

Overconfidence is not the illusion of control: Personality antecedents of chance beliefs among retail investors

Philip Weihrauch^{a,*}, Jonas Manske^b

^a FOM University of Applied Sciences for Economics and Management, Toulouser Allee 53, 40476 Düsseldorf, Germany

^b FOM University of Applied Sciences for Economics and Management, Department of Marketing, 45141 Essen, Germany

* Corresponding author.

E-mail address: philip.weihrauch@gmail.com (P. Weihrauch).

Overconfidence is not the illusion of control: Personality antecedents of chance beliefs among retail investors

Abstract

Behavioral finance routinely treats overconfidence and the illusion of control as facets of a single self-enhancement tendency. We test that assumption directly. Using an original survey of 177 self-directed retail investors, we jointly estimate the associations of impulsivity, overconfidence and sensation seeking with the illusion of control under both asymptotic and bootstrap inference. Impulsivity is the only trait positively related to the illusion of control, and under the double criterion we adopt the association is borderline rather than supported ($\beta = 0.224$, $p = 0.017$, bootstrap 95% confidence interval $[-0.001, 0.642]$). That verdict depends on the estimator: the same path reaches $p < 0.001$ under an ordinal specification and $p = 0.062$ under robust standard errors, while the point estimate stays between 0.22 and 0.30. The three traits jointly explain 6.1% of the variance. A second, exploratory set of results is more informative. Across two independent partitions of the sample, occupation and gender, overconfidence and the illusion of control move in opposite directions: finance-sector employees report 0.37 standard deviations more overconfidence yet 0.33 standard deviations less illusion of control than other investors. A confirmatory comparison of competing measurement models settles what that pattern can only raise: a two-factor specification of the two instruments fits significantly better than a single common factor, and their latent correlation is 0.089 with a 95% interval spanning zero. The gambling-derived instrument appears to capture superstitious rather than investment-specific control beliefs, so instruments imported from gambling research may not measure the construct finance theory specifies.

Keywords: illusion of control; overconfidence; impulsivity; sensation seeking; retail investors

JEL classification: G41; G11; G51; D91

1. Introduction

Retail investors trade too much, diversify too little, chase past returns and hold losers too long (Odean, 1999; Barber and Odean, 2000). A large behavioral literature attributes this pattern to a family of self-enhancement biases, among which overconfidence and the illusion of control are the most frequently invoked. The illusion of control, documented by Langer (1975), is the inappropriately higher expectancy of personal success probability than the objective probability warrants. Overconfidence, in its various forms, is the tendency to overrate one's own ability, precision or relative standing (Moore and Healy, 2008). In applied work their conceptual proximity is well established, and they are commonly treated as members of one family of self-enhancement biases.

The stakes of that conflation have risen. Commission-free execution, gamified mobile trading interfaces and the arrival of highly volatile asset classes have brought a new cohort of retail participants into the market whose behavior is characterized as unusually risk-

tolerant, transaction-intensive and sentiment-driven (Hasso et al., 2019; Barber et al., 2022; Klein, 2022). Eaton et al. (2022) show that the trading of this cohort is associated with lower market quality, higher volatility and stronger herding. The economic magnitudes are not trivial: Kumar (2009) estimates that the risk-adjusted underperformance associated with a lottery-stock preference ranges between 1.7% and 31.5% of annual income depending on the investor's income group.

The policy response to this cohort is already taking shape around dispositional screening. Suitability assessment under the European MiFID II regime obliges intermediaries to establish a client's knowledge, experience, financial situation and risk tolerance before recommending a product, and product governance rules require a defined target market. Both mechanisms presuppose that the relevant investor characteristics can be identified and measured. If regulators and platform designers are to target the dispositions that make investors vulnerable, they need to know which disposition they are targeting — and whether the instruments used to measure it capture what their labels claim. A screening instrument that is labeled “illusion of control” but in fact measures superstition would misallocate protective attention toward investors not necessarily most likely to need it.

This paper asks two questions. First, to what extent are impulsivity, overconfidence and sensation seeking jointly associated with the illusion of control among self-directed retail investors? Second, and consequentially, are overconfidence and the illusion of control empirically separable constructs, or do they behave as facets of one underlying tendency?

The theoretical case for a common structure is straightforward. Impulsivity, in its motor-spontaneous component, favors action before deliberation and thereby impedes the analytical engagement with randomness that would correct control beliefs (Patton et al. 1995). Overconfidence supplies the cognitive premise that one's own judgment is superior, reinforced by self-serving attribution (Miller and Ross, 1975). Heath and Tversky (1991) show that perceived competence in a domain raises willingness to bet within it. Sensation seeking motivates active, high-variance market participation (Grinblatt and Keloharju, 2009), for which the illusion of control can serve as a cognitive justification. Each of the three traits is time-stable rather than transient, which matters because Daniel et al. (2002) argue that behavioral distortions persist precisely when they are anchored in durable psychological characteristics rather than in situational moods. Existing work, however, has examined these predictors one at a time. Stefan and David (2013) note in their meta-analytic review that the set of studied antecedents of the illusion of control has grown substantially. Simultaneous quantitative analyses of several personality traits nonetheless remain rare, as far as we can establish. Subrahmanyam (2008) calls for closer attention to which investor types are subject to which distortions, rather than attributing biases to “the investor” in the aggregate, and Hirshleifer (2015) calls for a move from broad descriptions of distorted trading toward the analysis of individual psychological antecedents.

We cannot claim causal identification, and we do not. The design is a cross-sectional survey in which traits and beliefs are elicited simultaneously. No exogenous variation is available, and the directional arrows in our model rest on theory rather than on a research design. Two design choices compensate. Inference is reported under a double criterion applied uniformly to all three hypothesis paths: a hypothesis counts as supported only when the

asymptotic maximum-likelihood p-value falls below 0.05 *and* the percentile bootstrap 95% confidence interval from 5,000 replications excludes zero. Where the two diverge, we label the result borderline and report both. Measurement is assessed before estimation through exploratory factor analysis with a loading floor, three complementary reliability coefficients, the Fornell-Larcker discriminant criterion and Harman's single-factor test. Every eliminated item is disclosed. We also separate throughout between the three hypotheses stated in Section 2.3, which the design was built to test, and the subgroup analyses of Section 5, which are exploratory and were not specified in advance. What we exploit in the latter is the composition of our sample. Because our recruitment channel produced an unusually large subgroup of finance-sector professionals, we can ask whether two constructs that are said to be interchangeable move in the same direction when that subgroup is compared with the rest. Since this is a test of sign rather than magnitude, it is robust to the measurement limitations we document below.

Our sample comprises 177 self-directed German retail investors surveyed online between 18 April and 2 May 2026. Impulsivity is measured with the short German Barratt Impulsiveness Scale (BIS-15; Meule et al., 2011), sensation seeking with the German adaptation of the Arnett Inventory of Sensation Seeking (AISS-d; Roth and Mayerhofer, 2014), financial overconfidence with the scale of Costa et al. (2025), and the illusion of control with the Drake Beliefs about Chance inventory (DBC; Wood and Clapham, 2005). All constructs are modeled as reflective latent variables and estimated by maximum likelihood in a covariance-based structural equation model.

The joint personality model explains comparatively little on its own. Impulsivity carries a positive standardized coefficient of 0.224 ($p = 0.017$), but its bootstrap interval $[-0.001, 0.642]$ grazes zero, so under our double criterion the hypothesis is borderline rather than supported. Overconfidence ($\beta = 0.091$; $p = 0.315$) and sensation seeking ($\beta = 0.024$; $p = 0.823$) show no association under either criterion. The three traits jointly account for 6.1% of the variance in the illusion of control. An ordinary least squares regression on scale means reproduces this pattern (impulsivity $\beta = 0.258$, $p = 0.001$; overconfidence $\beta = 0.141$, $p = 0.063$; sensation seeking $\beta = 0.009$, $p = 0.906$; adjusted $R^2 = 0.069$).

The cross-sectional evidence is exploratory, and we report it as such. It is nonetheless the part of the study we consider most informative. Respondents employed in banking or financial services report significantly higher overconfidence ($M = 4.275$ versus 3.808; $d = 0.368$; $p = 0.031$) and simultaneously significantly lower illusion of control ($M = 2.676$ versus 3.077; $d = -0.330$; $p = 0.046$) than the remainder of the sample. In standardized terms the overconfidence gap corresponds to roughly 0.37 within-sample standard deviations, or about 12% of the sample mean. The illusion-of-control gap corresponds to roughly 0.33 standard deviations, or about 13.5% of the sample mean, in the opposite direction. Men report higher overconfidence ($d = 0.487$; $p = 0.003$) and lower illusion of control ($d = -0.505$; $p = 0.003$) than women — again in opposite directions, and with somewhat larger magnitudes. Both p-values in the banking comparison sit just below the conventional threshold and would not survive a Bonferroni-Holm correction across the full battery of demographic tests, so we read them as hypothesis-generating rather than confirmatory. But the direction is what matters here, not the magnitude, and it is consistent across two independent partitions of the sample: wherever overconfidence is elevated, DBC-measured illusion of control is depressed. That pattern is difficult to reconcile with a

reading of the two measures as interchangeable indicators of one self-enhancement tendency, and a confirmatory test agrees. Estimated jointly, a two-factor specification of the DBC and Costa items fits significantly better than a single common factor ($\Delta\chi^2(1) = 272.3$, $p < 0.001$), and in the one-factor solution the six DBC items load between -0.02 and 0.24 while the seven Costa items load between 0.63 and 0.93 . A single factor reproduces the overconfidence block and leaves the chance-belief block essentially unexplained.

We perform a battery of checks. Multivariate normality is violated by the Mardia test, which is why bootstrap intervals accompany every asymptotic p-value. No Heywood cases arise. The OLS diagnostics show no heteroskedasticity, no multicollinearity (maximum VIF 1.07) and no influential observations above Cook's threshold. A post hoc power analysis establishes that the design detects medium effects with power 0.99 but small effects ($f^2 = 0.02$) with power only 0.31. And a sensitivity analysis applying a post hoc completion-time filter shows that all path signs are preserved while the overconfidence path becomes significant in both filtered subsamples. We cannot rule out that the null on overconfidence is a power and response-noise artifact rather than a substantive absence.

Our study contributes to the literature in three ways. First, it supplies direct measurement-model evidence that the two instruments most commonly used to capture overconfidence and the illusion of control are not interchangeable in a retail-investor population, answering the call of Klayman et al. (1999) for a differentiated reading of overconfidence effects across competence domains. The finance-professional double finding — more overconfidence, less illusion of control — is difficult to reconcile with any account that treats the two as manifestations of one disposition, and it resolves an apparent tension with Fenton-O'Creevy et al. (2003), who document that individual differences in illusion of control among professional traders are inversely related to rated performance and remuneration. Their operationalization targets competence-based controllability beliefs, which is what the Costa instrument captures and what rises with professional exposure. The DBC targets magical and superstitious chance beliefs, which professional knowledge plausibly dispels.

Second, we identify a measurement problem. Research on control illusions in investment settings has, as far as we have been able to establish, no instrument developed and validated for that setting, and therefore borrows from gambling research, where the construct originated. Finance-specific operationalizations of adjacent constructs do exist — perceived control over market outcomes, perceived trading skill, self-attribution and overconfidence scales — but they are built for those constructs rather than for Langer's, which is what makes the borrowing attractive and what makes it costly. The DBC is a case in point: developed and validated on gambling populations, its item content is dominated by lucky charms, lucky numbers, rituals and gambler's-fallacy beliefs. Our gender and occupation results indicate what this borrowing costs — the instrument appears to track a superstition factor that professional financial knowledge dispels, rather than the competence-based steering illusion that finance theory specifies. Researchers importing gambling-derived instruments into investment settings should expect to measure something adjacent to, but not identical with, the construct their theory calls for.

Third, we contribute simultaneous rather than sequential estimation of three trait antecedents, and we report a null with full diagnostic disclosure. The impulsivity

association survives in both estimation approaches, but it applies to the motor-spontaneous subcomponent that remains after item elimination — four of the five eliminated BIS items were reverse-coded planning items — and it should not be generalized to impulsivity as a whole. Sensation seeking, measured by a domain-general instrument, contributes nothing. The economic content of this exercise is not psychometric refinement for its own sake. Suitability assessment and product-governance rules oblige intermediaries to identify the client characteristics that make an investor vulnerable, and interface design targets those characteristics directly. Which of two adjacent dispositions an instrument actually captures therefore determines who is classified as vulnerable and which intervention is chosen. Joint estimation is what separates the shared from the incremental variance on which such a classification would rest. Both results point toward the same methodological conclusion: progress in this area requires investment-specific adaptations of instruments that were developed for gambling and general personality research.

The remainder of the paper is organized as follows. Section 2 develops the hypotheses. Section 3 describes the data and the research design. Section 4 presents the main results. Section 5 reports the cross-sectional analyses that constitute our central evidence. Section 6 collects the robustness checks. Section 7 concludes.

2. Related literature and hypothesis development

2.1. Conceptual foundations and construct boundaries

The illusion of control belongs to the family of cognitive distortions that behavioral finance treats as systematic departures from the rational-agent benchmark (Kahneman and Tversky, 1979; Barberis and Thaler, 2003). Langer (1975) defines it as an expectancy of personal success probability inappropriately higher than the objective probability warrants. Theoretically it is anchored in Rotter's (1966) locus of control: individuals differ systematically in whether they attribute outcomes to their own actions or to external forces, and the illusion of control is the pathological extension of an internal attribution style into domains where outcomes are stochastically determined.

Three neighboring constructs require demarcation, because they are closely related and frequently co-occur. Overconfidence concerns the overestimation of one's own abilities, and Moore and Healy (2008) decompose it further into overestimation of one's absolute performance, overplacement relative to others, and overprecision in one's beliefs. The gambler's fallacy concerns erroneous beliefs about the serial dependence of random sequences. The illusion of control concerns specifically the belief that one's own action influences an objectively uncontrollable outcome (Presson and Benassi, 1996). All three are conceptually adjacent and are frequently treated as a family, but they are operationally distinguishable — and whether they are empirically distinguishable in retail-investor populations is the question this paper answers.

The distinction is not academic. If overconfidence and the illusion of control are one construct, then a single intervention — for instance, calibration feedback — addresses both, and a single measurement instrument suffices for either. If they are two, then interventions and instruments must be chosen separately, and evidence gathered on one cannot be transported to the other. Theory distinguishes the two, and the instruments developed for

them differ in content. Whether they are empirically separable in a retail-investor population has not, to our knowledge, been tested directly. Section 5 supplies that test.

2.2. The illusion of control in financial contexts

Research on the illusion of control in financial settings can be divided into three strands that have largely developed independently of one another.

Experimental work descends from Langer (1975), who demonstrated that participants demanded a substantially higher resale price for a lottery ticket they had selected themselves than for an assigned ticket (\$8.67 versus \$1.96), although the outcome was random in both conditions. Presson and Benassi (1996) synthesize 53 studies and report the effect as robust ($d = 0.68$). Dunn and Wilson (1990) show that domain expertise can paradoxically amplify rather than attenuate the illusion, because an informational advantage is read as evidence of controllability. Charness and Gneezy (2010), by contrast, find in an incentivized portfolio experiment that isolated manipulation of perceived control does not systematically alter investment behavior and that only a minority will pay for additional control. Control illusion effects are therefore context-dependent and not uniformly detectable across designs.

Field studies using actual trading records document trading patterns that the literature has associated with the illusion — excessive turnover, low diversification, and net-of-cost underperformance (Barber and Odean, 2000, 2001; Statman et al., 2006) — but they do not measure the psychological construct directly. The same patterns are equally consistent with overconfidence, with heterogeneous beliefs and with non-behavioral explanations, so we read them as suggestive of rather than diagnostic for the illusion of control. Kumar (2009) bridges the gap most explicitly: using portfolio and transaction data for 62,387 retail investors at a U.S. discount broker, he identifies lottery-type stocks and shows that the demographic profile of investors overweighting them mirrors that of lottery players, with risk-adjusted underperformance costs of between 1.7% and 31.5% of annual income depending on income group.

Survey-based studies that measure personality traits directly and relate them to investment distortions are, to our knowledge, considerably scarcer than experimental or archival work. A second strand classifies retail investors into personality profiles and recovers, repeatedly, the same three dimensions of differentiation: overconfidence, impulsivity, and risk appetite, with the last most commonly operationalized through sensation-seeking scales (Robba et al., 2024). Kim et al. (2020) find significantly elevated sensation seeking among cryptocurrency investors relative to equity investors. Delfabbro et al. (2021) identify, in a review of the psychology of crypto trading, overestimation of the role of knowledge or skill, fear of missing out, preoccupation and anticipated regret as risk factors that map structurally onto classical gambling patterns.

What none of this work does is estimate the three trait antecedents jointly. Doing so matters for a specific econometric reason: if the traits are correlated — and Table 7 shows that sensation seeking and overconfidence correlate at $r = 0.252$ in our sample — then single-predictor studies attribute shared variance to whichever trait happens to be included. Our estimation therefore recovers the incremental association of each trait conditional on the other two. The relevance of that quantity is not merely statistical: if

regulation or platform design is to act on a disposition, the association that survives conditioning on the adjacent traits is what identifies the disposition an intervention would have to address.

2.3. Hypotheses

Impulsivity in the Barratt tradition comprises attentional, motor and non-planning components (Patton et al., 1995; Stanford et al., 2009). Whiteside and Lynam (2001) derive a four-facet decomposition from the five-factor model — urgency, lack of premeditation, lack of perseverance and sensation seeking, and Moeller et al. (2001) and Verdejo-García et al. (2008) document its clinical relevance as a vulnerability marker across addictive behaviors. Two of these components bear on control beliefs through distinct channels. The motor-spontaneous component describes acting before deliberating: an investor who executes an order on impulse forgoes the reflective step at which the randomness of the price process might have been recognized. The non-planning component describes a weak orientation toward future consequences, which suppresses the retrospective evaluation through which repeated failures of prediction would otherwise correct an inflated sense of controllability. In Kahneman's (2011) dual-process terms, both components describe a decision environment in which System 1 operates without System 2 supervision — and the persistence of control illusions is a characteristic product of such an environment.

The most direct empirical support comes from clinical research on a behaviorally adjacent population. Michalczuk et al. (2011), in a case-control study of pathological gamblers at the UK National Problem Gambling Clinic, report a substantial correlation between impulsivity measures and cognitive distortions including the illusion-of-control component ($r \approx 0.64$). Mosenhauer et al. (2021) document the structural proximity of high-frequency stock trading to problem gambling, which supports transporting the clinical finding into the investment domain. Strömbäck et al. (2017) show, from the opposite direction, that self-control — impulsivity's conceptual inverse — predicts financial behavior and financial well-being.

H1. Impulsivity is positively associated with the illusion of control among retail investors.

Overconfidence supplies the cognitive premise on which beliefs about superior market-steering ability rest. The self-serving attribution bias reinforces it by assigning gains to skill and losses to circumstance (Miller and Ross, 1975), a mechanism that Daniel et al. (1998) formalize in an asset-pricing setting: biased self-attribution causes confidence to grow asymmetrically with public information that confirms private signals, which sustains mispricing. Heath and Tversky (1991) supply the direct bridge to controllability with their competence hypothesis — decision makers who feel competent in a domain accept larger bets within it, preferring to bet on their own judgment over an equiprobable chance event even though the judgmental probabilities are the more ambiguous ones. Statman et al. (2006) and Glaser and Weber (2007) document the resulting turnover, and Barber and Odean (2000) quantify its cost.

Whether this bridge extends all the way to the illusion of control is precisely what is contested. Despite conceptual proximity, the instruments are operationally distinct: Costa et al. (2025) measure finance-specific overconfidence — the conviction that one reads the

market better than other investors — while Wood and Clapham (2005) measure magical thinking under chance. The competence account of Heath and Tversky (1991) nonetheless predicts that they should covary. We assess discriminant validity formally in Section 4.2 and, more consequentially, test the covariation prediction across subgroups in Section 5.

H2. Overconfidence is positively associated with the illusion of control among retail investors.

Sensation seeking, as conceptualized by Zuckerman (1994) and operationalized for novelty and intensity by Arnett (1994), motivates active, high-variance market participation. The illusion of control plausibly functions here as a cognitive protective mechanism that reframes high-risk positions as manageable challenges rather than gambles — a framing without which sustained exposure would be harder to justify to oneself. Grinblatt and Keloharju (2009) show in Finnish register data that sensation-seeking investors trade significantly more frequently. Markiewicz and Weber (2013) find that the gambling subscale of the DOSPERT predicts excessive stock trading, and Wong and Carducci (1991) report the association in everyday financial risk-taking. Kim et al. (2020) find elevated sensation seeking specifically among cryptocurrency investors relative to equity investors, which matters because the crypto and meme-stock segments are where the retail cohort of interest is most concentrated.

H3. Sensation seeking is positively associated with the illusion of control among retail investors.

The research model comprises three direct paths onto the dependent variable, without mediators or moderators. Demographic characteristics are not entered as controls in the main model. They are analyzed separately in Section 5 as group comparisons and correlations, because the literature identifies them as potential predictors of the constructs studied here in their own right (Barber and Odean, 2001; Roth et al., 2005) rather than as nuisance parameters. This choice does not remove the risk of omitted-variable bias. Age, gender, financial-sector employment and investment experience may plausibly be related both to the trait measures and to the reported beliefs, so the conditional associations reported below are descriptive rather than bias-free. Section 5.1 therefore reports a specification that adds age, gender and financial-sector employment as covariates, and the two should be read side by side.

3. Data and research design

3.1. Survey design and data collection

The study uses a cross-sectional correlational design in which all four constructs are elicited through standardized self-report scales. No manipulation or experimental variation of the independent variables is undertaken. This follows established practice in personality-based behavioral finance research (Costa et al., 2025; Robba et al., 2024), and it bounds what the study can claim: associations, not effects.

Data were collected between 18 April and 2 May 2026 as an anonymous online survey administered through the open-source platform LimeSurvey. All construct items use a seven-point Likert format anchored at “strongly disagree” (1) and “strongly agree” (7). The

seven-point format was chosen because it permits sufficiently fine differentiation of response intensity without overtaxing respondents (Joshi et al., 2015; Norman, 2010).

A pretest with ten participants preceded the main survey. Items that produced comprehension or response difficulties were removed to keep completion time reasonable and protect response quality. The pretest was small and its purpose was comprehension testing, not psychometric validation. The removal criteria were qualitative rather than psychometric, and the German adaptations were not re-validated after the reduction. This is a limitation of the design, not a feature of it, and we return to it in Section 7. The most consequential pretest outcome concerned the dependent variable. The instrument originally specified was the four-item illusion-of-control subscale of the Gambling Related Cognitions Scale (GRCS; Raylu and Oei, 2004), chosen because it is well established in clinical gambling research and connects directly to Langer's construct. Pretest participants judged the GRCS items difficult to interpret because of their narrow gambling reference: the dice and lottery framing did not transfer intuitively to investment decisions for respondents without pronounced gambling experience. On this basis the operationalization was switched to the DBC (Wood and Clapham, 2005), which originates in the same research tradition but is formulated more broadly and offers six rather than four items for the control component. We flag at the outset — and return to it in Section 5.4 — that the DBC items nonetheless remain predominantly superstition-oriented (lucky charms, lucky numbers, rituals, gambler's-fallacy beliefs) and therefore capture a gambling-proximate rather than a genuinely investment-specific form of the construct. We state the resulting design problem plainly. The mismatch between the DBC content and the competence-based steering illusion that finance theory specifies was in principle visible before the main survey went into the field. The instrument was retained because no validated investment-specific alternative was available and because the DBC is the broader of the two gambling-derived options, but the consequence is that the dependent variable measures a gambling-proximate construct throughout. The interpretation we develop in Section 5.4 should therefore be read as a diagnosis reached after the fact, not as a hypothesis the design was built to test.

Following Podsakoff et al. (2003), several procedural remedies against common method bias were built into the instrument: independent and dependent variables were separated in the questionnaire layout, anonymity was assured explicitly, and respondents were told that no answers were right or wrong. An attention check item ("Please select 'mostly agree' here") was embedded in the sensation-seeking block.

3.2. Sample construction

The survey targeted retail investors who invest independently in equities or other exchange-traded securities. A screening item on the landing page permitted continuation only upon affirmation. Recruitment proceeded through social networks, investor forums, personal contacts, the survey-exchange platforms SurveyCircle and SurveySwap, and the employee environment of an online broker headquartered in Düsseldorf at which one of the authors is employed. No sample quotas were imposed. The last channel produces a disproportionate representation of banking and financial-sector employees. We treat this in Section 5.3 not merely as a limitation but as the source of our identifying variation for the construct-separation test.

Data cleaning proceeded in three sequential steps. Of 240 raw responses, 24 were excluded for incomplete processing of the six questionnaire pages, leaving 216 complete records. A further 38 were removed for failing the attention check, leaving 178. One record with an implausible age outside the interval [14, 100] was excluded, yielding a final analysis sample of $n = 177$.

We deliberately did not apply a completion-time filter in the main analysis. Of the 177 respondents who passed the three cleaning steps, 103 (58.2%) completed the questionnaire in under five minutes, so the threshold suggested in the survey-methodology literature would have left 74 respondents, or 42% of the analysis sample. The less restrictive median threshold of 4.43 minutes would have left 89. The cost is not only statistical power. The surviving subsample is compositionally different: the share of finance-sector employees rises from 29.4% to 41.9% and median age from 27 to 29, while the gender split is essentially unchanged (66.2% versus 65.0% male). Because the finance-sector subgroup carries the construct-separation evidence of Section 5.3, a filter that concentrates the sample on that subgroup is not a neutral robustness choice. We therefore retain the full attention-checked sample as the main specification and report the filtered estimates in Section 6.4.

3.3. Measures

All four constructs are measured with instruments validated in the literature and administered in German, drawing on published German adaptations where available. Table 1 summarizes the instruments together with original and administered item counts.

Table 1

Measurement instruments.

Construct	Instrument	Original items	Administered	Retained
Impulsivity	BIS-15 (Meule et al., 2011)	15	10	5
Sensation seeking	AISS-d (Roth and Mayerhofer, 2014)	14	14	6
Overconfidence	Costa et al. (2025)	30	7	7
Illusion of control	DBC (Wood and Clapham, 2005)	22	6	6

The table reports the four latent constructs with their measurement instruments. “Original items” is the item count of the published instrument; “administered” is the count used in the main survey after pretest reduction; “retained” is the count after elimination of items loading below $|\lambda| = 0.40$ in a one-factor EFA (Section 4.2). All items use a seven-point Likert format.

The administered item counts deliberately depart from the published originals. The reductions were determined across several pretest rounds in which the removed items repeatedly proved ambiguous in practice. Because this pretest reduction was followed by a second, data-driven elimination in the main sample (Section 4.2), the scales analyzed here are specific to this sample to a degree we cannot quantify without an independent

replication. All constructs are modeled as unidimensional reflective latent variables. Subscale analysis of the multidimensional original instruments was not undertaken, which we flag as a design limitation in Section 7.

Eight items across the four scales are reverse-keyed — four in the AISS-d and four in the BIS-15 — and were recoded as $x_{new} = 8 - x_{old}$ before scale construction. Scale scores are arithmetic means of the constituent items per respondent, consistent with established practice for reflective indicators (Hair et al., 2019).

Two features of the resulting measurement warrant disclosure up front. First, after pretest reduction the seven-item Costa version retains three overplacement items, three overestimation items and only one overprecision item. Under unidimensional aggregation the overprecision component therefore contributes structurally one seventh of the scale score, and an independent reliability assessment of that component is not possible with a single item. The reported overconfidence score predominantly reflects overplacement and overestimation content. Second, as reported in Section 4.2, four of the five BIS items eliminated on factor-loading grounds were reverse-coded planning items, so the retained impulsivity scale primarily captures the motor-spontaneous component.

Gender was recorded with a single self-report item (“Please select your gender”) offering the categories male, female and diverse. Throughout this paper the term therefore refers to self-identified gender rather than to sex assigned at birth. No biological sex variable was collected. Following the SAGER guidelines, the gender-based comparisons reported in Section 5.2 are confined to respondents identifying as male or female, because the single respondent identifying as diverse does not permit a group comparison, and their generalizability is limited accordingly.

3.4. Estimation strategy

Hypotheses are tested in a covariance-based structural equation model implemented in the R package lavaan (Rosseel, 2012). CB-SEM is chosen over regression on scale means because it accounts for indicator measurement error and corrects path coefficients for reliability, which is the methodologically superior treatment for multi-indicator models (Kline, 2016; Hair et al., 2019). The four constructs enter as latent variables with reflective indicators. Scaling uses the marker-variable method, fixing the first indicator loading per construct to unity.

The structural model comprises the three direct paths of Section 2.3 with no mediator or moderator. Estimation is by maximum likelihood. The distributional diagnostics in Section 4.3 show univariate indicator skewness and kurtosis within acceptable bounds (Kline, 2016), which permits ML as an estimator. The multivariate normality assumption is nonetheless violated (Section 6.1), which is precisely why standard errors and 95% confidence intervals for the three hypothesis paths are additionally obtained by non-parametric bootstrap with 5,000 replications using the percentile method (Preacher and Hayes, 2008). The case-to-indicator ratio of 177 to 24 equals 7.4 to 1, within the acceptable range (Bentler and Chou, 1987; Kline, 2016).

Two estimator choices warrant justification. We use covariance-based rather than partial least squares estimation because our constructs are reflective and theory-driven, because

model fit is itself informative for the measurement question this paper raises, and because PLS does not deliver the global fit statistics on which Section 4.4 turns. We retain maximum likelihood with bootstrap intervals rather than a robust estimator such as MLR because the two address different problems: MLR corrects standard errors for non-normality analytically under the assumption that the model is otherwise correctly specified, whereas the non-parametric bootstrap makes no distributional assumption at all and is the more conservative choice given the reliability limitations documented in Section 4.2. Reporting both an asymptotic and a bootstrap criterion, as the rule below does, makes the consequence of that choice visible rather than hidden in a single standard error. We do not claim more for the bootstrap than it delivers: it addresses non-normality, not misspecification, and it does not make seven-point indicators continuous. We therefore re-estimate the model with a robust (MLR) and with an ordinal (WLSMV) estimator and report both in Section 6.1. As it turns out, the choice of estimator matters for what the data support.

Model fit is assessed against conventional benchmarks (Bentler and Bonett, 1980; Browne and Cudeck, 1992; Hu and Bentler, 1999), reported alongside the estimated values in Table 8.

We report each hypothesis path under two inference criteria rather than one. This is a reporting convention adopted for transparency. The study was not pre-registered, and we do not present the rule as a pre-committed decision procedure. A hypothesis is labeled statistically supported only when both criteria coincide — the asymptotic maximum-likelihood p-value falls below 0.05 and the bootstrap 95% confidence interval excludes zero. Where the two diverge, the hypothesis is classified as borderline and both quantities are placed side by side in the discussion. The convention guards against overreading asymptotic significance in the presence of moderate reliability (Preacher and Hayes, 2008). Because it is a convention rather than a test, point estimates, standardized effect sizes and confidence intervals carry the interpretation throughout, and the verdict labels serve only as shorthand for them.

Scale quality is assessed in two stages before estimation. In the first stage a one-factor exploratory factor analysis with promax rotation and minimum-residual estimation is run per construct. Items with $|\lambda| < 0.40$ are removed as insufficient indicators, a threshold established in Hair et al. (2019). Data suitability is verified beforehand through the Kaiser-Meyer-Olkin measure (threshold $KMO \geq 0.50$) and Bartlett's test of sphericity ($p < 0.05$). Reliability is judged on three complementary coefficients: standardized Cronbach's alpha (Cronbach, 1951; threshold ≥ 0.70 per Nunnally, 1978), composite reliability (threshold ≥ 0.60 per Bagozzi and Yi, 1988) and average variance extracted (threshold ≥ 0.50 per Fornell and Larcker, 1981). McDonald's omega is added as a tau-equivalence-free reliability measure (McNeish, 2018). In the second stage the scales are re-formed from the retained items. A floor of four items per reflective unidimensional construct was set as a constant in the estimation script, following Kline (2016). Discriminant validity is assessed by the Fornell-Larcker criterion. Common method bias is examined through Harman's single-factor test. Item selection and model estimation are carried out in the same sample. This is common practice, but it makes both the reported fit and the reported paths optimistic, because the standard errors do not reflect the uncertainty introduced by the selection step.

A split-sample design or validation in an independent sample would be required to remove it, and we flag the point again in Section 7.

All computations were performed in R 4.6.0 using psych (scale assessment), lavaan and semTools (SEM), car and lmtest (assumption diagnostics), pwr (power analysis) and semPlot.

4. Main results

4.1. Sample description

Table 2 reports the demographic composition of the final sample of 177 retail investors. The sample is male-leaning (65.0% male, 34.5% female, 0.6% diverse) and young (median age 27; mean 32.7, SD 12.7, range 18 to 79). Just under three fifths hold an academic degree (57.1%), and 29.4% work in the banking or financial sector. The elevated share of finance-sector employees reflects the recruitment path described in Section 3.2. For the generalizability of the path estimates this is a distortion. For the construct-separation test in Section 5.3 it is what makes the analysis possible at all, because it delivers a subgroup of sufficient size ($n = 52$) to compare against the remainder.

Table 2

Sample composition ($n = 177$).

Characteristic	Category	n	Share
Gender	Male	115	65.0%
	Female	61	34.5%
	Diverse	1	0.6%
Education	Academic degree	101	57.1%
	Upper secondary	41	23.2%
	Completed vocational training	29	16.4%
	Lower secondary	5	2.8%
	Intermediate secondary	1	0.6%
Financial sector	No	125	70.6%
	Yes	52	29.4%
Age (years)	M = 32.7; SD = 12.7; median = 27; range 18–79		

Final analysis sample after the three-step cleaning described in Section 3.2 (240 raw responses → 216 complete → 178 after attention check → 177 after age plausibility). Shares may not sum to 100% due to rounding.

4.2. Scale assessment

In the first estimation step the four construct scales were examined on KMO, Bartlett's test, factor structure and three reliability coefficients. Items loading below $|\lambda| = 0.40$ were then removed and the scales re-formed. Table 3 reports the before-and-after comparison. In total 13 of the 37 administered items were eliminated.

Table 3**Scale assessment before and after item elimination.**

Construct	Items	KMO	α (after)	CR (after)	AVE (after)	ω (after)
Sensation seeking (AISS-d)	14 → 6	0.687	0.644	0.652	0.249	0.633
Impulsivity (BIS-15)	10 → 5	0.714	0.681	0.692	0.322	0.691
Illusion of control (DBC)	6 → 6	0.762	0.794	0.792	0.404	0.799
Overconfidence (Costa)	7 → 7	0.889	0.922	0.919	0.625	0.924

Items with $|\lambda| < 0.40$ in a one-factor EFA with promax rotation and minimum-residual estimation were removed; 13 of 37 administered items were eliminated in total. Thresholds: $KMO \geq 0.50$; $\alpha \geq 0.70$ (Nunnally, 1978); $CR \geq 0.60$ (Bagozzi and Yi, 1988); $AVE \geq 0.50$ (Fornell and Larcker, 1981). ω is McDonald's omega (McNeish, 2018). AVE improved through elimination from 0.13 to 0.25 (AISS-d) and from 0.20 to 0.32 (BIS-15); the pre-elimination values are available only to two decimals.

The overconfidence scale (Costa) clears every pre-specified reliability and validity threshold without difficulty ($KMO = 0.889$; $\alpha = 0.922$; $CR = 0.919$; $AVE = 0.625$; $\omega = 0.924$) and required no item elimination. The illusion-of-control scale (DBC) likewise required no elimination and attains satisfactory reliability ($KMO = 0.762$; $\alpha = 0.794$; $CR = 0.792$; $\omega = 0.799$), but it falls short of the conservative convergent-validity threshold with $AVE = 0.404$. Results concerning the illusion of control are therefore to be read subject to limited indicator homogeneity.

The AISS-d and BIS-15 fare worse. Elimination improved AVE noticeably for both (AISS 0.13 → 0.25; BIS 0.20 → 0.32), but neither scale reaches the established thresholds for Cronbach's alpha (0.644 and 0.681 respectively) or AVE after elimination. The explanation is structural: both original instruments are theoretically multidimensional — AISS distinguishes novelty from intensity, BIS distinguishes attentional, motor and non-planning components — and both were deliberately reduced to unidimensional constructs here. McDonald's omega (0.633 and 0.691) confirms the assessment. As noted, four of the five eliminated BIS items were reverse-coded planning items, which shifts the retained scale toward the motor-spontaneous component and constrains what H1 can be taken to mean.

The Fornell-Larcker comparison is passed across all four constructs: the square roots of AVE (sensation seeking 0.499, impulsivity 0.568, illusion of control 0.636, overconfidence 0.791) each exceed the maximum bivariate correlation with the other constructs (max $r = 0.256$). With AVE below 0.50 for three of the four constructs the criterion carries little diagnostic weight here, so we add two stronger diagnostics, both reported in Table 4. The heterotrait-monotrait ratio of correlations remains far below the conservative 0.85 threshold (Henseler et al., 2015) for every pair of constructs, reaching a maximum of 0.288 for impulsivity and the illusion of control and standing at 0.152 for overconfidence and the illusion of control, the pair on which our argument turns. Estimated as latent variables in a four-factor confirmatory model, those same two constructs correlate at 0.089 (SE 0.083;

95% CI [-0.074, 0.253]; $p = 0.283$), an interval comfortably containing zero. Harman's single-factor test yields a first factor explaining 13.7% of total variance, far below the majority share that would indicate a dominant method factor.

Table 4

Discriminant validity: heterotrait-monotrait ratios and latent correlations.

Construct pair	HTMT	Latent r	95% CI	p
Sensation seeking – impulsivity	0.247	0.193	[-0.005, 0.391]	0.056
Sensation seeking – illusion of control	0.186	0.098	[-0.094, 0.289]	0.318
Sensation seeking – overconfidence	0.282	0.329	[0.163, 0.495]	< 0.001
Impulsivity – illusion of control	0.288	0.225	[0.048, 0.402]	0.013
Impulsivity – overconfidence	0.107	-0.042	[-0.215, 0.131]	0.633
Illusion of control – overconfidence	0.152	0.089	[-0.074, 0.253]	0.283

HTMT is the heterotrait-monotrait ratio of correlations; values below 0.85 indicate discriminant validity under the conservative criterion (Henseler et al., 2015). The ratios reported here are computed in the geometric-mean form HTMT2, which is appropriate for congeneric indicators (Roemer et al., 2021); the arithmetic-mean form of Henseler et al. (2015) yields somewhat higher ratios and the same conclusion. Latent correlations are standardized estimates from a four-factor confirmatory model with correlated factors, estimated by maximum likelihood on $n = 177$; confidence intervals are asymptotic. The pair on which the paper's argument turns is the last row.

The decisive test compares the competing measurement models directly, and Table 5 reports it. Estimating the six DBC items and the seven Costa items in one confirmatory model, a two-factor specification ($\chi^2(64) = 253.67$; CFI = 0.859; TLI = 0.828; RMSEA = 0.129; SRMR = 0.095) fits significantly better than a single common self-enhancement factor ($\chi^2(65) = 525.95$; CFI = 0.657; TLI = 0.589; RMSEA = 0.200; SRMR = 0.177): $\Delta\chi^2(1) = 272.28$, $p < 0.001$, $\Delta\text{CFI} = 0.202$. The one-factor solution shows why. Its standardized loadings run from -0.015 to 0.235 across the six DBC items, only one of which reaches significance, and from 0.625 to 0.930 across the seven Costa items: a single factor reproduces the overconfidence block and leaves the chance-belief block essentially unexplained. We treat this comparison as the primary evidence for construct separation and read the subgroup patterns of Section 5 as corroboration rather than as the test itself. One qualification belongs with it. The two-factor model, though clearly superior, does not itself reach the conventional CFI threshold under maximum likelihood, so the comparison establishes relative and not absolute model adequacy.

Table 5

Competing measurement models for overconfidence and the illusion of control.

Model	χ^2	df	CFI	TLI	RMSEA	SRMR
Two-factor model	253.67	64	0.859	0.828	0.129	0.095
One-factor model	525.95	65	0.657	0.589	0.200	0.177

Confirmatory factor models estimated by maximum likelihood on the six DBC items and the seven Costa items retained after elimination ($n = 177$). The two-factor model specifies the illusion of control and overconfidence as correlated latent factors; the one-factor model specifies a single common self-enhancement factor. Model comparison: $\Delta\chi^2(1) = 272.28, p < 0.001$; $\Delta CFI = 0.202$. In the one-factor solution standardized loadings range from -0.015 to 0.235 across the DBC items and from 0.625 to 0.930 across the Costa items. Neither model reaches the conventional CFI threshold; the comparison establishes relative, not absolute, model adequacy.

4.3. Descriptive statistics and correlations

Table 6 reports location and dispersion measures for the four construct scales after item elimination, together with skewness and kurtosis. All skewness values lie below $|1|$ and all kurtosis values below $|2|$, within the range treated as unremarkable by Kline (2016) and West et al. (1995), which justifies standard ML estimation in the SEM.

Table 6

Descriptive statistics.

Construct	M	SD	Median	Min	Max	Skewness	Kurtosis
Sensation seeking	4.55	1.04	4.67	1.00	6.83	-0.45	0.31
Impulsivity	3.64	1.02	3.60	1.00	6.80	0.12	0.08
Illusion of control	2.96	1.23	3.00	1.00	5.83	0.01	-0.94
Overconfidence	3.95	1.28	4.14	1.00	7.00	-0.32	-0.40

Scale means computed from the items retained after elimination (Table 3). All items are seven-point Likert. All skewness values fall below $|1|$ and all kurtosis values below $|2|$, within the range treated as unremarkable by Kline (2016) and West et al. (1995).

Two features of the descriptive distribution deserve note. The illusion-of-control scale has the lowest mean of the four constructs ($M = 2.96$ on a seven-point scale) and the most negative kurtosis (-0.94), indicating a broad, flat distribution centered well below the scale midpoint. Overconfidence, by contrast, sits close to the midpoint ($M = 3.95$) with the widest dispersion ($SD = 1.28$). The two constructs do not share a distributional profile.

Table 7 reports the bivariate correlation matrix with Spearman coefficients in brackets as a distribution-robust comparison. The central bivariate relationship between impulsivity and the illusion of control is already clearly detectable here ($r = 0.256, p < 0.001$).

Overconfidence relates to the illusion of control only weakly and non-significantly ($r = 0.138, p = 0.068$), and sensation seeking not at all ($r = 0.089$). Notably, impulsivity and overconfidence are essentially orthogonal ($r = -0.020$), while sensation seeking correlates with overconfidence ($r = 0.252, p < 0.001$). Pearson and Spearman coefficients agree closely for the three paths of interest, which argues against distribution-driven artifacts.

Table 7

Correlation matrix.

	Sensation seeking	Impulsivity	Illusion of control	Overconfidence
Sensation seeking	(0.499)	0.173* [0.096]	0.089 [0.039]	0.252*** [0.222]
Impulsivity	0.173* [0.096]	(0.568)	0.256*** [0.276]	-0.020 [-0.025]
Illusion of control	0.089 [0.039]	0.256*** [0.276]	(0.636)	0.138† [0.129]
Overconfidence	0.252*** [0.222]	-0.020 [-0.025]	0.138† [0.129]	(0.791)

*Pearson coefficients with Spearman coefficients in brackets. Significance markers refer to the Pearson coefficient. Diagonal in parentheses reports $\sqrt{AVE}$. *** $p < 0.001$; * $p < 0.05$; † $p < 0.10$. Discriminant validity holds under the Fornell-Larcker criterion: each $\sqrt{AVE}$ exceeds the maximum bivariate correlation of the respective construct (max $r = 0.256$).*

4.4. Structural equation model

The CB-SEM with ML estimation converged admissibly. In the bootstrap with 5,000 replications, 60 replications produced non-convergent or inadmissible solutions (1.2%), below the customary 5% tolerance and consistent with a stable bootstrap. Table 8 reports the fit indices.

The model fit is not adequate by conventional standards. χ^2/df (2.018), RMSEA (0.076, 90% CI [0.066, 0.085]) and SRMR (0.083) fall in the acceptable range, but CFI (0.849) and TLI (0.830) remain below the conventional 0.90 acceptance threshold (Bentler and Bonett, 1980) and well below the value close to 0.95 recommended by Hu and Bentler (1999). Strictly read, the measurement model is not adequately fitted. This weakness traces directly to the reliability problems documented in Section 4.2: items of the AISS-d and BIS-15 contribute heterogeneously to their respective factors across a wide loading range (AISS 0.349 to 0.699; BIS 0.462 to 0.809), which weakens fit in those measurement blocks. Loadings are considerably tighter for the DBC (0.438 to 0.803) and the Costa scale (0.623 to 0.932). The structural parameters are to be interpreted with corresponding restraint. That the model meets the thresholds on the remaining fit dimensions, and that the correlation structure of Section 4.3 delivers consistent point estimates, argues against gross misspecification but does not substitute for a fit demonstration. Two qualifications temper that verdict. First, the statistics reported here treat seven-point Likert indicators as continuous. Re-estimated with an estimator built for ordered categorical indicators, the same measurement model clears the conventional thresholds, as Section 6.1 shows. Second, a substantial part of the residual misfit is local rather than diffuse. The largest modification index in the model, by a wide margin, is the residual covariance between DBC2 and DBC6 (MI = 45.0; residual correlation 0.373), the two items that measure the gambler's fallacy rather than superstition. We return to what that implies in Section 5.4. We nonetheless treat the maximum-likelihood path estimates as descriptive associations, and we let the measurement-model comparison of Section 4.2 rather than the structural parameters carry the argument about construct structure.

Table 8**Model fit.**

Fit index	Value	Acceptable	Good
χ^2 (df = 246)	496.35 (p < 0.001)	–	–
χ^2/df	2.018	≤ 3	≤ 2
CFI	0.849	≥ 0.90	≥ 0.95
TLI	0.830	≥ 0.90	≥ 0.95
RMSEA (90% CI)	0.076 [0.066, 0.085]	≤ 0.08	≤ 0.06
SRMR	0.083	≤ 0.10	≤ 0.08
AIC / BIC	14,883.7 / 15,055.2	–	–
R ² (illusion of control)	0.061	–	–

Covariance-based SEM estimated by maximum likelihood in lavaan (Rosseel, 2012) on n = 177 with 24 indicators across four latent constructs. Acceptable thresholds follow Bentler and Bonett (1980) and Browne and Cudeck (1992); the stricter benchmarks follow Hu and Bentler (1999). CFI and TLI fall below the acceptance threshold; see Section 4.4.

Table 9 reports the standardized path coefficients with asymptotic p-values and percentile bootstrap intervals.

Under the double rule of Section 3.4, H1 is not unambiguously supported: asymptotic inference indicates a significant association (p = 0.017), while the conservative bootstrap reading shows a narrow contact with zero at the lower bound (–0.001). We classify the result as borderline and report both quantities. The wider set of estimators in Section 6.1 confirms that classification: the impulsivity path is significant under maximum likelihood (p = 0.017) and under an ordinal specification (p < 0.001), returns p = 0.022 in a likelihood-ratio test, and falls short under robust standard errors (p = 0.062). The point estimate is stable across all of them, between 0.22 and 0.30, while the inferential verdict is not. H2 and H3 are clearly unsupported under both criteria. The R² of the dependent variable is 0.061 — the three traits jointly explain a small share of individual variance in the illusion of control, which points to further drivers outside the model such as investment experience, cognitive style, market conditions or self-efficacy.

Table 9**Structural paths.**

Hypothesis	β	SE	p	Bootstrap 95% CI	Verdict
H1: Impulsivity	0.224	0.094	0.017	[–0.001, 0.642]	borderline
H2: Overconfidence	0.091	0.090	0.315	[–0.084, 0.197]	not supported
H3: Sensation seeking	0.024	0.107	0.823	[–0.395, 0.515]	not supported

Dependent variable in all three paths is the illusion of control. Standardized coefficients (std.all specification). Bootstrap intervals are percentile-method from 5,000 replications; 60 replications (1.2%) returned non-convergent or inadmissible solutions. Under the double rule

of Section 3.4 a hypothesis counts as supported only when $p < 0.05$ and the bootstrap interval excludes zero. Standardized factor loadings range from 0.349 to 0.699 (AISS-d), 0.462 to 0.809 (BIS-15), 0.438 to 0.803 (DBC) and 0.623 to 0.932 (Costa). Bootstrap intervals refer to the unstandardized solution. Because the two scalings differ by a strictly positive factor in every replication, the assessment of whether an interval covers zero is unaffected.

Two features of these estimates deserve emphasis before we turn to the cross-sectional evidence.

First, the direction of the H1 result is more informative than its magnitude, because it survives specification change. The bivariate Pearson correlation is 0.256 ($p < 0.001$), the OLS coefficient on scale means is 0.258 ($p = 0.001$), and the SEM coefficient is 0.224. Three estimators that differ in how they treat measurement error deliver point estimates within 0.034 of one another. Whatever the inferential verdict, the association itself is not an artifact of the latent-variable specification.

Second, and most consequentially for what follows, the near-zero coefficient on overconfidence is not explained by collinearity with impulsivity. The two predictors are essentially orthogonal in this sample ($r = -0.020$), and the maximum variance inflation factor across the OLS specification is 1.07. The conditional coefficient on overconfidence is therefore close to its unconditional counterpart: overconfidence simply does not track the illusion-of-control measure in this dataset, whether or not the other traits are held constant. If the two were expressions of a single self-enhancement tendency, this would be difficult to accommodate. Section 5 takes up that pattern.

5. Cross-sectional analyses

This section reports the exploratory analyses that motivated the measurement-model comparison of Section 4.2 and that corroborate it in the observed data. The reasoning is the following. If the two measures were interchangeable indicators of one self-enhancement tendency, with comparable loadings and comparable measurement properties across groups, then subgroups elevated on one should not be systematically depressed on the other. Observing the opposite is evidence against interchangeability, though not on its own a test of factor structure: group-specific loadings, an additional group-specific factor, or a failure of measurement invariance could produce the same pattern. That test is supplied by the model comparison in Section 4.2, and what follows shows how the separation it establishes appears in the observed data. The comparisons reported below are post hoc and hypothesis-generating throughout, and we examine them in two approximately independent partitions of the sample.

5.1. Robustness regression

Before turning to the subgroup evidence we verify that the SEM path pattern is not an artifact of the latent-variable specification. A multiple OLS regression of the illusion-of-control scale mean on the three predictor scale means was estimated as a robustness check. All classical assumptions were verified in advance. Variance inflation factors reach a maximum of 1.07, far below the conservative threshold of 5 (O'Brien, 2007), so multicollinearity is immaterial. The Breusch-Pagan test ($BP = 3.42$, $df = 3$, $p = 0.331$) indicates no heteroskedasticity. The Shapiro-Wilk test on residuals ($W = 0.982$, $p = 0.021$)

marginally violates normality, but inference is robust to this violation at $n = 177$. Cook's distance identifies no observation above unity, four observations exceed $4/n = 0.023$ and were not removed.

The regression coefficients confirm the SEM pattern: impulsivity ($\beta = 0.258$, $p = 0.001$) shows a substantial association, while overconfidence ($\beta = 0.141$, $p = 0.063$) and sensation seeking ($\beta = 0.009$, $p = 0.906$) remain non-significant. The adjusted R^2 of 0.069 corresponds closely to the SEM value of 0.061. Adding the demographic characteristics as covariates changes the picture in a way that matters for H2, as Table 10 shows. With age, gender and financial-sector employment in the regression, impulsivity remains significant ($\beta = 0.235$, $p = 0.001$) and overconfidence becomes so ($\beta = 0.216$, $p = 0.004$), while sensation seeking stays null ($\beta = 0.074$, $p = 0.329$). Adjusted R^2 rises from 0.069 to 0.158. The same specification estimated as a structural model reproduces the pattern (impulsivity $\beta = 0.188$, $p = 0.034$; overconfidence $\beta = 0.175$, $p = 0.037$) and raises the explained variance in the illusion of control from 0.061 to 0.205. Being male ($\beta = -0.280$) and working in the financial sector ($\beta = -0.185$) are both associated with lower DBC scores, while age is not ($\beta = -0.022$, $p = 0.770$). The H2 null of the main model is therefore not robust to composition. It holds under maximum likelihood without covariates and not under the specifications that condition on who is in the sample, and we treat it accordingly — as specification-dependent rather than as established.

Table 10

Specification with demographic control variables.

Predictor	SEM β	SEM p	OLS β	OLS p
Impulsivity	0.188	0.034	0.235	0.001
Overconfidence	0.175	0.037	0.216	0.004
Sensation seeking	0.084	0.404	0.074	0.329
Age (standardized)	-0.022	0.770	0.012	0.869
Male	-0.280	< 0.001	-0.564	< 0.001
Financial sector	-0.185	0.011	-0.406	0.009

Dependent variable is the illusion of control. The SEM columns report standardized coefficients from the structural model of Section 3.4 augmented by the three covariates ($\chi^2(315) = 639.26$; $CFI = 0.814$; $TLI = 0.795$; $RMSEA = 0.076$; $SRMR = 0.091$); R^2 of the dependent variable rises from 0.061 without covariates to 0.205 with them. The OLS columns report coefficients from a regression on z-standardized scale means, with the two binary indicators left in 0/1 form; adjusted R^2 rises from 0.069 to 0.158. Age enters standardized in both specifications.

5.2. Gender

Welch t-tests by gender (men versus women; the single diverse-identifying respondent is excluded) reveal substantial differences in three of the four constructs, reported in Table 11. Men report significantly higher sensation seeking ($M = 4.742$ versus 4.202; $t = 3.377$; $p < 0.001$; $d = 0.534$) and significantly higher overconfidence ($M = 4.160$ versus 3.548; $t =$

3.024; $p = 0.003$; $d = 0.487$), while women report significantly higher illusion of control ($M = 3.355$ for women versus 2.751 for men; $t = -3.079$; $p = 0.003$; $d = -0.505$). Impulsivity does not differ ($p = 0.287$). These are comparisons of observed scale means, not of latent means, and the distinction carries weight here. A multi-group test across gender, reported in Table 13 together with the corresponding test for the occupation split, supports metric invariance ($\Delta\chi^2(20) = 23.59$, $p = 0.261$; $\Delta CFI = -0.002$) but rejects scalar invariance ($\Delta\chi^2(20) = 41.67$, $p = 0.003$; $\Delta CFI = -0.012$). Loadings are therefore comparable between men and women while intercepts are not, so part of the mean difference may reflect differential item functioning rather than a difference in the underlying constructs. We report the comparison as a difference in observed scores and do not interpret it at the level of latent means.

Table 11

Gender differences (Welch t-tests).

Construct	M (male, n = 115)	M (female, n = 61)	t	p	Cohen's d
Sensation seeking	4.742	4.202	3.377	< 0.001	0.534
Impulsivity	3.577	3.754	-1.070	0.287	-0.173
Illusion of control	2.751	3.355	-3.079	0.003	-0.505
Overconfidence	4.160	3.548	3.024	0.003	0.487

The single respondent identifying as diverse is excluded. Note the opposing signs for overconfidence and illusion of control.

The first two findings align with the established literature. Barber and Odean (2001) and Cueva et al. (2019) describe men as more overconfident in financial decision contexts, Roth et al. (2005) document higher male sensation seeking in a representative German population sample, and Bannier et al. (2019) report a gender gap in Bitcoin literacy and in both actual and perceived financial literacy.

The third finding is the striking one. Women reporting *higher* illusion of control contradicts the conventional narrative of a more measured female investment psychology — and, more importantly, it runs directly counter to the gender pattern in overconfidence within the very same sample. A consistent explanation follows from the operational content of the DBC. Wood and Clapham (2005) measure magical thinking under chance — lucky charms, lucky numbers, rituals — not finance-related self-assessment. Superstitious content of this kind is reported more frequently by women than by men in the general population, as Ward and King (2020) demonstrate across four studies with more than 2,500 participants combined. The gender result therefore says little about financial self-assessment, but a great deal about what the DBC is measuring.

5.3. Employment in the financial sector

Welch t-tests comparing respondents employed in banking or financial services ($n = 52$) against the remainder ($n = 125$) deliver a coherent double finding, reported in Table 12. Finance-sector employees report significantly higher overconfidence ($M = 4.275$ versus 3.808; $t = 2.194$; $p = 0.031$; $d = 0.368$) and simultaneously significantly lower illusion of

control ($M = 2.676$ versus 3.077 ; $t = -2.019$; $p = 0.046$; $d = -0.330$). Sensation seeking and impulsivity do not differ between the groups ($p = 0.556$ and $p = 0.510$). The same qualification as in Section 5.2 applies, and more sharply. Across this partition, as Table 13 shows, even metric invariance is rejected ($\Delta\chi^2(20) = 32.39$, $p = 0.039$; $\Delta CFI = -0.007$), as is scalar invariance ($\Delta\chi^2(20) = 42.76$, $p = 0.002$; $\Delta CFI = -0.012$). These are differences in observed scale means and nothing beyond that. With 52 respondents in the smaller group against 78 free parameters per group, the invariance test is itself imprecise, so we read it as a caution rather than as a verdict. Two further limitations are specific to this partition. Financial-sector employment is recorded as a single binary item, without function, seniority or front-office versus back-office detail, so the group is heterogeneous in ways we cannot observe. And it was reached largely through one recruitment channel, so self-selection into the sector and into our sample cannot be separated from any effect of professional exposure. No causal reading of the difference is available.

Table 12

Differences by financial-sector employment (Welch t-tests).

Construct	M (yes, n = 52)	M (no, n = 125)	t	p	Cohen's d
Sensation seeking	4.625	4.513	0.591	0.556	0.107
Impulsivity	3.723	3.608	0.662	0.510	0.113
Illusion of control	2.676	3.077	-2.019	0.046	-0.330
Overconfidence	4.275	3.808	2.194	0.031	0.368

Post hoc subgroup analysis within a series of demographic tests. Neither p-value remains uniformly significant under Bonferroni-Holm correction. The opposing signs for overconfidence and illusion of control replicate the pattern in Table 11 across an orthogonal partition.

Table 13

Measurement invariance across gender and financial-sector employment.

Model	χ^2	df	CFI	$\Delta\chi^2$ (Δdf)	p	ΔCFI
Gender: configural	859.55	492	0.790	—	—	—
Gender: metric	883.14	512	0.788	23.59 (20)	0.261	-0.002
Gender: scalar	924.81	532	0.776	41.67 (20)	0.003	-0.012
Financial sector: configural	923.19	492	0.770	—	—	—
Financial sector: metric	955.58	512	0.763	32.39 (20)	0.039	-0.007
Financial sector: scalar	998.34	532	0.751	42.76 (20)	0.002	-0.012

Multi-group confirmatory factor models with mean structure, estimated by maximum likelihood. Gender: 115 men against 61 women, the single diverse-identifying respondent excluded. Financial sector: 52 respondents employed in banking or financial services against 125 others. Each Δ compares a model with the one directly above it within the same partition. Under the χ^2 criterion metric invariance holds across gender and is rejected across occupation, and scalar invariance is rejected in both partitions; under the $\Delta CFI \leq 0.01$ criterion (Cheung

and Rensvold, 2002) only the scalar steps are flagged. With 78 free parameters per group the tests are imprecise in the smaller groups.

Both p-values sit just below the conventional 0.05 threshold, and this is a post hoc subgroup analysis within a series of demographic tests. Under correction for multiple testing (Bonferroni-Holm) neither would remain uniformly significant. The finding is hypothesis-generating, not confirmatory.

Two further checks bear on its interpretation. First, the two partitions are close to independent. Of the men in the sample 31.3% work in the financial sector, against 24.6% of the women, and the association is not significant ($\chi^2(1) = 0.577$, $p = 0.447$, $\phi = 0.057$). The occupation split therefore does not simply reproduce the gender split. Second, when both indicators are entered jointly in a regression on the scale means, the occupation difference in overconfidence remains significant ($b = 0.440$, $p = 0.035$), while the difference in the illusion of control falls to $b = -0.361$ ($p = 0.070$). Adding age leaves both coefficients unchanged. The overconfidence leg of the double finding is thus robust to mutual control, the illusion-of-control leg is not, and we treat the latter as suggestive.

We do not claim that this result outweighs the pre-specified path estimates. It is a post hoc subgroup comparison, it does not survive correction for multiple testing, and its illusion-of-control leg falls to $p = 0.070$ under mutual control. It belongs to hypothesis generation, not to confirmation. What makes it worth reporting is that it concerns the sign of a difference rather than its magnitude, and that the sign opposition recurs in the second of two approximately independent partitions, the first being gender in Section 5.2. Read together, the two partitions make it difficult to treat the two measures as interchangeable. They do not establish that the underlying constructs are distinct, and we do not read them as doing so.

The consistent reading is that professional financial exposure goes together with a stronger competence-based component — market acumen, belief in superior investment decisions, which is what the Costa instrument captures — and a weaker magical-superstitious component of the kind the DBC captures. This connects directly to the call of Klayman et al. (1999) for a differentiated reading of overconfidence effects across competence domains.

5.4. Reconciling the finding with the trading-on-illusions literature

At first sight our finance-professional result appears to contradict Fenton-O'Creevy et al. (2003), who study 107 professional securities traders and find that traders with higher perceived control over objectively uncontrollable price movements performed worse on performance, risk and remuneration assessments. If professional exposure lowered the illusion of control, their result should not obtain.

The apparent contradiction dissolves once the two operationalizations are separated. Fenton-O'Creevy et al. measure competence-based controllability beliefs about market processes — conceptually the Costa content, which in our sample *rises* rather than falls with professional exposure. The DBC measures superstitious chance beliefs, which professional financial knowledge plausibly dispels. Our double finding is therefore not a counterexample to the trading-on-illusions literature but an empirical demonstration that its construct and

the DBC construct are separable and move in opposite directions across the same population split.

The implication is a caution about instrument transfer. The DBC was developed and validated on gambling populations, and our results indicate that it retains that character when it is carried into an investment setting. A study adopting it to test an economic hypothesis about trading behavior would therefore be measuring a superstition factor under a label that theory assigns to a steering illusion, and a null result in such a design would be as consistent with the instrument missing the construct as with the construct missing the outcome. The internal structure of the DBC in our data sharpens the point. Its largest source of residual misfit is the covariance between the two gambler's-fallacy items — DBC2, on a coin that has landed heads ten times in a row, and DBC6, on the chance of winning after repeated losses at dice — with a modification index of 45.0 and a residual correlation of 0.373 once the common factor is accounted for. The lucky-charm, lucky-number and ritual items form the other pole. None of the six items refers to an investment, a financial market or a trading decision, and none describes influencing a random outcome through one's own action, which is the expectancy Langer (1975) defines as the illusion of control. As administered, then, the scale is not the unidimensional construct our model treats it as: it combines a superstition facet with a negative-recency facet, and neither of them is the competence-based steering illusion that finance theory specifies. We report this as a diagnosis and do not free the residual covariance, since a data-driven modification in the same sample would compound the selection problem disclosed in Section 3.4.

5.5. Age and education

Table 14 reports Pearson and Spearman correlations with age. Only sensation seeking shows a weak but consistent negative association (Spearman $\rho = -0.190$, $p = 0.011$). This is consistent with developmental evidence that sensation-seeking tendencies decline with age (Roth et al., 2005; Steinberg, 2008).

Table 14

Correlations with age.

Construct	Pearson r	p	Spearman ρ	p
Sensation seeking	-0.136	0.072	-0.190	0.011
Impulsivity	-0.092	0.225	-0.047	0.533
Illusion of control	-0.005	0.946	0.028	0.715
Overconfidence	0.072	0.338	0.084	0.268

One-way ANOVA across education categories (Table 15) yields a significant main effect only for sensation seeking ($F(4, 172) = 2.852$; $p = 0.025$; $\eta^2 = 0.062$). Tukey HSD post hoc comparisons, restricted to categories with $n \geq 2$, indicate that respondents with lower secondary education report lower sensation seeking than the remaining education groups ($\Delta = -1.51$ to -1.57 ; $p = 0.008$ to 0.011). Given the subgroup size of $n = 5$ in the lower-secondary category, this comparison is methodologically fragile and we do not build on it.

Table 15

One-way ANOVA across education categories.

Construct	F (df = 4, 172)	p	η^2
Sensation seeking	2.852	0.025	0.062
Impulsivity	0.420	0.794	0.010
Illusion of control	0.973	0.424	0.022
Overconfidence	0.508	0.730	0.012

Tukey HSD post hoc comparisons restricted to categories with $n \geq 2$ indicate lower sensation seeking among respondents with lower secondary education ($\Delta = -1.51$ to -1.57 ; $p = 0.008$ to 0.011). Given the subgroup size of $n = 5$ in the lower-secondary category, this comparison is fragile.

The age result deserves one further comment, because it is a null with content. If the DBC were capturing an investment-specific steering illusion, one would expect it to decline with market experience, which correlates with age. List (2003) shows that market experience attenuates the endowment effect. It does not decline: the illusion-of-control measure is the most age-invariant of the four constructs (Pearson $r = -0.005$; Spearman $\rho = 0.028$). Superstitious beliefs, by contrast, have no particular reason to erode with trading experience. The age pattern is not inconsistent with the gender and occupation results, but given the limited power for small effects documented in Section 6.3 it carries little independent weight.

5.6. Implications for measurement, disclosure and platform design

For empirical research, the immediate implication is that the illusion of control cannot be treated as observationally equivalent to overconfidence in retail-investor populations, and that instrument choice determines which of the two a study is actually measuring. Studies that report an overconfidence-turnover relationship (Statman et al., 2006; Glaser and Weber, 2007, for the better-than-average facet) and studies that report a control-illusion-performance relationship (Fenton-O'Creevy et al., 2003) are not replications of one another and should not be cited as though they were. Where a study using a gambling-derived instrument reports a null on an economic outcome, the null is at least as likely to indicate that the instrument missed the construct as that the construct missed the outcome.

For investor protection, the implication cuts against a natural intuition. Financial literacy and professional exposure are the standard remedies proposed for behavioral vulnerability. In our sample, exposure is associated with the two components in opposite directions, lower on the superstitious component and higher on the competence-based one. Self-selection into the financial sector is an equally plausible reading of that pattern and cannot be separated from it in a cross-section. We draw no policy conclusion from this. A convenience sample of 177 investors, a cross-sectional design without trading outcomes, a measurement model that does not meet conventional fit standards and an exploratory subgroup comparison cannot support a recommendation about how suitability assessment, disclosure or training programs should be designed. What the pattern supports is a research implication: whether behavioral training and disclosure aimed at experienced investors should address the belief in one's own superior judgment rather than the belief in luck, and whether professional exposure moves the two components in opposite directions

at all, is a question that larger and outcome-linked samples would have to settle. This aligns with Dunn and Wilson (1990), who show that domain expertise can amplify rather than attenuate control illusions when the illusion is competence-based.

For platform design and its supervision, the impulsivity result is the operative one, subject to the inferential caveats above. The retained impulsivity scale captures the motor-spontaneous component, and it is precisely that component which contemporary trading interfaces are built to engage: real-time push notifications on price movements, audiovisual feedback on completed trades, level systems and badges tied to trading activity, social leaderboards and time-limited trading challenges. Barber et al. (2022) document empirically that attention-induced trading of the kind such features generate raises trading activity among Robinhood users, and Eaton et al. (2022) show that the associated order flow coincides with reduced market quality. Whether supervisory practice under the MiFID II conduct-of-business rules should extend to gamified interface elements is a question our data raises but cannot settle. Our data cannot establish how dispositional screening should be designed. At most it suggests that impulsivity is the more promising candidate for further study in that context than chance beliefs are.

6. Robustness checks

6.1. SEM assumptions

The Mardia test indicates significant multivariate skewness across the 24 indicators ($\chi^2(2600) = 3638.71$; $p < 0.001$) and significant multivariate kurtosis ($z = 10.95$; $p < 0.001$). The multivariate normality assumption is therefore violated and the asymptotic ML standard errors stand under reservation. This is precisely why the main analysis was supplemented with bootstrap confidence intervals, and it is why our double reporting rule is not a cosmetic device: for H1 the two criteria genuinely diverge, and the divergence is informative rather than incidental. Two re-estimations bear directly on the assumption. Under MLR, which corrects standard errors analytically for non-normality, the scaled fit improves slightly (scaling factor 1.077; scaled $\chi^2 = 460.82$; RMSEA = 0.070) and the standard errors widen: impulsivity falls to $p = 0.062$, overconfidence to $p = 0.477$ and sensation seeking to $p = 0.862$. Under WLSMV, which treats the seven-point indicators as ordered categorical and works from polychoric correlations, the measurement model improves substantially (scaled $\chi^2(246) = 495.26$; CFI = 0.956; TLI = 0.951; RMSEA = 0.076; SRMR = 0.088) and both trait paths reach significance (impulsivity $\beta = 0.298$, $p < 0.001$; overconfidence $\beta = 0.168$, $p = 0.041$; sensation seeking $\beta = -0.040$, $p = 0.703$). Neither result should be over-read. Fit indices are not comparable across estimators, because CFI and TLI are referenced to different baseline models. RMSEA is unchanged at 0.076 and SRMR is marginally worse, so what the ordinal specification shows is that part of the misfit under maximum likelihood is an artifact of treating ordinal indicators as continuous, not that the model fits well in absolute terms. The MLR standard errors, for their part, are estimated from 300 distinct moments on 177 observations and are themselves imprecise. Read together, the three estimators agree on the sign and the rough magnitude of the impulsivity association and disagree on whether it clears conventional significance, which is precisely what our borderline classification was meant to convey.

No Heywood cases arise in any specification. All 24 item residual variances and all four factor variances remain positive. The smallest item residual is 0.311, the smallest factor variance 0.331, and the smallest eigenvalue of the covariance matrix 0.227. Mahalanobis distance diagnostics at the item level identify 21 records above the conservative $\chi^2(0.95)$ threshold of 36.42 and one extreme record (response ID 167; $D^2 = 60.36$) exceeding the stricter $\chi^2(0.999)$ threshold of 51.18. No deletion was performed, because a plausibility inspection of the self-report Likert data revealed no substantive irregularities. In a replication on a larger sample the question of a leverage-based sensitivity analysis would arise again.

6.2. Common method bias

Common method bias was addressed both procedurally and statistically. Following Podsakoff et al. (2003), the questionnaire separated the independent from the dependent constructs in its layout, assured respondents of anonymity, stated explicitly that no answers were right or wrong, and embedded an attention check. Statistically, Harman's single-factor test yields a first unrotated factor explaining 13.7% of total variance, far below the majority share that would indicate a dominant method factor. Podsakoff et al. (2003) describe the test as an insensitive diagnostic rather than a control. It cannot demonstrate the absence of common method bias, and we do not read it as doing so. All four constructs were elicited from the same respondents, in one session, by the same self-report method, which leaves common method variance as a material identification concern rather than a residual one. No marker variable was collected. A method-factor specification, or a marker-variable design in a replication, is what would be required to bound the problem.

6.3. Statistical power

Power computed after the fact from conventional effect sizes adds little to what the reported intervals already convey, so we report the sensitivity of the design instead: the smallest effect it could have detected at conventional power. At $n = 177$, $\alpha = 0.05$ and three predictors, the minimum detectable f^2 at power 0.80 is approximately 0.063, corresponding to an R^2 of roughly 0.06 (pwr.f2.test; Cohen, 1992). For reference, the power attained at the conventional benchmarks is 0.31 for small effects ($f^2 = 0.02$), 0.99 for medium effects ($f^2 = 0.15$) and above 0.99 for large effects ($f^2 = 0.35$).

The design is thus well dimensioned for medium and large effects but clearly underpowered for small individual effects. This bears directly on the interpretation of H2 and H3. Neither should be read as falsified, both should be read as not securable with this sample and this operationalization. The bootstrap intervals in Table 9 are the relevant quantities: each includes zero, and each is wide enough to remain consistent with an association of a size this design could not have detected.

6.4. Sensitivity to a completion-time filter

The main analysis was run on the full attention-check-cleaned sample. The survey methodology literature nonetheless recommends flagging conspicuously fast responses as potential insufficient effort responding, since even attention-checked respondents may answer systematically distorted at sub-threshold reading and processing times (Huang et al., 2012; Meade and Craig, 2012; Leiner, 2019). Huang et al. (2012) adopt an ad hoc floor of

roughly two seconds minimum reading time per item. Leiner (2019) develops the Relative Speed Index, whose empirically derived cut-off lies near 1.3, with a more lenient pragmatic recommendation of 2.0.

Completion time was computed from the start and submission timestamps of the cleaned raw data. The distribution is right-skewed (median 4.43 minutes, 95th percentile 10.1). Because the timestamps record elapsed time between starting and submitting rather than active processing time, the upper tail reflects sessions left open rather than unusually thorough responding. The two records above three hours lie far above any threshold considered here and are retained under every filter. Of the 177 respondents, 103 (58.2%) fall below five minutes and 74 (41.8%) at or above. We re-estimate the main analysis on two additional subsamples: an absolute threshold of at least five minutes ($n = 74$, corresponding to roughly 8 seconds per item across the 37 items administered, well above the Huang et al. floor), and a less restrictive median threshold of 4.43 minutes ($n = 89$). Tables 16, 17 and 18 report the comparison.

Table 16

Sensitivity analysis: standardized SEM path coefficients.

Path	Main ($n = 177$)	≥ 5 min ($n = 74$)	$\geq$ median ($n = 89$)
Impulsivity	0.224 ($p = 0.017$)	0.520 ($p < 0.001$)	0.465 ($p < 0.001$)
Overconfidence	0.091 ($p = 0.315$)	0.442 ($p < 0.001$)	0.438 ($p = 0.001$)
Sensation seeking	0.024 ($p = 0.823$)	-0.225 ($p = 0.134$)	-0.167 ($p = 0.290$)

Dependent variable is the illusion of control. Completion-time filters applied post hoc. Median completion time 4.43 minutes; 103 respondents (58.2%) fall below five minutes. The filter was not applied in the main analysis (Section 3.2). Main-column values are those reported in Table 9. All p -values are those of the standardized solution, obtained by the delta method, and are therefore consistent with the standardized coefficients they accompany; the corresponding unstandardized p -values differ materially in the filtered subsamples, as discussed in Section 6.4.

Table 17

Sensitivity analysis: fit indices and R^2 .

Subsample	n	CFI	TLI	RMSEA	SRMR	R^2
Main analysis	177	0.849	0.830	0.076	0.083	0.061
Completion time ≥ 5 min	74	0.802	0.778	0.084	0.094	0.345
Completion time $\geq$ median (4.43 min)	89	0.808	0.784	0.080	0.092	0.264

The rise in R^2 across subsamples is partly attributable to upward small-sample bias in unadjusted R^2 , as the more moderate increase in OLS adjusted R^2 in Table 18 shows.

Table 18

Sensitivity analysis: OLS robustness regression (z-standardized scales).

Predictor	Main (n = 177)	≥ 5 min (n = 74)	≥ median (n = 89)
Impulsivity	0.258 (p = 0.001)	0.336 (p = 0.004)	0.317 (p = 0.003)
Overconfidence	0.141 (p = 0.063)	0.252 (p = 0.035)	0.268 (p = 0.014)
Sensation seeking	0.009 (p = 0.906)	-0.067 (p = 0.563)	-0.059 (p = 0.576)
Adjusted R ²	0.069	0.103	0.102

Dependent variable is the illusion-of-control scale mean. Diagnostics for the main model: maximum VIF 1.07; Breusch-Pagan BP = 3.42, df = 3, p = 0.331; Shapiro-Wilk W = 0.982, p = 0.021; no observation above Cook's distance of 1.

The sensitivity analysis preserves the direction of every path. The H1 coefficient remains highly significant in all three OLS models ($\beta = 0.258$ to 0.336 ; $p \leq 0.004$) and carries a substantially larger point estimate in both filtered SEM subsamples ($\beta = 0.520$ and 0.465 ; $p < 0.001$ in each). One point of reporting hygiene belongs here, because it governs how Table 16 is read. Standardized and unstandardized coefficients do not share a p-value: the standardized solution is tested by the delta method, the unstandardized one by the ratio of the raw coefficient to its standard error. At $n = 74$ the two diverge sharply for impulsivity ($p < 0.001$ standardized against $p = 0.104$ unstandardized), because the latent variance of that construct is weakly determined in the reduced sample. We report the standardized quantity throughout, consistently with the coefficients it accompanies, and note that the likelihood-ratio test, which does not depend on this parameterization, agrees with it ($\chi^2(1) = 12.73$, $p < 0.001$ in the five-minute subsample).

The H2 path is where the analysis becomes genuinely informative. In the main analysis the overconfidence path fell short of significance in both the SEM ($\beta = 0.091$; $p = 0.315$; Table 9) and the OLS ($\beta = 0.141$; $p = 0.063$). In both filtered subsamples it becomes significant in the SEM ($\beta = 0.442$ and 0.438 ; $p < 0.001$ and $p = 0.001$) and in the OLS robustness regression ($p = 0.014$ to 0.035). Under our double rule this is emphatically not confirmatory support for H2, because the filter was chosen post hoc. Nor do we read it as evidence for H2. A coefficient that moves from 0.091 to 0.442 when 58% of the sample is removed is in the first instance a warning about specification sensitivity and about response quality in the full sample, and only in the second instance a substantive signal. With 103 of 177 respondents below five minutes, we examined response quality directly rather than relying on the time cut-off alone. Intra-individual response variability across the 24 retained indicators averages 1.62 (SD 0.48) and falls below 0.5 for five respondents. The longest run of identical consecutive answers has a median of 4, reaches the full 24 in at least one case, and is eight or longer for seventeen respondents (9.6%). Respondents below the five-minute threshold show slightly lower response variability (1.57 against 1.69) and longer identical runs (5.29 against 4.32), but the difference is not significant (Welch $t = -1.70$, $p = 0.090$). The completion-time filter therefore does not separate the sample cleanly by response quality, which is a reason to treat the filtered estimates as a bound on the confidence the H2 null can carry rather than as a cleaner measurement. The exercise does establish one thing outright: at least one respondent gave the identical answer to all 24 indicators and still passed the attention check, which is a limitation of single-item attention screening that replication designs should address.

The H3 path remains clearly non-significant in all three OLS models ($p = 0.563$ to 0.906). In the filtered SEM subsamples the point estimate turns negative ($\beta = -0.167$ and -0.225) without approaching conventional significance in any specification.

Fit indices for the subsample SEMs (CFI 0.802 and 0.808; TLI 0.778 and 0.784) remain below the 0.90 threshold and in fact fall slightly below the main analysis, an expected consequence of reduced sample size at unchanged model complexity. The apparently sharp rise in the R^2 of the illusion of control (0.061 in the main analysis versus 0.264 to 0.345 in the subsamples) should not be read as a substantive gain in explanatory power. Unadjusted R^2 is upward biased in small samples, and the subsample estimates are unstable at $n = 74$, where the measurement model fits least well. The far more moderate increase in the OLS adjusted R^2 (0.069 to 0.102 and 0.103) is the more informative comparison. Substantively, the explained variance remains at a level customary for personality-based models in behavioral research (Costa et al., 2025; Robba et al., 2024).

The five-minute filter is post hoc and reduces the sample by 58%, which materially weakens statistical power. The less restrictive median filter ($n = 89$) yields qualitatively identical results and thereby supports the robustness claim additionally. The analysis should be understood as a supplementary check in the sense of Meade and Craig (2012) and Leiner (2019), not as a confirmatory re-analysis.

7. Conclusion

Overconfidence and the illusion of control move in opposite directions across two independent partitions of our sample. Investors employed in the financial sector report more overconfidence and less illusion of control than other investors, and the same opposition appears across gender. That pattern is corroborative rather than decisive on its own: both comparisons are post hoc, neither survives correction for multiple testing, and measurement invariance holds only partially across gender and not at all across occupation. What settles the question is the measurement model. A two-factor specification of the two instruments fits significantly better than a single common factor ($\Delta\chi^2(1) = 272.3$, $p < 0.001$), their latent correlation is 0.089 with an interval spanning zero, and their heterotrait-monotrait ratio is 0.15. On that evidence the two are not facets of a single self-enhancement tendency.

The joint personality model that motivated the study explains comparatively little by itself. Impulsivity carries a positive coefficient that survives every estimation approach we try but grazes zero under conservative bootstrap inference and under robust standard errors, while reaching $p < 0.001$ under an ordinal specification, overconfidence and sensation seeking show no detectable association, and the three traits together account for 6.1% of the variance in the dependent measure. Where the impulsivity association does hold, it applies to the motor-spontaneous subcomponent that remains after item elimination, not to impulsivity as construed by the full Barratt taxonomy. Read against the pattern above, that is less a failure of the model than an indication that the dependent measure is not capturing the construct the model was built for. The measurement-model comparison rules out the reading that both instruments tap one latent tendency. What it cannot settle is whether the separation is one of constructs or one of item content, and our reading is the latter: the Drake Beliefs about Chance inventory captures superstitious chance beliefs, which

professional financial knowledge dispels and which women report more frequently in the general population, rather than the competence-based steering illusion that finance theory has in mind. On that reading our finding reconciles with Fenton-O'Creevy et al. (2003) rather than contradicting it, and it would imply that studies deploying gambling-derived instruments to test economic hypotheses about trading may be reporting nulls that belong to the instrument.

Four features of the design bound what these results can support. The design is cross-sectional and self-reported, so directional claims rest on theory alone. The sample is a convenience sample of 177 investors, male-leaning, young, and heavily weighted toward finance-sector employment. It is not representative of German retail investors, though that same weighting is what enables the construct-separation test. Two of the four scales fall short of conventional reliability thresholds, the measurement model does not meet the CFI and TLI benchmarks under maximum likelihood, and the design is underpowered for small effects. Item selection and model estimation were carried out in the same sample, so the reported fit and paths are optimistic to an unquantified degree. Measurement invariance is only partial across gender and absent across occupation, so the subgroup results are comparisons of observed scale means rather than of latent means. We did not use trading data, so the illusion of control is never linked to realized behavior or returns. Gender is measured by a single self-report item, and the gender comparisons cover respondents identifying as male or female only, which bounds their generalizability.

Four implications follow for future work. Investment-specific adaptation of the illusion-of-control instrument is the precondition for progress in this area. An adaptation along the lines of Fenton-O'Creevy et al. (2003) would allow the trading-specific steering illusion to be measured directly. Attention-sensitive survey administration — multiple distributed attention items, per-page minimum processing times, item-level response latencies — would remove a noise source that our sensitivity analysis shows to be material. Linkage of trait measurement to brokerage records would supply the outcome variable this literature needs to establish economic relevance rather than merely psychological association. And a replication large enough to establish measurement invariance across gender and occupation would allow the subgroup patterns reported here to be read at the level of latent means, which our sample does not permit.

Ethics approval and informed consent

This study did not require approval by an ethics committee. At FOM University of Applied Sciences, where the research was carried out, the Ethics Committee reviews projects involving human participants against the institution's checklist of features of low-risk research, and projects that meet those features may be conducted without a formal vote. This study meets them. It collected only the information needed to answer the research questions, involved no burden beyond everyday experience, no physiological or diagnostic data, no invasive procedures and no deception, and surveyed adults only. No incentive was offered that might have induced participation. Respondents recruited through the survey-exchange platforms SurveyCircle and SurveySwap received a participation code for reciprocal participation in their own surveys and no monetary compensation. Responses were collected anonymously and treated confidentially, so that no respondent was exposed

to legal, financial, professional or reputational risk, and all data protection provisions were observed.

Participation was voluntary and anonymous, and respondents could discontinue at any point. The introductory page of the questionnaire identified the study as part of a bachelor's thesis at FOM University of Applied Sciences on the decision styles and personality traits of retail investors, stated the expected duration of eight to ten minutes and the anonymity of the responses, and informed respondents that by clicking "next" they consented to the anonymous processing of their answers. No names, contact details or other personal identifiers were collected, and the data contain no information that would allow individual respondents to be identified. The study was conducted in accordance with the ethical principles of the Declaration of Helsinki.

Data availability

The anonymized survey data and the R code used for all analyses reported in this article can be made available upon reasonable request.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used Claude Opus 5 (Anthropic) in order to support the statistical analysis and to improve the wording and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

- Arnett, J., 1994. Sensation seeking: a new conceptualization and a new scale. *Pers. Individ. Differ.* 16, 289–296.
- Bagozzi, R.P., Yi, Y., 1988. On the evaluation of structural equation models. *J. Acad. Mark. Sci.* 16, 74–94.
- Bannier, C.E., Meyll, T., Röder, F., Walter, A., 2019. The gender gap in "Bitcoin literacy". *J. Behav. Exp. Finance* 22, 129–134.
- Barber, B.M., Odean, T., 2000. Trading is hazardous to your wealth: the common stock investment performance of individual investors. *J. Finance* 55, 773–806.

Barber, B.M., Odean, T., 2001. Boys will be boys: gender, overconfidence, and common stock investment. *Q. J. Econ.* 116, 261–292.

Barber, B.M., Huang, X., Odean, T., Schwarz, C., 2022. Attention-induced trading and returns: evidence from Robinhood users. *J. Finance* 77, 3141–3190.

Barberis, N., Thaler, R., 2003. A survey of behavioral finance. In: *Handbook of the Economics of Finance*, vol. 1, pp. 1053–1128.

Bentler, P.M., Bonett, D.G., 1980. Significance tests and goodness of fit in the analysis of covariance structures. *Psychol. Bull.* 88, 588–606.

Bentler, P.M., Chou, C.-P., 1987. Practical issues in structural modeling. *Sociol. Methods Res.* 16, 78–117.

Browne, M.W., Cudeck, R., 1992. Alternative ways of assessing model fit. *Sociol. Methods Res.* 21, 230–258.

Charness, G., Gneezy, U., 2010. Portfolio choice and risk attitudes: an experiment. *Econ. Inq.* 48, 133–146.

Cheung, G.W., Rensvold, R.B., 2002. Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equ. Model.* 9, 233–255.

Cohen, J., 1992. A power primer. *Psychol. Bull.* 112, 155–159.

Costa, D.F., Soares, C.C., Moreira, B.C.M., Tonelli, A.O., 2025. Construction and validation of an overconfidence scale in investment decisions. *Future Bus. J.* 11, article 6.

Cronbach, L.J., 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16, 297–334.

Cueva, C., Iturbe-Ormaetxe, I., Ponti, G., Tomás, J., 2019. Boys will still be boys: gender differences in trading activity are not due to differences in (over)confidence. *J. Econ. Behav. Organ.* 160, 100–120.

Daniel, K., Hirshleifer, D., Teoh, S.H., 2002. Investor psychology in capital markets: evidence and policy implications. *J. Monet. Econ.* 49, 139–209.

Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and overreactions. *J. Finance* 53, 1839–1885.

Delfabbro, P., King, D.L., Williams, J., 2021. The psychology of cryptocurrency trading: risk and protective factors. *J. Behav. Addict.* 10, 201–207.

Dunn, D.S., Wilson, T.D., 1990. When the stakes are high: a limit to the illusion-of-control effect. *Soc. Cogn.* 8, 305–323.

Eaton, G.W., Green, T.C., Roseman, B.S., Wu, Y., 2022. Retail trader sophistication and stock market quality: evidence from brokerage outages. *J. Financ. Econ.* 146, 502–528.

- Fenton-O'Creevy, M., Nicholson, N., Soane, E., Willman, P., 2003. Trading on illusions: unrealistic perceptions of control and trading performance. *J. Occup. Organ. Psychol.* 76, 53–68.
- Fornell, C., Larcker, D.F., 1981. Evaluating structural equation models with unobservable variables and measurement error. *J. Mark. Res.* 18, 39–50.
- Glaser, M., Weber, M., 2007. Overconfidence and trading volume. *Geneva Risk Insur. Rev.* 32, 1–36.
- Grinblatt, M., Keloharju, M., 2009. Sensation seeking, overconfidence, and trading activity. *J. Finance* 64, 549–578.
- Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., 2019. *Multivariate Data Analysis*, eighth ed. Cengage Learning, Andover.
- Hasso, T., Müller, D., Pelster, M., Warkulat, S., 2019. Who trades cryptocurrencies, how do they trade it, and how do they perform? Evidence from brokerage accounts. *J. Behav. Exp. Finance* 23, 64–74.
- Heath, C., Tversky, A., 1991. Preference and belief: ambiguity and competence in choice under uncertainty. *J. Risk Uncertain.* 4, 5–28.
- Henseler, J., Ringle, C.M., Sarstedt, M., 2015. A new criterion for assessing discriminant validity in variance-based structural equation modeling. *J. Acad. Mark. Sci.* 43, 115–135.
- Hirshleifer, D., 2015. Behavioral finance. *Annu. Rev. Financ. Econ.* 7, 133–159.
- Hu, L., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Modeling* 6, 1–55.
- Huang, J.L., Curran, P.G., Keeney, J., Poposki, E.M., DeShon, R.P., 2012. Detecting and deterring insufficient effort responding to surveys. *J. Bus. Psychol.* 27, 99–114.
- Joshi, A., Kale, S., Chandel, S., Pal, D.K., 2015. Likert scale: explored and explained. *Br. J. Appl. Sci. Technol.* 7, 396–403.
- Kahneman, D., 2011. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, New York.
- Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. *Econometrica* 47, 263–292.
- Kim, H.J., Hong, J.S., Hwang, H.C., Kim, S.M., Han, D.H., 2020. Comparison of psychological status and investment style between bitcoin investors and share investors. *Front. Psychol.* 11, 1–9.
- Klayman, J., Soll, J.B., González-Vallejo, C., Barlas, S., 1999. Overconfidence: it depends on how, what, and whom you ask. *Organ. Behav. Hum. Decis. Process.* 79, 216–247.
- Klein, T., 2022. A note on GameStop, short squeezes, and autodidactic herding: an evolution in financial literacy? *Finance Res. Lett.* 46, article 102229.

- Kline, R.B., 2016. Principles and Practice of Structural Equation Modeling, fourth ed. Guilford Press, New York.
- Kumar, A., 2009. Who gambles in the stock market? *J. Finance* 64, 1889–1933.
- Langer, E.J., 1975. The illusion of control. *J. Pers. Soc. Psychol.* 32, 311–328.
- Leiner, D.J., 2019. Too fast, too straight, too weird: non-reactive indicators for meaningless data in internet surveys. *Surv. Res. Methods* 13, 229–248.
- List, J.A., 2003. Does market experience eliminate market anomalies? *Q. J. Econ.* 118, 41–71.
- Markiewicz, Ł., Weber, E.U., 2013. DOSPERT's gambling risk-taking propensity scale predicts excessive stock trading. *J. Behav. Finance* 14, 65–78.
- McNeish, D., 2018. Thanks coefficient alpha, we'll take it from here. *Psychol. Methods* 23, 412–433.
- Meade, A.W., Craig, S.B., 2012. Identifying careless responses in survey data. *Psychol. Methods* 17, 437–455.
- Meule, A., Vögele, C., Kübler, A., 2011. Psychometrische Evaluation der deutschen Barratt Impulsiveness Scale – Kurzversion (BIS-15). *Diagnostica* 57, 126–133.
- Michalczyk, R., Bowden-Jones, H., Verdejo-Garcia, A., Clark, L., 2011. Impulsivity and cognitive distortions in pathological gamblers attending the UK National Problem Gambling Clinic: a preliminary report. *Psychol. Med.* 41, 2625–2635.
- Miller, D.T., Ross, M., 1975. Self-serving biases in the attribution of causality: fact or fiction? *Psychol. Bull.* 82, 213–225.
- Moeller, F.G., Barratt, E.S., Dougherty, D.M., Schmitz, J.M., Swann, A.C., 2001. Psychiatric aspects of impulsivity. *Am. J. Psychiatry* 158, 1783–1793.
- Moore, D.A., Healy, P.J., 2008. The trouble with overconfidence. *Psychol. Rev.* 115, 502–517.
- Mosenhauer, M., Newall, P.W.S., Walasek, L., 2021. The stock market as a casino: associations between stock market trading frequency and problem gambling. *J. Behav. Addict.* 10, 683–689.
- Norman, G., 2010. Likert scales, levels of measurement and the “laws” of statistics. *Adv. Health Sci. Educ.* 15, 625–632.
- Nunnally, J.C., 1978. *Psychometric Theory*, second ed. McGraw-Hill, New York.
- O'Brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* 41, 673–690.
- Odean, T., 1999. Do investors trade too much? *Am. Econ. Rev.* 89, 1279–1298.
- Patton, J.H., Stanford, M.S., Barratt, E.S., 1995. Factor structure of the Barratt Impulsiveness Scale. *J. Clin. Psychol.* 51, 768–774.

Podsakoff, P.M., MacKenzie, S.B., Lee, J.-Y., Podsakoff, N.P., 2003. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *J. Appl. Psychol.* 88, 879–903.

Preacher, K.J., Hayes, A.F., 2008. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav. Res. Methods* 40, 879–891.

Presson, P.K., Benassi, V.A., 1996. Illusion of control: a meta-analytic review. *J. Soc. Behav. Pers.* 11, 493–510.

Raylu, N., Oei, T.P.S., 2004. The Gambling Related Cognitions Scale (GRCS): development, confirmatory factor validation and psychometric properties. *Addiction* 99, 757–769.

Robba, M., Richetin, J., Preti, E., 2024. Investor profiles, financial literacy, and overconfidence: a latent profile analysis. *J. Behav. Exp. Finance* 42, 1–12.

Roemer, E., Schubert, F., Henseler, J., 2021. HTMT2 — an improved criterion for assessing discriminant validity in structural equation modeling. *Ind. Manag. Data Syst.* 121, 2637–2650.

Rosseel, Y., 2012. lavaan: an R package for structural equation modeling. *J. Stat. Softw.* 48, 1–36.

Roth, M., Mayerhofer, D., 2014. Deutsche Version des Arnett Inventory of Sensation Seeking (AISS-d). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS), GESIS – Leibniz-Institut für Sozialwissenschaften, Mannheim. <https://doi.org/10.6102/zis73> (accessed 11 March 2026).

Roth, M., Schumacher, J., Brähler, E., 2005. Sensation seeking in the community: sex, age and sociodemographic comparisons on a representative German population sample. *Pers. Individ. Differ.* 39, 1261–1271.

Rotter, J.B., 1966. Generalized expectancies for internal versus external control of reinforcement. *Psychol. Monogr.* 80, 1–28.

Stanford, M.S., Mathias, C.W., Dougherty, D.M., Lake, S.L., Anderson, N.E., Patton, J.H., 2009. Fifty years of the Barratt Impulsiveness Scale: an update and review. *Pers. Individ. Differ.* 47, 385–395.

Statman, M., Thorley, S., Vorkink, K., 2006. Investor overconfidence and trading volume. *Rev. Financ. Stud.* 19, 1531–1565.

Stefan, S., David, D., 2013. Recent developments in the experimental investigation of the illusion of control: a meta-analytic review. *J. Appl. Soc. Psychol.* 43, 377–386.

Steinberg, L., 2008. A social neuroscience perspective on adolescent risk-taking. *Dev. Rev.* 28, 78–106.

Strömbäck, C., Lind, T., Skagerlund, K., Västfjäll, D., Tinghög, G., 2017. Does self-control predict financial behavior and financial well-being? *J. Behav. Exp. Finance* 14, 30–38.

Subrahmanyam, A., 2008. Behavioural finance: a review and synthesis. *Eur. Financ. Manag.* 14, 12–29.

Verdejo-García, A., Lawrence, A.J., Clark, L., 2008. Impulsivity as a vulnerability marker for substance-use disorders: review of findings from high-risk research, problem gamblers and genetic association studies. *Neurosci. Biobehav. Rev.* 32, 777–810.

Ward, S.J., King, L.A., 2020. Examining the roles of intuition and gender in magical beliefs. *J. Res. Pers.* 86, article 103956.

West, S.G., Finch, J.F., Curran, P.J., 1995. Structural equation models with nonnormal variables: problems and remedies. In: Hoyle, R.H. (Ed.), *Structural Equation Modeling: Concepts, Issues, and Applications*. Sage, Thousand Oaks, pp. 56–75.

Whiteside, S.P., Lynam, D.R., 2001. The five factor model and impulsivity: using a structural model of personality to understand impulsivity. *Pers. Individ. Differ.* 30, 669–689.

Wong, A., Carducci, B.J., 1991. Sensation seeking and financial risk taking in everyday money matters. *J. Bus. Psychol.* 5, 525–530.

Wood, J.M., Clapham, M.M., 2005. Development of the Drake Beliefs about Chance inventory. *J. Gambl. Stud.* 21, 411–430.

Zuckerman, M., 1994. *Behavioral Expressions and Biosocial Bases of Sensation Seeking*. Cambridge University Press, Cambridge.