality of political decision-making.

Beyond issues of operationalization, another factor that could stand in the way of clearer evidence and has yet to be addressed, is data quality. While participants who sampled too little and trials where participants spent less than 1000ms exposed to a trait were removed prior to statistical analyses, the temporal exclusion criterion might have been too lenient, given that a typical trait looked something like this: “The candidate repeatedly seeks out conversations with experts in order to stay fully informed about current developments.” In total, for almost 15% of traits sampled, participants took less than 3000ms (3 seconds) before moving on to the next trait or candidate. This motivated a reprocessing and reanalysing of collected data. To counteract this in future research, a cooldown period between sampled traits could be introduced to facilitate information integration and memory consolidation processes. Manually inspecting and removing hundreds of problematic traits, 7 participants were removed from the final analysis due to lack of remaining traits sampled, leaving us with 96 participants and four traits sampled per candidate per participant on average. Running a post-hoc power analysis adjusted for the lower sample size and data points yielded a power of 82.20% [79.69; 84.52]. Dishearteningly, after rerunning our set of analyses, results were practically identical to before.

Thus, we can confidently conclude that (1) either along the way of tinkering with the original operationalizations of Ziegler & Fielder (2024) and Biella & Hütter (2024), a central structural component was altered to a degree that explains these deviations in our nomological network’s dynamics, or (2), the political context seems to decidedly differ from the more neutral impression formation exercises lacking strong implicit associations with confounding factors. While the bridge from social psychology to political science was not crossed in this attempt, light was perhaps shed on some of the intricate interplays between shared concepts, such as political efficacy, judgment extremity and sampling behavior, hopefully inspiring and informing future research on what changes need to be made to better understand political impression formation – a matter that affects us all, not just on a scientific, but also on a moral dimension.

References

Anderson, N. H. (1981). *Foundations of information integration theory*. Academic Press.

Arigapudi, S., Heller, Y., & Schreiber, A. (2023). Heterogeneous noise and stable miscoordination. *arXiv*. https://doi.org/10.48550/arXiv.2305.10301

Arnold, M., Goldschmitt, M., & Rigotti, T. (2023). Dealing with information overload: A comprehensive review. *Frontiers in Psychology, 14*, Article 1122200. https://doi.org/10.3389/fpsyg.2023.1122200

Asch, S. E. (1946). Forming impressions of personality. Journal of Abnormal and Social Psychology, 41(3), 258–290. https://doi.org/10.1037/h0055756

Biella, M., & Hütter, M. (2024). Navigating the social environment: Linking motivations, impressions, and behaviors via a sampling approach. Personality and Social Psychology Bulletin, Early View. https://doi.org/10.1177/01461672241273237

Bromme, L., Rothmund, T., & Caprara, G. V. (2020). A translation and validation of the Perceived Political Self-Efficacy (P-PSE) Scale for the use in German samples. *Measurement Instruments for the Social Sciences, 3*, Article 13. https://doi.org/10.1186/s42409-020-00013-4

Brunswik, E. (1956). Perception and the representative design of psychological experiments. University of California Press. https://www.jstor.org/stable/jj.8501445.4

Caplin, A., Dean, M., & Leahy, J. (2017). Rationally inattentive behavior: Characterizing and generalizing Shannon entropy. NBER Working Paper Series, No. 23652. National Bureau of Economic Research. https://www.nber.org/papers/w23652

Caplin, A., Dean, M., & Martin, D. (2011). Search and satisficing. American Economic Review, 101(7), 2899–2922. https://doi.org/10.1257/aer.101.7.2899

Carroll, J. S. (1982). [Review of the book Foundations of information integration theory, by N. H. Anderson]. The American Journal of Psychology, 95(4), 708–711. https://doi.org/10.2307/1422202

Chen, F., & Krajbich, I. (2018). Biased sequential sampling underlies the effects of time pressure and delay in social decision making. Nature communications, 9(1), 3557. https://doi.org/10.1038/s41467-018-05994-9

Chomsky, N. (1996). *Powers and prospects: Reflections on human nature and the social order*. Pluto Press.

Condorcet, M. de. (1785/1994). *Essay on the application of analysis to the probability of majority decisions* (I. McLean & F. Hewitt, Trans.). In I. McLean & A. B. Urken (Eds.), *Classics of social choice* (pp. 91–112). University of Michigan Press.

Cottam, M. L., Mastors, E., Preston, T., & Dietz, B. (2022). Introduction to political psychology (4th ed.). Routledge. https://doi.org/10.4324/9780429244643

Fiedler, K. (2000). Beware of samples! A cognitive-ecological sampling approach to judgment biases. Psychological Review, 107(4), 659–676. https://doi.org/10.1037/0033-295X.107.4.659

Denrell, J. (2005). Why most people disapprove of me: Experience sampling in impression formation. Psychological Review, 112(4), 951–978. https://doi.org/10.1037/0033-295X.112.4.951

De Soete, G., & Winsberg, S. (1993). A Thurstonian pairwise choice model with univariate and multivariate spline transformations. Psychometrika, 58(2), 233–256. https://doi.org/10.1007/BF02294575

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. Psychological Bulletin, 130(6), 959–988. https://doi.org/10.1037/0033-2909.130.6.959

Dhami, S., al-Nowaihi, A., & Sunstein, C. R. (2019). Heuristics and public policy: Decision-making under bounded rationality. Studies in Microeconomics, 7(1), 7–58. https://doi.org/10.1177/2321022219832148

Ditto, P. H., Pizarro, D. A., & Tannenbaum, D. (2009). Motivated moral reasoning. In D. M. Bartels, C. W. Bauman, L. J. Skitka, & D. L. Medin (Eds.), *Moral judgment and decision making* (Vol. 50, pp. 307–338). Academic Press. https://doi.org/10.1016/S0079-7421(08)00410-6

Druckman, J. N., Peterson, E., & Slothuus, R. (2013). How elite partisan polarization affects public opinion formation. American Political Science Review, 107(1), 57–79. https://doi.org/10.1017/S0003055412000500

Fiedler, K. (2017). What Constitutes Strong Psychological Science? The (Neglected) Role of Diagnosticity and A Priori Theorizing. *Perspectives on Psychological Science*, *12*(1), 46-61. https://doi.org/10.1177/1745691616654458

Giavazzi, M. (2023). The epistemic responsibilities of voters: Towards an assertion-based account. Journal of Moral Philosophy, 20(1), 111–131. https://doi.org/10.1163/17455243-20223791

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. Psychological Review, 103(4), 650–669. https://doi.org/10.1037/0033-295X.103.4.650

Goodin, R. E., & Spiekermann, K. (2018). An epistemic theory of democracy. Oxford University Press. https://doi.org/10.1093/oso/9780198823452.001.0001

Groskurth, K., Nießen, D., Rammstedt, B., & Lechner, C. M. (2021). An English-language adaptation and validation of the Political Efficacy Short Scale (PESS). *Measurement Instruments for the Social Sciences, 3*, Article 1. https://doi.org/10.1186/s42409-020-00018-z

Hannon, M. (2023). The politics of post-truth. *Critical Review, 35*(1–2), 40–62. https://doi.org/10.1080/08913811.2023.2194109

Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication, 59*(1), 19–39. https://doi.org/10.1111/j.1460-2466.2008.01402.x

Jung, J., & Mittal, V. (2020). Political identity and the consumer journey: A research review. Journal of Retailing, 96(1), 55–73. https://doi.org/10.1016/j.jretai.2019.09.003

Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. Psychological Review, 104(2), 344–366. https://doi.org/10.1037/0033-295X.104.2.344

Kubrick, S. (1968). *2001: A space odyssey*. Metro-Goldwyn-Mayer.

Lau, R. R., Andersen, D. J., & Redlawsk, D. P. (2008). An exploration of correct voting in recent U.S. presidential elections. *American Journal of Political Science, 52*(2), 395–411. https://doi.org/10.1111/j.1540-5907.2008.00319.x

Lau, R. R., & Redlawsk, D. P. (1997). Voting correctly. American Political Science Review, 91(3), 585–598. https://doi.org/10.2307/2952076

List & Goodin (2002). Epistemic democracy: Generalizing the Condorcet Jury Theorem. Journal of Political Philosophy, 10(1), 19–34. https://doi.org/10.1111/1467-9760.00128

Mckelvey, R.D., Palfrey, T.R. Quantal Response Equilibria for Extensive Form Games. *Experimental Economics* **1**, 9–41 (1998). https://doi.org/10.1023/A:1009905800005

Mechtenberg, L., & Tyran, J.-R. (2019). Voter motivation and the quality of democratic choice. Games and Economic Behavior, 116, 241–259. https://doi.org/10.1016/j.geb.2018.12.002

Norton, M. I., Frost, J. H., & Ariely, D. (2007). Less is more: The lure of ambiguity, or why familiarity breeds contempt. *Journal of Personality and Social Psychology, 92*(1), 97–105. https://doi.org/10.1037/0022-3514.92.1.97

Norton, M. I., Frost, J. H., & Ariely, D. (2013). Less is often more, but not always: Additional evidence that familiarity breeds contempt and a call for future research. *Journal of Personality and Social Psychology, 105*(6), 921–923. https://doi.org/10.1037/a0034379

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Paulos, J. A. (2003). *A Mathematician Plays the Stock Market*. Basic Books.

Peng, M., Xu, Z., & Huang, H. (2021). How does information overload affect consumers' online decision process? An event related potentials study. Frontiers in Neuroscience, 15, Article 695852. https://doi.org/10.3389/fnins.2021.695852

Pothos, E. M., & Busemeyer, J. R. (2022). Quantum cognition. Annual Review of Psychology, 73, 749–778. https://doi.org/10.1146/annurev-psych-033020-123501

Poulsen, O., & Sonntag, A. (2020). Time pressure and focal points in coordination games: Experimental evidence. Games and Economic Behavior, 123, 149–166.

Prager, J., & Fiedler, K. (2021). Forming impressions from self-truncated samples of traits: Interplay of Thurstonian and Brunswikian sampling effects. Journal of Personality and Social Psychology, 121(3), 474–497. https://doi.org/10.1037/pspa0000274

Prager, J., Krueger, J. I., & Fiedler, K. (2018). Towards a deeper understanding of impression formation – New insights gained from a cognitive ecological perspective. Journal of Personality and Social Psychology, 115(3), 379–397. https://doi.org/10.1037/pspa0000123

Prats, M., Smid, S., & Ferrin, M. (2024). *Lack of trust in institutions and political engagement: An analysis based on the 2021 OECD Trust Survey* (OECD Working Papers on Public Governance, No. 75). OECD Publishing. https://doi.org/10.1787/83351a47-en

Reeder, G. D., & Brewer, M. B. (1979). A schematic model of dispositional attribution in interpersonal perception. *Psychological Review, 86*(1), 61–79. https://doi.org/10.1037/0033-295X.86.1.61

Regenwetter, M., Ho, M.-H. R., & Tsetlin, I. (2007). Sophisticated approval voting, ignorance priors, and plurality heuristics: A behavioral social choice analysis in a Thurstonian framework. Psychological Review, 114(4), 994–1014. https://doi.org/10.1037/0033-295X.114.4.994

Salant, Y., & Cherry, J. (2020). Statistical inference in games. *Econometrica, 88*(5), 1961–1986. https://doi.org/10.3982/ECTA17105

Simon, H. A. (1978). Rational decision-making in business organizations. Nobel Lecture, December 8, 1978. Retrieved from https://www.nobelprize.org/prizes/economic-sciences/1978/simon/lecture/

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. Journal of Personality and Social Psychology, 52(4), 689–699. https://doi.org/10.1037/0022-3514.52.4.689

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*(3), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

Ullrich, J., Krueger, J. I., Brod, A., & Groschupf, F. (2013). More is not less: Greater information quantity does not diminish liking. *Journal of Personality and Social Psychology, 105*(6), 909–920. https://doi.org/10.1037/a0033183

Unkelbach, C., Fiedler, K., Bayer, M., Stegmüller, M., & Danner, D. (2008). Why positive information is processed faster: the density hypothesis. *Journal of personality and social psychology*, *95*(1), 36–49. https://doi.org/10.1037/0022-3514.95.1.36

Ziegler, J., & Fiedler, K. (2024). Small sample size and group homogeneity: A crucial ingredient to inter-group bias. Personality and Social Psychology Bulletin, 50(1), 1–17. https://doi.org/10.1177/01461672231223335

Appendix

Project files: https://osf.io/sdwpb/?view_only=4fe29756f5804ed0a7030aed4f0c5159

The preregistration is available under: https://aspredicted.org/jjfp-p39y.pdf

Statutory Declaration

I hereby declare that I have written the present thesis independently and have not used any sources or aids other than those stated. Thoughts taken directly or indirectly from other sources have been identified as such. This thesis has not previously been submitted to any examination authority and has not been published.

Munich, 9th of July, 2025

Konstantinos-Nestor Germanos