Dipartimento di / Department of

Dottorato di Ricerca in / PhD program Psychology, Linguistics and Cognitive Neuroscience XXXVIII Ciclo / Cycle

Dear Diary: Advances in Experience Sampling Methodology Studies

TESI IN COTUTELA

Cognome / Surname Galkina Nome / Name Anastasiia

Matricola / Registration number 851053

Tutore / Tutor: Prof. Giulio Costantini

Cotutore / Co-tutor: Prof. Marco Perugini
(se presente / if there is one)

Cotutore / Co-tutor: Prof. Wolfgang Viechtbauer
(se presente / if there is one)

Coordinatore / Coordinator: Prof. Simona Sacchi

ANNO ACCADEMICO / ACADEMIC YEAR 2022-2025

Dear Diary: Advances in Experience Sampling Methodology Studies

DISSERTATION

To obtain the double degree of Doctor at Maastricht University and at Università degli Studi di Milano-Bicocca,

on the authority of the Rectores Magnifici,

Prof. Dr. Pamela Habibović and Prof. Dr. Marco Orlandi,

in accordance with the decision of the Board of Deans,

to be defended in public on Thursday 16 April 2026 at 10:00 hours CET at Maastricht University

and on 12 May 2026 at 14:00 hours at Università degli Studi di Milano-Bicocca

by Anastasiia Galkina

Supervisors:

Dr. Wolfgang Viechtbauer, Maastricht University

Prof. Dr. Marco Perugini, University of Milano-Bicocca

Co-supervisor:

Dr. Giulio Costantini, University of Milano-Bicocca

Assessment Committee:

Chair:

Prof. Dr. Philippe Delespaul, Maastricht University

Members:

Prof. Dr. Marcello Gallucci, University of Milano-Bicocca

Dr. Marlies Houben, Tilburg University

Dr. Marco di Sarno, University of Milano-Bicocca

Dr. Laura Sels, Ghent University

Dr. Claudia Simons, Maastricht University

Financial contributions

The project described in Part I of this dissertation was made possible through the support of a grant from The Honesty Project at Wake Forest University and the John Templeton Foundation (#61842), awarded to Giulio Costantini and Marco Perugini (#21-021). The opinions expressed in this work are those of the authors and do not necessarily reflect the views of The Honesty Project, Wake Forest University, or John Templeton Foundation.

There are two possible outcomes: if the result confirms the hypothesis, then you've made a measurement. If the result is contrary to the hypothesis, then you've made a discovery.
Enrico Fermi

1

General Introduction

Experience Sampling Methodology and Its Challenges

The Experience Sampling Method (ESM), also known as Ecological Momentary Assessment (EMA), has gained increasing prominence in psychological research over recent years (Bolger & Laurenceau, 2013; Hamaker & Wichers, 2017; Trull & Ebner-Priemer, 2014). The last decade has witnessed a particularly marked proliferation of ESM research publications, with growth rates that appear to exceed the broader expansion of scientific publishing as a whole (Fritz et al., 2024). The ESM approach offers significant advantages over cross-sectional studies, as it enables researchers to collect real-time data on variables of interest as they naturally occur in participants' daily lives, typically through brief, repeated assessments completed on smartphones at multiple points throughout the day. By capturing participants' emotions, behaviors, experiences, and contextual circumstances in the moment, ESM provides high ecological validity and substantially reduces the retrospective recall biases inherent in traditional assessment methods (Lucas et al., 2021). ESM approaches have revealed crucial insights into personality processes that would not be detectable through cross-sectional or less intensive designs (Trull & Ebner-Priemer, 2020). They allow researchers to distinguish between-person differences (how people differ from each other on average) from within-person processes (how an individual's states fluctuate over time and in response to situations) (Conner et al., 2009; Viechtbauer, 2022). This distinction is fundamental for understanding personal dynamics and has important implications for interventions (Schuurman, 2023).

However, ESM research also presents unique analytical challenges. The intensive longitudinal data generated by ESM studies involve complex nested structures, including repeated observations within individuals, that can be further nested in different groups. These complex data structures necessitate the application of specialized statistical methods (Viechtbauer, 2022). In fact, studying dyadic or even more complex interpersonal (e.g., family) processes benefits significantly from the use of ESM methodology (Ebner-Priemer et al., 2009;

Laurenceau & Bolger, 2005). However, there are still many unresolved issues, such as defining the levels of analysis and disentangling within-dyad from between-dyad effects, and the choice of the most appropriate analytical approaches that would satisfy both theoretical and statistical requirements (Thorson et al., 2018). Another demanding aspect of ESM research is designing relevant questionnaires, which requires not only conceptual precision and methodological rigor but also the ability to anticipate the practical challenges and patterns that may emerge throughout the data collection process (Eisele et al., 2022). An additional methodological gap in ESM studies stems from the fact that many variables of interest in ESM research, especially in the field of psychopathology, exhibit highly skewed distributions with floor or ceiling effects (Haslbeck et al., 2023). These distributional characteristics can violate assumptions of standard statistical models, potentially leading to biased estimates and incorrect conclusions (Baird et al., 2006; Dejonckheere et al., 2024). The aim of this thesis is to apply advanced ESM methods and data-analytical approaches to the study of personality and emotional dynamics, considering the abovementioned challenges

Development and Validation of ESM Measurements

Traditional self-report questionnaires, despite typically undergoing comprehensive validation, are generally incompatible with ESM research due to their extensive length and fundamentally different measurement objectives. Specifically, conventional instruments focus on retrospective trait assessment, whereas ESM questionnaires are designed to capture momentary experiences and states (Eisele et al., 2025). However, ESM questionnaire development cannot be approached only by shortening and adapting existing retrospective scales. The method demands principled decisions about item content, temporal contextualization, response formats, and sampling schemes (Wright & Zimmermann, 2019). Additionally, the need for brief, contextually appropriate items suitable for repeated administration without causing participant fatigue creates tension between comprehensiveness

and practicality (Fritz et al., 2024).

Assessing the psychometric properties of new ESM questionnaires is an important analytical step, which, however, is rarely done in the existing literature (Degroote et al., 2020; Hall et al., 2021). Both basic and applied research frequently employ the problematic practice of selecting items based on their presumed relevance to target constructs, often without conducting a comprehensive psychometric evaluation to establish their validity and reliability (Leertouwer et al., 2025; Wright & Zimmermann, 2019). Hall and colleagues (2021) found that over half of published ESM studies failed to report reliability or validity indices, with many providing insufficient detail about item content, response formats, and temporal instructions. The validation of dynamic statistics has been conducted almost exclusively with momentary ratings of emotions (Houben et al., 2015), leaving their validity largely undemonstrated for other psychological domains such as personality trait expression, interpersonal behavior, or situational experiences. The heterogeneity in measurement approaches across studies, while partly reflecting diverse research questions, often stems from inadequate attention to psychometric principles rather than theoretical considerations (Fritz et al., 2024). A lot of confusion also comes from the need to associate the levels of reliability estimations with the levels of the main analysis (between and/or within), as well as from the multiple techniques to assess reliability that exist in the literature (Castro-Alvarez et al., 2025). The lack of both expert consensus about how to determine such properties and already validated ESM questionnaires creates the need for studies where ESM questionnaires are created with transparency and based on empirical evidence (Dejonckheere et al., 2022; Fritz et al., 2024).

Traditional psychometric frameworks prove inadequate for ESM data because they were designed for between-person comparisons rather than within-person fluctuations over time (Dejonckheere et al., 2022). ESM items possess distinct characteristics that necessitate

specialized methodological considerations compared to traditional questionnaire items (Degroote et al., 2020; Horstmann & Ziegler, 2020). A fundamental distinction is that ESM items are administered repeatedly within real-world contexts, which can magnify methodological challenges that may also affect traditional questionnaires but remain less apparent in single-administration formats (Eisele et al., 2025).

The importance of rigorous ESM questionnaire development extends beyond purely methodological concerns. Unreliable measurements critically compromise data quality, leading to biased parameter estimates, attenuated effect sizes, reduced statistical power, and increased sample size requirements (Dejonckheere et al., 2022; Wright & Zimmermann, 2019). There are relatively few studies that demonstrate the necessary attention to these methodological issues (e.g., Gander et al., 2025; Lenferink et al., 2022), but they still represent rare examples of proper psychometric evaluation of ESM items. Insufficient reporting on ESM item development and evaluation practices undermines researchers' ability to assess the validity of study conclusions, enables the proliferation of poorly constructed items across subsequent studies, and consequently impedes the advancement of cumulative scientific knowledge (Vogelsmeier et al., 2024). These issues impede scientific progress in understanding psychological dynamics and limit the development of personalized interventions that depend on accurate real-time assessment.

Addressing these ESM implementation challenges requires confronting a fundamental but often neglected prerequisite: the availability of psychometrically validated measures specifically designed for intensive longitudinal assessment. While ESM research has proliferated across psychological domains, the field has frequently relied on ad hoc adaptation of existing trait questionnaires to momentary formats, with limited systematic investigation of whether such translations preserve construct validity, maintain factorial structure, or capture

meaningful within-person variation rather than measurement artifacts (Hamaker, 2012; Wright & Zimmermann, 2019). This practice becomes particularly problematic for complex personality constructs that encompass multiple facets, involve both trait-like dispositions and process-like mechanisms, and manifest differently across temporal scales. The development of ESM-appropriate measures therefore requires a systematic, multi-stage approach: first establishing comprehensive cross-sectional instruments through transparent, bottom-up procedures that ensure ecological validity and minimize researcher bias; second, identifying the motivational or process-level mechanisms through which stable traits translate into behavioral tendencies; and third, validating adapted measures at the momentary level while demonstrating that they capture genuine within-person dynamics rather than merely reflecting participant confusion about temporal framing or statistical artifacts of repeated measurement. This thesis adopts precisely such a sequential framework, using Honesty-Humility as an empirical case study to demonstrate a generalizable methodology for developing theoretically grounded, psychometrically robust, and behaviorally predictive measures suitable for intensive longitudinal research.

Interpersonal ESM: The Case of Dyadic Relationships

Beyond individual personality processes, ESM research enables the investigation of interpersonal dynamics — how people's psychological states are linked to those of close others. Romantic relationships provide a particularly important context for such an investigation (Liu et al., 2021; Rauers & Riediger, 2023). Partners in romantic relationships continuously influence each other's emotional experiences through daily interactions, creating dynamic emotional systems that extend beyond individual boundaries (Butler, 2011).

ESM has substantially enriched couple research by enabling the capture of naturally occurring relationship dynamics within their ecological contexts, providing unprecedented insights into daily-level processes such as emotional transmission, coregulation, and temporal

fluctuations in intimacy that traditional cross-sectional approaches cannot detect (Laurenceau & Bolger, 2005; Schoebi & Randall, 2015). This methodology allows researchers to examine how intimate relationships shape partners' experiences over time while minimizing retrospective biases and capturing contextual influences as they naturally unfold (Butler & Barnard, 2019).

However, ESM in dyadic research presents substantial analytical and methodological challenges. The nested structure of dyadic data — with observations nested within persons, and persons nested within couples — requires sophisticated approaches, such as multilevel modeling (MLM) or multilevel structural equation modeling (SEM), to address non-independence between partners (Ledermann & Kenny, 2017). Critical technical considerations include distinguishing between distinguishable versus indistinguishable dyads, selecting appropriate lag lengths and interval timing to capture underlying processes, handling missing data effectively, and choosing the appropriate analytical model considering its underlying assumptions (Helm et al., 2018). Moreover, dyadic ESM studies require careful choices in terms of theoretical framework, as well as in terms of statistical decisions, such as handling autocorrelation, or estimation of a single or separate models (Thorson et al., 2018). Additional challenges encompass participant compliance and potential reactivity to intensive measurement, power requirements for detecting often small effect sizes, and the computational complexity of analyzing higher-order interactions (Laurenceau & Bolger, 2005). Considering all these complexities, more dyadic ESM studies are needed to provide evidence on the feasibility of this type of research.

The Problem of Skewed Data in ESM Studies

ESM studies routinely encounter right-skewed distributions when measuring psychological constructs such as negative affect, psychopathological symptoms, and other bounded emotional states (Haslbeck et al., 2023; Mestdagh & Dejonckheere, 2021; Siepe et

al., 2025). These distributional patterns manifest as pronounced floor effects, where substantial proportions of observations cluster at minimum scale values, with ceiling effects occasionally observed at the opposite extreme (e.g., Wang et al., 2008). The prevalence of such skewness stems from multiple sources, including scale boundary constraints, item formulation considerations, temporal sampling parameters, and insufficient construct relevance to participants' daily experiences (Dejonckheere et al., 2024; Haslbeck et al., 2023).

The consequences of ignoring these distributional characteristics extend far beyond statistical technicalities to fundamental misinterpretations of psychological phenomena. Most critically, bounded rating scales can create artificial structural dependencies between mean levels and maximum possible variability, generating spurious associations that can lead to entirely incorrect scientific conclusions (Baird et al., 2006; Dejonckheere et al., 2024; Mestdagh et al., 2018). A compelling example is provided by Kalokerinos et al. (2020), who demonstrated that the previously established association between neuroticism and emotional variability may have been largely artifactual: when controlling for mean levels of negative emotions, this relationship disappeared entirely, revealing that the apparent association was apparently just a consequence of bounded scale measurement rather than a genuine psychological process.

Standard linear mixed-effects models, a common statistical tool in ESM research, can be problematic when applied to such skewed data because core statistical assumptions of such models, including normality of residuals and random effects, are often violated when floor or ceiling effects are present, leading to systematically biased parameter estimates, attenuated effect sizes, and distorted intercept-slope correlations (Terluin et al., 2016). This problem, therefore, necessitates an alternative statistical approach that can effectively handle the nature of bounded psychological data and ensure more reliable estimates.

Aims and Outline of This Thesis

This thesis examines the recent applications of ESM in various domains. First, it demonstrates how intensive longitudinal methods can advance our understanding of personality dynamics through the creation and validation of an ESM questionnaire in a bottom-up procedure. Second, it illustrates the application of ESM in dyadic research by focusing on the emotional interdependence of partners in romantic relationships. Finally, the work addresses methodological issues of applying the standard linear mixed-effects model to skewed ESM data and proposes an alternative theoretical and statistical approach. The thesis is organized into five empirical chapters, grouped in three parts.

Part I: Systematic Bottom-Up Development of Honesty-Humility Measures

Chapters 2, 3, and 4 form an integrated sequence focused on developing and validating measures of Honesty-Humility suitable for both cross-sectional trait assessment and intensive longitudinal research. These chapters share a common methodological philosophy: using systematic, bottom-up approaches to ensure that measures capture constructs as they naturally manifest in language and experience.

The development of appropriate measurement instruments represents a foundational challenge for ESM research, particularly when investigating constructs that have traditionally been assessed through standard questionnaires in a cross-sectional paradigm. **Chapters 2** and **3** address the initial stage of this challenge systematically by developing two complementary cross-sectional measures of Honesty-Humility that are theoretically grounded, psychometrically robust, and possess characteristics that make them particularly suitable for subsequent adaptation to intensive longitudinal contexts. Both chapters adopt a transparent, bottom-up methodology that prioritizes participant-generated content over researcher-imposed theoretical frameworks, thereby minimizing investigator bias and enhancing ecological validity. These measures establish the empirical and conceptual foundation that **Chapter 4** then

leverages to construct a fully validated ESM questionnaire capable of capturing momentary fluctuations in honest and dishonest states. This sequential approach—developing rigorously validated cross-sectional instruments before adapting them to ESM contexts—represents a paradigm shift from traditional top-down instrument construction, offering a systematic template for ESM measure development across personality domains. The focus on Honesty-Humility—the sixth dimension of the HEXACO personality model (Ashton & Lee, 2008)—is particularly important for the personality psychology field due to its demonstrated relevance to consequential real-world behaviors including ethical decision-making, interpersonal trust, and dishonest conduct (Ashton & Lee, 2008; Hilbig et al., 2013), yet existing measures present significant practical limitations for repeated momentary assessment.

Chapter 2 presents the development and validation of the Adjective Checklist of Honesty (ACH), a comprehensive adjective-based measure that captures the four main facets of Honesty-Humility: Sincerity, Fairness, Modesty, and Greed-Avoidance. We began with an extensive corpus of Italian adjectives potentially related to Honesty-Humility, drawn from previous lexical studies and expanded through systematic examination of semantic spaces and dictionaries. Five independent expert raters then judged how well each adjective represented each facet, providing an objective basis for item classification. This bottom-up procedure yielded a 22-item scale, which we validated across multiple studies using self-reports, informant-reports, and associations with theoretically relevant constructs. Importantly, our work addresses ongoing debates about the relationship between Honesty-Humility and Truthfulness, demonstrating that Truthfulness is conceptually indistinguishable from the Sincerity facet in natural language use – a finding with important implications for how we understand and measure this trait.

Chapter 3 addresses a fundamental limitation in contemporary personality research: the predominant focus on what people are like (trait descriptions) at the expense of understanding why they behave as they do (motivational processes). While traits such as Honesty-Humility describe stable patterns of individual differences, they provide limited insight into the goal-directed processes through which dispositions translate into concrete actions in specific situations. According to integrative frameworks such as Whole Trait Theory (Fleeson & Jayawickreme, 2015) personal goals represent a critical intermediate level of analysis between abstract trait descriptions and observable behaviors—they capture what individuals are striving toward or attempting to avoid, thereby serving as the proximal motivational mechanisms that govern behavioral enactment.

Chapter 3 extends the systematic approach from **Chapter 2** to uncover the motivational core of Honesty-Humility, developing the Goals for Honesty and Goals for Dishonesty (GH/GD) questionnaire. Following a similar bottom-up methodology, we used personality-descriptive adjectives as prompts to elicit goal statements from 250 participants, generating over 9,000 textual responses. We then employed a combination of natural language processing algorithms and independent rater classifications to identify and organize these goals into meaningful categories. This process revealed 48 goal classes potentially related to Honesty-Humility, for which we developed 191 preliminary questionnaire items. In a second study with 300 participants, we refined this set by examining empirical associations between goals and adjective-based personality assessments, retaining 21 goals (17 related to honesty, 4 to dishonesty) assessed by 78 items that showed specific associations with Honesty-Humility. A third validation study (N = 300) demonstrated that these goals predicted self-reported behavioral tendencies, informant-reported lying behavior, and actual dishonest behavior in an incentivized cheating task. This finding indicates that motivational constructs capture variance in behavioral outcomes that stable dispositional measures systematically miss. Crucially, we

demonstrated that goals for honesty and goals for dishonesty represent distinct motivational orientations rather than opposite poles of a single continuum, indicating that individuals can simultaneously endorse approach-oriented strivings toward honesty and possess varying levels of temptation or motivation toward dishonesty. These findings provide empirical support for the theoretical proposition that comprehensive understanding of personality requires moving beyond trait taxonomies to examine the goal structures and motivational processes through which dispositions become manifest in everyday actions.

Chapter 4 completes this measurement development sequence by adapting the GH/GD questionnaire for use in intensive longitudinal research. We created an ESM/EMA version (GH/GD-EMA) by selecting 21 goal items suitable for capturing momentary variations in goal importance and reformulating them to assess importance "during the last hour" rather than in general. In a 15-day ESM study involving 198 participants who received five daily prompts, we systematically examined the psychometric properties of this adapted measure. Multilevel confirmatory factor analyses evaluated the factor structure at both the within- and between-person levels, revealing a two-factor structure (GH and GD) that replicated across both levels of analysis. We estimated reliability separately for the within- and between-person components using multilevel factor analysis frameworks, demonstrating acceptable reliability at both levels for honest goals, though lower within-person reliability for dishonest goals. Most importantly, we validated the measure by demonstrating that momentary goal importance robustly predicted Honesty-Humility states both contemporaneously and temporally, and that changes in goals during the ESM period predicted trait-level change from pre- to post-assessment through their effects on state changes. This validation provides strong evidence that the GH/GD-EMA successfully captures the dynamic motivational processes underlying Honesty-Humility as they unfold in daily life.

Together, these three chapters demonstrate the value of systematic, bottom-up approaches to measurement development. By starting from natural language use (**Chapter 2**) and participant-generated goal statements (**Chapter 3**), ensuring alignment with theoretical definitions through expert ratings, and rigorously validating measures through multiple methods, including intensive longitudinal designs (**Chapter 4**), we developed a suite of instruments that capture Honesty-Humility at multiple levels—from stable traits to dynamic goals to momentary states. This integrated measurement approach provides the foundation for examining personality processes as they unfold in daily life.

It is important to mention that the integrated approach across **Chapters 2, 3**, and **4** also addresses a critical theoretical gap in contemporary personality research: the need to understand personality constructs not merely as stable between-person differences, but as multi-level phenomena encompassing traits, motivational processes, and momentary states that dynamically unfold in daily life. This comprehensive examination of Honesty-Humility holds particular theoretical significance for several reasons. First, Honesty-Humility represents the most recent major addition to structural models of personality, yet remains underexplored relative to the Big Five dimensions, with substantial debate continuing regarding its facet structure, discriminant validity from related constructs, and the psychological processes through which it manifests in behavior (Ashton & Lee, 2020; Thielmann et al., 2020). Second, unlike traits such as Extraversion or Neuroticism, where the translation from disposition to behavior is relatively straightforward, Honesty-Humility encompasses a complex interplay between moral cognition, self-interest regulation, and interpersonal orientation that requires examination at multiple levels of analysis to be fully understood (Hilbig et al., 2013). The distinction between having honest traits, endorsing honest goals, and enacting honest behaviors in specific moments represents fundamentally different psychological phenomena that existing research has often conflated through exclusive reliance on trait-level assessment. Third,

understanding the motivational architecture underlying Honesty-Humility—particularly the empirical demonstration that goals for honesty and goals for dishonesty constitute separate dimensions rather than bipolar opposites—challenges traditional assumptions about the structure of moral personality and has direct implications for predicting when individuals will behave honestly versus dishonestly in real-world contexts. By systematically developing measures that capture Honesty-Humility at the trait level (stable individual differences), process level (motivational goals and intentions), and state level (momentary fluctuations in daily life), this research provides a comprehensive empirical test of integrative personality frameworks such as Whole Trait Theory (Fleeson & Jayawickreme, 2015) and Cybernetic Big Five Theory (DeYoung, 2015), which propose that complete understanding of personality requires examining characteristic adaptations, goal structures, and dynamic within-person processes alongside traditional trait descriptions. Furthermore, the bottom-up, participant-centered methodology employed across all three chapters ensures that the resulting theoretical insights are grounded in how individuals actually experience and express honesty in natural language and everyday contexts, rather than being constrained by a top-down framework.

Part II: Interpersonal Emotion Dynamics

Chapter 5 shifts focus to interpersonal processes, examining emotional interdependence in romantic relationships. Using data from 76 couples who completed ESM assessments five times daily over a 30-day period, we investigate both contemporaneous and temporal patterns of emotional interdependence between partners. Results reveal significant interdependence for positive affect but not negative affect, with modest partner effects that persist across time. Substantial between-couple variability suggests that emotional interdependence is heterogeneous, highlighting the need for more nuanced approaches to understanding interpersonal emotion dynamics.

Part III: Methodological Advances

Chapter 6 addresses a fundamental methodological challenge in ESM research: the analysis of right-skewed data. Many variables of interest in ESM studies exhibit floor or ceiling effects, which can lead to biased parameter estimates when analyzed with standard linear mixed-effects models. We propose conceptualizing such data as censored manifestations of underlying latent variables and demonstrate that mixed-effects tobit models provide more accurate parameter recovery than standard approaches. Using both artificially censored data and naturally skewed negative affect data, we show that tobit models better preserve true effect sizes and reveal greater individual variability.

In **Chapter 7**, the main findings of this thesis are summarized and discussed. This is done in the light of its role in addressing methodological and applied challenges of ESM research. The thesis will conclude with directions for future ESM studies.

References

Ashton, M. C., & Lee, K. (2008). The HEXACO Model of Personality Structure and the Importance of the H Factor. *Social and Personality Psychology Compass*, *2*(5), 1952–1962. https://doi.org/10.1111/j.1751-9004.2008.00134.x

Ashton, M. C., & Lee, K. (2020). Objections to the HEXACO Model of Personality Structure—And why those Objections Fail. *European Journal of Personality*, *34*(4), 492–510. https://doi.org/10.1002/per.2242

Baird, B. M., Le, K., & Lucas, R. E. (2006). On the nature of intraindividual personality variability: Reliability, validity, and associations with well-being. *Journal of Personality and Social Psychology*, *90*(3), 512–527. https://doi.org/10.1037/0022-3514.90.3.512

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research.* The Guilford Press.

Butler, E. A. (2011). Temporal Interpersonal Emotion Systems: The "TIES" That Form Relationships. *Personality and Social Psychology Review*, *15*(4), 367–393. https://doi.org/10.1177/1088868311411164

Butler, E. A., & Barnard, K. J. (2019). Quantifying Interpersonal Dynamics for Studying Socio-Emotional Processes and Adverse Health Behaviors. *Psychosomatic Medicine*, *81*(8), 749–758. https://doi.org/10.1097/PSY.0000000000000703

Castro-Alvarez, S., Bringmann, L. F., Back, J., & Liu, S. (2025). The Many Reliabilities of Psychological Dynamics: An Overview of Statistical Approaches to Estimate the Internal Consistency Reliability of Intensive Longitudinal Data. Preprint. https://osf.io/preprints/psyarxiv/qyk2r_v4

Conner, T. S., Tennen, H., Fleeson, W., & Barrett, L. F. (2009). Experience Sampling Methods: A Modern Idiographic Approach to Personality Research. *Social and Personality Psychology Compass*, *3*(3), 292–313. https://doi.org/10.1111/j.1751-9004.2009.00170.x

Degroote, L., DeSmet, A., De Bourdeaudhuij, I., Van Dyck, D., & Crombez, G. (2020). Content validity and methodological considerations in ecological momentary assessment studies on physical activity and sedentary behaviour: A systematic review. *The International Journal of Behavioral Nutrition and Physical Activity*, *17*(1), 35. https://doi.org/10.1186/s12966-020-00932-9

Dejonckheere, E., Demeyer, F., Geusens, B., Piot, M., Tuerlinckx, F., Verdonck, S., &

Mestdagh, M. (2022). Assessing the reliability of single-item momentary affective measurements in experience sampling. *Psychological Assessment*, *34*(12), 1138–1154. https://doi.org/10.1037/pas0001178

Dejonckheere, E., Penne, I., Briels, L., & Mestdagh, M. (2024). For Better or for Worse? Visualizing Previous Intensity Levels Improves Emotion (Dynamic) Measurement in Experience Sampling. *Psychological Assessment*, *36*(3), 215–234. https://doi.org/10.1037/pas0001296

DeYoung, C. G. (2015). Cybernetic Big Five Theory. *Journal of Research in Personality*, *56*, 33–58. https://doi.org/10.1016/j.jrp.2014.07.004

Ebner-Priemer, U. W., Eid, M., Kleindienst, N., Stabenow, S., & Trull, T. J. (2009). Analytic strategies for understanding affective (in)stability and other dynamic processes in psychopathology. *Journal of Abnormal Psychology*, *118*(1), 195–202. https://doi.org/10.1037/a0014868

Eisele, G., Hiekkaranta, A., Kunkels, Y. K., Rot, M. A. H., Van Ballegooijen, W., Bartels, S. L., Bastiaansen, J. A., Beymer, P. N., Bylsma, L. M., Carpenter, R. W., Ellison, W. D., Fisher, A. J., Forkmann, T., Frumkin, M. R., Fulford, D., Naragon-Gainey, K., Greene, T., Heininga, V. E., Jones, A., … Kirtley, O. J. (2025). ESM-Q: A consensus-based quality assessment tool for experience sampling method items. *Behavior Research Methods*, *57*(4), 124. https://doi.org/10.3758/s13428-025-02626-1

Eisele, G., Kasanova, Z., & Houben, M. (2022). Questionnaire Design and Evaluation. In *I. Myin-Germeys & P. Kuppens (Eds.), The open handbook of Experience Sampling Methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies* (2nd ed., pp. 71–90). Center for Research on Experience Sampling and Ambulatory Methods Leuven.

Fleeson, W., & Jayawickreme, E. (2015). Whole Trait Theory. *Journal of Research in Personality*, *56*, 82–92. https://doi.org/10.1016/j.jrp.2014.10.009

Fritz, J., Piccirillo, M. L., Cohen, Z. D., Frumkin, M., Kirtley, O., Moeller, J., Neubauer, A. B., Norris, L. A., Schuurman, N. K., Snippe, E., & Bringmann, L. F. (2024). So You Want to Do ESM? 10 Essential Topics for Implementing the Experience-Sampling Method. *Advances in Methods and Practices in Psychological Science*, *7*(3), 25152459241267912. https://doi.org/10.1177/25152459241267912

Hall, M., Scherner, P. V., Kreidel, Y., & Rubel, J. A. (2021). A Systematic Review of Momentary Assessment Designs for Mood and Anxiety Symptoms. *Frontiers in Psychology*, *12*, 642044. https://doi.org/10.3389/fpsyg.2021.642044

Hamaker, E. L. (2012). Why researchers should think "within-person": A paradigmatic rationale. In M. R. Mehl & T. S. Conner (Eds.), *Handbook of research methods for studying daily life* (pp. 43–61). The Guilford Press.

Hamaker, E. L., & Wichers, M. (2017). No Time Like the Present: Discovering the Hidden Dynamics in Intensive Longitudinal Data. *Current Directions in Psychological Science*, *26*(1), 10–15. https://doi.org/10.1177/0963721416666518

Haslbeck, J., Ryan, O., & Dablander, F. (2023). Multimodality and skewness in emotion time series. *Emotion*, *23*(8), 2117–2141. https://doi.org/10.1037/emo0001218

Helm, J. L., Miller, J. G., Kahle, S., Troxel, N. R., & Hastings, P. D. (2018). On Measuring and Modeling Physiological Synchrony in Dyads. *Multivariate Behavioral Research*, *53*(4), 521–543. https://doi.org/10.1080/00273171.2018.1459292

Hilbig, B. E., Zettler, I., Leist, F., & Heydasch, T. (2013). It takes two: Honesty–Humility and Agreeableness differentially predict active versus reactive cooperation. *Personality and Individual Differences*, *54*(5), 598–603. https://doi.org/10.1016/j.paid.2012.11.008

Horstmann, K. T., & Ziegler, M. (2020). Assessing Personality States: What to Consider when Constructing Personality State Measures. *European Journal of Personality*, *34*(6), 1037–1059. https://doi.org/10.1002/per.2266

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*(4), 901–930. https://doi.org/10.1037/a0038822

Kalokerinos, E. K., Murphy, S. C., Koval, P., Bailen, N. H., Crombez, G., Hollenstein, T., Gleeson, J., Thompson, R. J., Van Ryckeghem, D. M. L., Kuppens, P., & Bastian, B. (2020). Neuroticism may not reflect emotional variability. *Proceedings of the National Academy of Sciences*, *117*(17), 9270–9276. https://doi.org/10.1073/pnas.1919934117

Laurenceau, J.-P., & Bolger, N. (2005). Using Diary Methods to Study Marital and Family Processes. *Journal of Family Psychology*, *19*(1), 86–97. https://doi.org/10.1037/0893-3200.19.1.86

Ledermann, T., & Kenny, D. A. (2017). Analyzing dyadic data with multilevel modeling versus structural equation modeling: A tale of two methods. *Journal of Family Psychology*, *31*(4), 442–452. https://doi.org/10.1037/fam0000290

Leertouwer, Ij., Keijsers, L., van Roekel, E., & Schuurman, N. (2025). A Practical Guide to Estimating Reliability of Intensive Longitudinal Data. *Preprint*.

Liu, D. Y., Strube, M. J., & Thompson, R. J. (2021). Interpersonal Emotion Regulation: An Experience Sampling Study. *Affective Science*, *2*(3), 273–288.

https://doi.org/10.1007/s42761-021-00044-y

Lucas, R. E., Wallsworth, C., Anusic, I., & Donnellan, M. B. (2021). A direct comparison of the day reconstruction method (DRM) and the experience sampling method (ESM). *Journal of Personality and Social Psychology*, *120*(3), 816–835. https://doi.org/10.1037/pspp0000289

Mestdagh, M., & Dejonckheere, E. (2021). Ambulatory assessment in psychopathology research: Current achievements and future ambitions. *Current Opinion in Psychology*, *41*, 1–8. https://doi.org/10.1016/j.copsyc.2021.01.004

Mestdagh, M., Pe, M., Pestman, W., Verdonck, S., Kuppens, P., & Tuerlinckx, F. (2018). Sidelining the mean: The relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychological Methods*, *23*(4), 690–707. https://doi.org/10.1037/met0000153

Rauers, A., & Riediger, M. (2023). Ease of Mind or Ties That Bind? Costs and Benefits of Disclosing Daily Hassles in Partnerships. *Social Psychological and Personality Science*, *14*(5), 551–561. https://doi.org/10.1177/19485506221112252

Schoebi, D., & Randall, A. K. (2015). Emotional Dynamics in Intimate Relationships. *Emotion Review*, *7*(4), 342–348. https://doi.org/10.1177/1754073915590620

Schuurman, N. K. (2023). *A "Within/Between Problem" Primer: About (Not) Separating Within-Person Variance and Between-Person Variance in Psychology.* https://doi.org/10.31234/osf.io/7zgkx

Siepe, B. S., Rieble, C. L., Tutunji, R., Rimpler, A., März, J., Proppert, R. K. K., & Fried, E. I. (2025). Understanding Ecological-Momentary-Assessment Data: A Tutorial on Exploring Item Performance in Ecological-Momentary-Assessment Data. *Advances in Methods and Practices in Psychological Science*, *8*(1), 25152459241286877. https://doi.org/10.1177/25152459241286877

Terluin, B., De Boer, M. R., & De Vet, H. C. W. (2016). Differences in Connection Strength between Mental Symptoms Might Be Explained by Differences in Variance: Reanalysis of Network Data Did Not Confirm Staging. *PLOS ONE*, *11*(11), e0155205. https://doi.org/10.1371/journal.pone.0155205

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological Bulletin*, *146*(1), 30–90. https://doi.org/10.1037/bul0000217

Thorson, K. R., West, T. V., & Mendes, W. B. (2018). Measuring physiological influence in dyads: A guide to designing, implementing, and analyzing dyadic physiological studies.

Psychological Methods, *23*(4), 595–616. https://doi.org/10.1037/met0000166

Trull, T. J., & Ebner-Priemer, U. (2014). The Role of Ambulatory Assessment in Psychological Science. *Current Directions in Psychological Science*, *23*(6), 466–470. https://doi.org/10.1177/0963721414550706

Trull, T. J., & Ebner-Priemer, U. (2020). Ambulatory Assessment in Psychopathology Research: A Review of Recommended Reporting Guidelines and Current Practices. *Journal of Abnormal Psychology*, *121*(1), 56–63. https://doi.org/10.1037/abn0000473

Viechtbauer, W. (2022). Statistical methods for ESM data. In *I. Myin-Germeys & P. Kuppens (Eds.), The open handbook of Experience Sampling Methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies* (2nd ed., pp. 153–183). Center for Research on Experience Sampling and Ambulatory Methods Leuven.

Vogelsmeier, L. V. D. E., Jongerling, J., & Maassen, E. (2024). Assessing and accounting for measurement in intensive longitudinal studies: Current practices, considerations, and avenues for improvement. *Quality of Life Research*, *33*(8), 2107–2118. https://doi.org/10.1007/s11136-024-03678-0

Wang, L., Zhang, Z., McArdle, J. J., & Salthouse, T. A. (2008). Investigating Ceiling Effects in Longitudinal Data Analysis. *Multivariate Behavioral Research*, *43*(3), 476–496. https://doi.org/10.1080/00273170802285941

Wright, A. G. C., & Zimmermann, J. (2019). Applied ambulatory assessment: Integrating idiographic and nomothetic principles of measurement. *Psychological Assessment*, *31*(12), 1467–1480. https://doi.org/10.1037/pas0000685

Part I

2

A lexical examination of the facets of Honesty-Humility: The Adjective Checklist of Honesty

Amenta, S.*, Galkina, A.*, Romano, D., Perugini, M., & Costantini, G. (2025). A lexical examination of the facets of honesty-humility: The adjective checklist of honesty. *European Journal of Personality*, *0*(0). https://doi.org/10.1177/08902070251394755

* The two authors contributed equally to this publication.

Abstract

The HEXACO model describes human personality through six major domains, each one further specified by four facets. Honesty-Humility is a key component of the model, and subsumes facets Sincerity, Fairness, Modesty, and Greed-Avoidance. Although established and comprehensive measurement tools of HEXACO traits and facets through sentence items exist, a comprehensive assessment of Honesty-Humility through adjectives has not yet been developed. We present a novel instrument, the Adjective Checklist of Honesty (ACH), designed to enhance the assessment of Honesty by incorporating lexical descriptors of its facets, developed using a systematic approach. In Study 1, we identified a broad set of potential descriptors of Honesty-Humility facets, and refined it through independent raters to identify candidate items. In Study 2 (N = 266), we examined the factorial structure of the candidate items and developed the final 22-item version of the ACH. In Study 3 (N = 300), we confirmed the factorial structure of the shortened ACH questionnaire and collected validity evidence for the new scale. Our studies have important implications for the assessment of Honesty-Humility, shed light on the relationships between Honesty-Humility and Truthfulness and, more generally, showcase the importance of a systematic approach in constructing assessment scales in personality psychology.

Keywords: Honesty-Humility, Truthfulness, Psycholexical approach, Adjective Checklist of Honesty

Introduction

According to the psycholexical hypothesis, personality traits are encoded in lexical descriptors (mostly adjectives) found in natural language use (Allport & Odbert, 1936; Goldberg, 1993, 2002). Looking to define the factorial structure of personality, current models relied, in fact, on the analysis of person-descriptive adjectives (e.g., Ashton & Lee, 2001; Caprara & Perugini, 1994; Costa Jr. & McCrae, 1988; De Raad et al., 1998) to identify a five-factor (e.g., Caprara & Perugini, 1994; McCrae & Costa, 1987) or six-factor (e.g., Ashton, Lee, & Goldberg, 2004) solution. The six-factor solution is at the basis of the HEXACO model (Ashton, Lee, & Goldberg, 2004) and has been consistently replicated across languages and cultures (e.g., Ashton et al., 2006; Ashton & Lee, 2005; Lee & Ashton, 2008) showing to portray the most reliable factorial structure (Ashton & Lee, 2008, 2020).

Differences between the five-factor solution, also known as Big Five (Goldberg, 2002), and the HEXACO model, can be traced, besides the obvious number of factors, also in the *semantic content* captured by each factor. The Big Five model encompasses five factors labeled as Extraversion, Agreeableness, Conscientiousness, Neuroticism vs. Emotional Stability, and Openness to Experience or Intellect, while the HEXACO model includes Honesty-Humility, Emotionality, eXtraversion, Agreeableness vs. Anger, Conscientiousness, and Openness to Experience (or Intellect/Imagination/Unconventionality). Some content overlaps between models: in particular the Extraversion, Conscientiousness, and Openness to Experience are similar between models (Lee & Ashton, 2008). Some differences appear in the Agreeableness and Emotionality content, which present a few shifts between the five- and six-factor models (see Lee & Ashton, 2008, for a detailed discussion of the differences between Big Five and HEXACO). The most notable difference is, however, the addition of a sixth factor that emerged from adjectives like *honest*, *sincere*, *humble*, *modest*, etc. (and their contraries: *dishonest*, *insincere*, *immodest*, etc.), labeled Honesty-Humility (Lee & Ashton, 2008).

The HEXACO Personality Inventory–Revised (HEXACO-PI-R), as well as the shorter HEXACO-100 and HEXACO-60, are widespread questionnaires that have been developed to assess the six HEXACO traits and a set of four facets for each major trait, plus Altruism vs. Antagonism as an interstitial facet (Ashton & Lee, 2009; Lee & Ashton, 2004, 2006, 2018). In the HEXACO-PI-R, Honesty-Humility is operationalized in terms of four facets: Sincerity, Fairness, Modesty, and Greed-Avoidance. The role of these facets is to capture specific aspects of Honesty-Humility (e.g., being generous or manipulating others), which reflects a tendency towards prosocial behavior, cooperation, altruism, and avoiding exploiting others (Ashton & Lee, 2008, 2020; Lee & Ashton, 2008).

Although the Honesty-Humility factor incorporates some aspects previously coded in the Agreeableness factor (McCrae & Costa, 2008), Agreeableness of the five-factor does not encompass the full content of Honesty-Humility (see Ashton & Lee, 2007, 2020). In the HEXACO model, Honesty-Humility and Agreeableness versus Anger are seen as complementary traits rather than hierarchically ordained (where one encapsulates the other). In fact, while Honesty-Humility can be ascribed to a general tendency to avoid exploiting others (active cooperation), Agreeableness versus Anger has been linked to an individual response to being exploited by others (reactive cooperation vs. retaliation; Ashton & Lee, 2007; Hilbig et al., 2013). This is particularly evident in the facet-level description of the two traits in the HEXACO-PI-R model, which helps to better understand the breadth of content they represent and the differences between the two traits (Ashton et al., 2009, 2014; Ashton & Lee, 2007, 2020). On the one hand, the facets of Agreeableness versus Anger denote how one person reacts to being exploited by others (Patience and Forgiveness), or during negotiations or interactions with others (Flexibility and Gentleness). On the other hand, the facets of Honesty-Humility denote the tendency to exploit others through lying or manipulating (low Sincerity and low Fairness), or to feeling entitled to a gain, to the point of being willing to exploit others to obtain

it (low Modesty and high Greed – the negative polarity of Greed-Avoidance). Positive polarities of Honesty-Humility facets are instead reflective of positive and non-exploitative interactions.

In terms of assessment, the phrased items included in the HEXACO-PI-R sometimes represent specific situations that might not apply to the personal experience of the person filling in the questionnaire, but by being asked in a conditional format (e.g., "I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed"), require respondents to engage in reasoning about their potential behaviors in specified situations. Recently, the HEXACO Adjective Scale (HAS; Romano et al., 2023) has been introduced as an alternative measure that assesses each HEXACO trait with 10 person-descriptive adjective items (e.g., "attentive"). Adjectives offer a different and more decontextualized way to assess personality, with adjectival self-ratings presumably representing an implicit aggregation of behavioral patterns demonstrated across diverse circumstances and contexts.

While the HAS includes adjectives related to Honesty-Humility, it, however, does not cover the facet structure of the trait in its full extension. Among the HAS adjectives, a few can be associated with Honesty-Humility facets: *sincere* is clearly related to Sincerity, while *humble, haughty,* and *snob* can be related to Modesty, even if *humble* maintains also some connections with Greed-Avoidance, which seems to be represented solely by *greedy.* There seems to be no adjective assessing specifically facet Fairness in the current HAS scales. The remaining adjectives, like *honest*, *dishonest*, *loyal, faithful,* and *hypocrite* seem to relate to a more general conceptualization of Honesty, which cannot be directly connected to a specific facet. Given the relevance of the facets for the correct conceptualization and measurement of Honesty-Humility, the first aim of this work is developing the Adjective Checklist[1] of Honesty

[1] We would like to clarify that, consistent with what we have done in the past (Costantini et al., 2015), we used the term "checklist" to indicate an instrument in which the respondents indicate the extent to which each adjective describes them on a multi-point scale, rather than an assessment in which the respondents indicate which items

(ACH), an analog of the HAS focusing specifically on lexical descriptors of honesty and its facets.

Since the identification of a sixth broad trait across languages and cultures, alternative labels have been considered for this dimension, until “Honesty-Humility” was chosen as a suitable descriptive name that could capture the essence of this personality dimension (Ashton et al., 2000; Ashton, Lee, Perugini, et al., 2004). Whereas substantial portions of this factor’s content align with common conceptualizations of honesty, certain aspects that individuals might consider central to honesty may remain tangential to this dimension. This debate still reverberates in the current discussion around the precise conceptualization of Honesty-Humility (Miller et al., 2021). A key issue concerns the discrepancy between the everyday meaning of "honesty" — which primarily relates to truthfulness, authenticity, directness, and avoiding deception—and its conceptualization and measurement within Honesty-Humility. For instance, the HEXACO-PI-R items assessing the Sincerity and Fairness facets, which are conceptually closest to Truthfulness, tend to focus more on exploitative behaviors than on truthfulness itself (Fleeson et al., 2022). This has led some scholars to argue that Truthfulness may be underrepresented in assessments of Honesty-Humility, prompting suggestions to rename the trait as "Benevolence / Simplicity" (Fleeson, 2020).

Empirical findings on the relationship between Honesty-Humility and Truthfulness have been mixed. Some studies suggest that Honesty-Humility is associated with Truthfulness regardless of prosocial considerations and benevolence (Thielmann et al., 2024). Others indicate that individuals high in Honesty-Humility may lie for prosocial reasons (Choshen-Hillel et al., 2020; McArthur et al., 2022, 2024; Paul et al., 2022; Ścigała et al., 2020). The second aim of this study is thus to further examine the relationships between Truthfulness and

---

they consider accurate (i.e., by giving a checkmark to some adjectives and not others), as the term checklist might suggest.

Honesty-Humility. To this end, we considered Truthfulness alongside the four existing Honesty-Humility facets and investigated whether an adjective-based assessment of Truthfulness could be generated without excessive conceptual overlap with respect to the other facets. Furthermore, we decided to include in the ACH a larger set of adjectives representing Truthfulness, to obtain a measure that would represent this aspect more clearly.

Throughout our work, we followed a transparent and documented item selection process (Flake & Fried, 2020). Studies 1 and 3 were preregistered. Study 2 was part of a larger study that included other measures not analyzed here. Preregistrations, data, and analysis code are available at https://osf.io/94w6r/overview?view_only=df7bea97fc5145ee8bb9ad3f73bdfe25.

Study 1: Identification of descriptors of Honesty-Humility facets

The aim of Study 1 was to generate an initial set of candidate items to assess honesty and its facets. Personality-descriptive adjectives have been previously used to investigate different aspects of personality, including HEXACO traits (e.g., Ashton, Lee, & Goldberg, 2004; Ashton & Lee, 2005; Costantini & Perugini, 2018; Perugini & Leone, 2009; Romano et al., 2023). Previous attempts, however, did not consider the full complexity of the Honesty-Humility trait or its relationship with the general construct of honesty and truthfulness. We, therefore, started by identifying a comprehensive set of adjectives related to honesty and its facets, and we aimed at reducing this number, relying on the evaluation of expert judges in order to identify adjectives that could describe specific facets of honesty. In doing so, we closely examined the conceptual overlap between Truthfulness and the four Honesty-Humility facets to examine whether Truthfulness could be assessed as an additional facet of Honesty-Humility or whether Truthfulness-related content should be embedded within the existing facets.

Materials and Methods

Materials

We identified an initial list of adjectives related to Truthfulness (or lack thereof) and the facets of Honesty-Humility as defined in the HEXACO (Ashton & Lee, 2008), starting from previous lexical work (Ashton, Lee, & Goldberg, 2004; Ashton et al., 2006; Caprara & Perugini, 1994; Lee & Ashton, 2008; Romano et al., 2023). Adjectives extracted from previous work on the English language were translated by the authors. Some adjectives were translated into multi-word expressions in order to preserve the original meaning (e.g., "goodhearted" translates as "una persona di buon cuore" - a person with a good heart). We further expanded the initial list inspecting one of the available semantic spaces for the Italian language (WEISS, Marelli, 2017) and dictionaries. We identified an initial list of 120 adjectives potentially related to Honesty-Humility and its facets. The preregistration file mentions 122 adjectives, because two adjectives ("altruista/altruistic" and "degno di fiducia/trustworthy") were included twice due to an oversight in programming the study: We only retained the ratings of their first occurrence.

We further included six items from each HAS scale, excluding Honesty-Humility, balanced by trait pole. The resulting 30 adjectives served as controls, to better identify adjectives that were specifically relevant to Honesty-Humility and not also to other traits. See OSF Supplement for the full list of adjectives.

Procedure

Five independent experts with a record of publication in Personality Psychology received a definition of the four main facets of Honesty-Humility and of Truthfulness. The definition of Honesty-Humility facets were adapted from Lee and Ashton (2008), whereas the definition of Truthfulness was adapted from Miller et al. (2021, pp. 7–8), considering the

tripartite conception of honesty introduced by Cooper and colleagues (2023). The raters indicated the extent to which each of 150 adjectives was representative of the negative vs. the positive pole of different facets of Honesty-Humility plus Truthfulness, on an 11-point scale. For example, in the case of sincerity, the scale ranged between -5 (*[representative of] insincerity*) and +5 (*[representative of] sincerity*). The midpoint 0 was used to indicate that the adjective was *non-representative* [of sincerity]. Study materials are available in the OSF Supplement S1. All studies were implemented in the QualtricsXM platform, and data were analyzed in the R environment (4.2.2; R Core Team, 2022).

Results

First, we assessed inter-rater agreement by computing the intraclass correlations across facets and for each facet separately. The ratings were highly consistent (overall across-facet ICC(2,5) = .93; Truthfulness ICC (2,5) = .96; Sincerity ICC(2,5) = .96; Fairness ICC(2,5) = .91; Greed-Avoidance ICC(2,5) = .88; and Modesty ICC(2,5) = .91).

We averaged judgments across raters and, for ease of interpretation, we normalized them to vary between -1 (the adjective represents the negative pole of the facet) and 1 (the adjective represents the positive pole of the facet). The ratings of adjectives on the four facets showed positive correlations, as it could be expected since they pertained to the same overall dimension of Honesty-Humility (Table 1). Notably, the ratings of Truthfulness and Sincerity showed a nearly perfect correlation of r = .97. This result indicated that Truthfulness and Sincerity could not be distinguished. For this reason, we decided to deviate from the preregistration and merge the ratings of Truthfulness and Sincerity into a single rating, thus focusing on a four-dimensional space: Greed-Avoidance, Modesty, Fairness, and Sincerity, instead of a five-dimensional one separating Truthfulness and Sincerity.

Table 1. Study 1, Correlations among average ratings.

	Greed-Avoidance	Truthfulness	Modesty	Fairness	Sincerity
Greed-Avoidance	1				
Truthfulness	.49***	1			
Modesty	.73***	.32***	1		
Fairness	.78***	.75***	.49***	1	
Sincerity	.56***	**.97***	.36***	.78***	1

Note. * $p < .05$, ** $p < .01$, *** $p < .001$

We projected each adjective into the four-dimensional space and estimated the Euclidean distance of each adjective from the center of the space. For brevity, we call this index Center-distance (C-distance). The C-distance allowed quantifying the extent to which each adjective was generally representative of Honesty-Humility: Adjectives close to the center of the space (with a C-distance near zero) were not representative of any facet of Honesty-Humility, whereas those distant from the center could represent a facet or a combination of facets. We also quantified the Euclidean distance between each adjective and the pole of a facet that was the closest to the adjective. Lower values indicate adjectives that are very representative of the pole of a facet. For brevity, we call this index Facet-distance (F-distance; for a conceptually similar approach, see Gallucci & Perugini 2007). The F- and C-distances of adjectives are represented in Figure 1, exact values are reported in OSF Supplement S2.

Our pre-registered hypothesis was that the C-distance of the 120 adjectives selected to represent Honesty-Humility would be, on average, larger than that of the 30 control adjectives. The hypothesis was confirmed, with average C-distance = .85 (*SD* = .30) for honesty-related adjectives and .10 (*SD* = .12) for controls, as indicated by a Welch's independent samples t-test, $t(119.77) = 21.52$, $p < .001$, to correct for unequal variances.

We selected a subset of adjectives that were the best representative of each of the four facets of Honesty-Humility considered. First, we decided to limit our search to adjectives

whose C-distance was larger than that of each of the 30 control adjectives. We then retained adjectives starting from those with the lowest F-distance to the pole of the facet, but we also kept an eye on breadth of content (i.e., we avoided adjectives that were too close in meaning, if adjectives that captured other nuances of the facet were also available at a relatively short distance, see Table S1 in Supplementary Materials). Six adjectives were retained for each of the facets (three for the positive and three for the negative pole of each facet) and twelve adjectives to represent Sincerity, encompassing Truthfulness (also balanced by pole). Of them, three were part of the HAS, and 27 were new. These adjectives served as the initial set of items for developing the ACH (see Table 2).

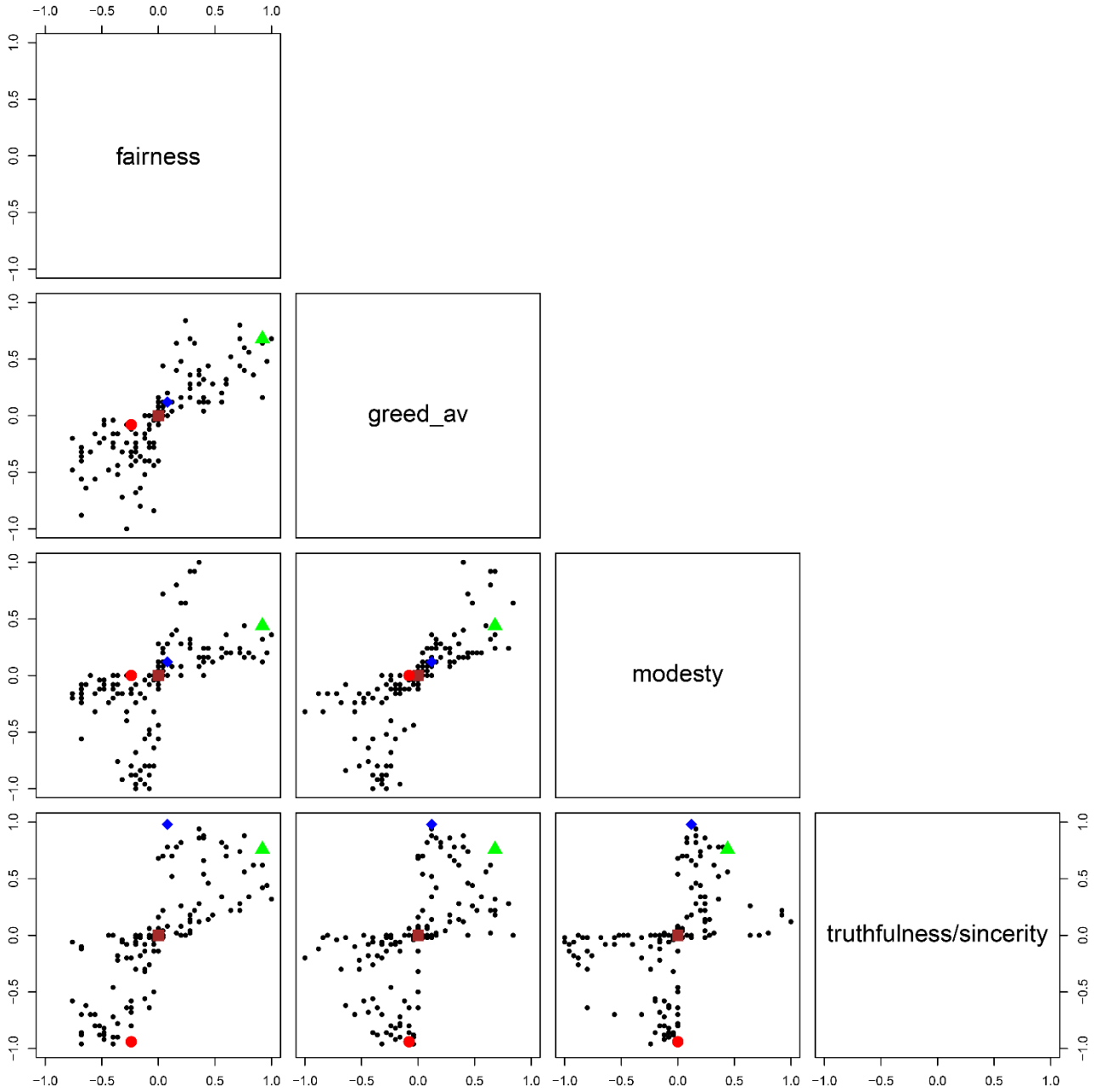


Figure 1. Visualization of the 150 adjectives in the four-dimensional space. Each dot represents a single adjective. The blue diamond is the adjective "vero" (true), which was identified as representative of Sincerity for its low F-distance from positive pole of this facet; the red circle is the adjective "menzognero" (liar), representative of Insincerity for its low F-distance from negative pole of the facet. The green triangle is the adjective "righteous" (retto) representative of many aspects of honesty at once, as indicated by its high C-distance, whereas the brown square is the control adjective "calmo" (calm), which is not representative of any aspect of honesty, as indicated by its null C-distance.

Table 2. *Study 1. List of 30 adjectives to serve as an initial pool for developing the ACH.*

Facet	Selected Adjectives	Notes
Fairness	imparziale (impartial), obiettivo (objective/unbiased), giusto (just/fair)	
Unfairness	ingiusto (unfair), prepotente (bully), immorale (immoral)	
Modesty	sobrio (sober/moderate), modesto (modest), semplice (simple)	
Immodesty	sbruffone (boaster), immodesto (immodest), vanitoso (vain)	
Greed-Avoidance	senza pretese (unpretentious), generoso (generous), non avaro (not stingy)	Since there were only two adjectives that were closer to Greed-Avoidance than to other facets, we selected *avaro (stingy)* from the opposite pole, greed, and preceded it with a negation
Greed	ingordo (greedy), venale (mercenary), avido*(avid)	
Sincerity (incl. Truthfulness)	vero (truthful), diretto (direct), franco (frank/candid), autentico (authentic/genuine), sincero*(sincere), amante della verità (truth-lover)	we decided not to pick *verace (truthful)* and *schietto (candid)*, because they were very close to *vero (truthful)* and *franco (frank/candid)* in content, and *attendibile (reliable)* which was also included in the Adjective Checklist of Conscientiousness (Costantini et al., 2015)
Insincerity (incl. falsehood)	menzognero (lying), in malafede (in bad faith), falso (false), ipocrita*(hypocritical), insincero (insincere), bugiardo (liar)	we decided not to pick *dissimulatore (dissembler)*, *mendace (mendacious)* and *ciarlatano (charlatan)*, because they are rarely used words in spoken Italian

Note. *This adjective is also part of the HEXACO Adjective Scales (Romano et al., 2023)

Discussion

The aim of the study was to select an initial set of items covering the facets of Honesty-Humility. To do so, we collected a list of 120 candidate descriptors of Honesty. Our initial set of adjectives included all Italian adjectives relevant to Honesty-Humility identified in previous lexical work, plus several additional ones. We intentionally broadened the scope of the item pool because our work was guided by the goal of creating a scale that would align as much as possible with the four Honesty-Humility facet labels, Sincerity, Fairness, Modesty, and Greed-Avoidance, while including more clearly content related to Truthfulness. For this reason, in Study 1, we decided to expand the set of descriptors using other sources, such as dictionaries and lexical spaces, and employed independent raters to judge the alignment of each descriptor with each facet. This approach is different compared to previous studies, somehow reversing the approach that starts from factors emerging from psycho-lexical studies and then identifies labels best describing them. By ensuring alignment in meaning between personality descriptors and facet labels, we wanted to assuage jingle-jangle issues (Flake & Fried, 2020; Wulff & Mata, 2025) and ensure that Truthfulness-related content would not be neglected in the process (Fleeson, 2020). We see these two approaches as potentially complementary and both useful to improve assessment in psychology. Our decision to expand the set of descriptors also reflects the fact that language changes over time: Some of the original studies performed on Italian Language (Caprara & Perugini, 1994; Di Blas & Forzi, 1998) are now quite old and may not fully reflect current language use.

Out of the initial set of adjectives, we selected 30 adjectives covering the breadth of content of trait honesty, including both Honesty-Humility facets in the HEXACO model (Ashton & Lee, 2020) and Truthfulness (Fleeson, 2020; Miller et al., 2021). Notably, our data indicated a nearly perfect conceptual overlap between adjectives describing Truthfulness and Sincerity. An inspection of specific adjectives suggested that, of the 120 items examined, no single one could be identified as representative of Truthfulness and not of Sincerity, or

representative of Sincerity and not of Truthfulness. This result suggests that in language use, Truthfulness is virtually indistinguishable from the Sincerity facet of Honesty-Humility. Hence, the lack of conceptual overlap between Honesty-Humility and Truthfulness (Fleeson, 2020; Fleeson et al., 2022) might be due to the operationalization of the trait in the HEXACO-PI-R items, which emphasize other aspects of the Honesty-Humility (see Ashton & Lee, 2020), rather than to the failure to include an additional and specific Truthfulness facet within the assessment. In light of this result, in the process of developing the ACH, we decided to include Truthfulness-related content within facet Sincerity rather than as a separate facet.

Study 2. Development of the Adjective Checklist of Honesty

Study 2 was aimed at further refining a subset of items for the ACH, inspecting their factorial structure and convergence with the HEXACO-PI-R.

Materials and Methods

Participants

Two hundred sixty-six participants (60 men, 206 women; mean age = 31, *SD* =12.5) took part in the study. All participants gave informed consent before starting the survey and took part in the study on a voluntary basis. They were recruited through the university's Sona System. An additional 14 participants were excluded for reporting inattentive responding, as indicated by a Self-Reported Single-Item (Meade & Craig, 2012). The study was run as part of a larger study that included other measures not analyzed here. The final sample size allows detecting a correlation as small as r = .17 with 80% power in a two-tailed test at the conventional α level of .05 (Perugini et al., 2018). In the current study, the main goal was testing the factorial structure of ACH using CFA. One hundred participants are often considered the minimum sample size for CFA, with N = 200 being the 'typical' sample size

(Kline, 2011), albeit samples of at least 300 participants are desirable (Comrey & Lee, 1992). To obtain an estimate of the power for this analysis, we used the PAMLj package (Gallucci, 2026) in Jamovi (The jamovi project, 2025). The package requires specifying a hypothesized factorial structure and one or more parameters of interest to test. Power is then estimated relying on a model comparison between a model that constrains the parameters of interests to zero and a model in which those parameters are freely estimated. Our model included 30 loadings on the general factor, 30 on the specific factors, and six correlations among specific factors. Using a conservative estimate, we assumed all standardized loadings on the general factor to be .2, all loadings on the specific factors .3, and all correlations to be .5. We then compared a model in which one loading for each specific factor was fixed to zero versus a model in which all loadings were freely estimated. With 266 participants, this resulted in 87% power, suggesting that our sample size was adequate for this analysis.

Materials and Procedures

Participants completed the 30 candidate ACH items identified in Study 1. They indicated the extent to which each of the adjectives (e.g., "honest") described themselves, on a scale from 1 (*it does not describe me at all*) to 7 (*it describes me completely*). Cronbach's α reliabilities ranged between .79 and .91. They also completed the HEXACO Adjective Scale (HAS; Romano et al., 2023), an instrument featuring 60 adjectives meant to cover the six main dimensions of personality as reflected in the HEXACO model (see Romano et al., 2023). Three adjectives that were both part of the candidate ACH items and of the HAS (avid, sincere, and hypocritical; see Table 2) were administered only once. Since the HAS and ACH items shared the same format and response options, they were administered as a single questionnaire. Cronbach α reliabilities for the HAS scales ranged between .73 and .91.

Subsequently, participants filled in the 60-item version of the HEXACO-PI-R (Ashton & Lee, 2009) that assesses six major personality traits with 10 items each. The trait Honesty-

Humility was investigated in more detail with the full 32-item HEXACO-PI-R scale, which also assesses the four main facets of the trait (Ashton et al., 2006). Cronbach's α reliabilities ranged between .71 and .90 for traits, and between .76 and .82 for Honesty-Humility facets. Participants indicated their level of agreement with each item (e.g., "I would be quite bored by a visit to an art gallery"), on a scale from 1 (*strongly disagree*) to 5 (*strongly agree*).

At the end of the survey, participants answered to a Self-Report Single Item investigating to which they could either report inattentive responding, or that their data should be used in the study (Meade & Craig, 2012). Participants who reported careless responding were excluded from the analyses.

Results

We first examined the factorial structure of the 30 candidate ACH items. To this aim, a CFA analysis was run in R using the *lavaan* package (Rosseel, 2012). Mirroring the HAS structure (Romano et al., 2023), we fitted a bifactor model (Model 1) in which all items loaded on the facet that was the closest to them in Study 1, as well as on a general factor that encapsulated response factors[2]. The model fit was acceptable, and superior to the fit of three alternative plausible models (see Table 3), namely: A model with four correlated facets, that did not include the general response factor (Model 2); a hierarchical-factor model with four facets and a superordinate Honesty-Humility factor (Model 3, $\Delta\chi^2 = 400.4, \Delta df = 32, p < .001$); and a single-factor model, in which all items loaded on the same Honesty-Humility factor without a specified facet structure (Model 4, $\Delta\chi^2 = 640.1, \Delta df = 3, p < .001$).

[2] Though bifactor models usually assume the orthogonality of specific factors (Markon, 2019), several studies allow them to correlate for theoretical or empirical reasons (e.g., Anglim et al., 2017; Hörz-Sagstetter et al., 2021; Sharp et al., 2015). In our study, specific factors were correlated following the factor solution defined for HAS in a previous study (Romano et al., 2023).

Table 3. *Study 2. CFA model comparison.*

Model	χ^2	*df*	*p*	cfi	rmsea	srmr	AIC	BIC	$\Delta\chi^2$	Δdf	*p*
1	843.5	369	< .001	0.799	0.070	0.065	23856.8	24200.8	-	-	-
2	1239.8	399	< .001	0.644	0.089	0.092	24193.1	24429.6	396.3	30	< .001
3	1243.9	401	< .001	0.643	0.089	0.093	24193.2	24422.6	4.1	2	.127
4	1483.6	405	< .001	0.543	0.100	0.102	24424.9	24639.9	239.7	4	< .001

The fit of Model 1 was acceptable but not optimal. We hence decided to retain for each scale only the four best items, based on modification indices, but also on semantic considerations (e.g. removing polysemous items which can be misinterpreted out of a sentence context like "sobrio" – sober), and keeping in sight the breadth of the content of each facet. Removed items were: "franco" (frank) and "in malafede" (in bad faith) for Sincerity; "prepotente" (bully) and "imparziale" (impartial) for Fairness; "immodesto" (immodest) and "sobrio" (sober) for Modesty; "senza pretese" (unpretentious) and "ingordo" (greedy) for Greed-Avoidance. A total of 22 adjectives were hence selected for the final ACH. After removing these items, we refitted the model. Model fit improved substantially (CFI=0.928; RMSEA=0.050; SRMR = .051; $\chi2(181) = 300.643$, $p < .001$; see the OSF Supplement S3 and Figure 2 for individual loadings).

Table 4 reports correlations with the HAS and HEXACO-PI-R scales. The ACH facets significantly correlated with the Honesty-Humility scale of the HAS (after removing three overlapping items in the case of the HAS) and of the HEXACO-PI-R. However, some substantial correlations emerged with Conscientiousness, with Fairness and Sincerity showing stronger correlations with HEXACO-PI-R Conscientiousness than with Honesty-Humility and Fairness showing a similarly sized correlation with HAS Honesty-Humility and Conscientiousness. Some correlations emerged between HEXACO-PI-R Honesty-Humility

facets scales and Conscientiousness, albeit they were generally weaker than those with Honesty-Humility. At the facet level, ACH Fairness showed a similar correlation with HEXACO-PI-R Sincerity and Fairness, and ACH Greed-Avoidance showed a stronger correlation with HEXACO-PI-R Sincerity and Fairness than with Greed-Avoidance.

Table 4. *Study 2. Correlations between ACH and HEXACO-PI-R Honesty-Humility facets scores (columns) and HEXACO-PI-R / HAS scales (rows).*

	Adjective Checklist of Honesty (ACH)				HEXACO-PI-R Honesty-Humility facets			
	SI	FA	MO	GR	SI	FA	MO	GR
HAS traits								
Honesty-Humility[1]	.57***	.43***	.54***	.47***	.47***	.42***	.34***	.31***
Emotionality	.03	-.03	.23***	.16**	.00	.01	.20**	-.02
eXtraversion	.23***	.15*	-.01	.21***	.07	.09	-.02	.00
Agreeableness vs. Anger	.09	.16*	.37***	.27***	.15*	.18**	.23***	.16**
Conscientiousness	.35***	.38***	.18**	.13*	.20**	.30***	.09	.13*
Openness to experience	.20**	.29***	.03	.13*	.01	.06	-.10	-.09
HEXACO-PI-R traits								
Honesty-Humility[2]	.30***	.26***	.33***	.35***	.60***	.65***	.61***	.63***
Emotionality	.12+	.09	.21***	.17**	.08	.14*	.24***	.09
eXtraversion	.21***	.23***	-.06	.17**	.01	.02	-.12*	.00
Agreeableness vs. Anger	.00	.08	.24***	.21***	.12*	.15*	.25***	.21***
Conscientiousness	.36***	.37***	.16**	.15*	.27***	.34***	.16**	.17**
Openness to experience	.11+	.22***	.01	.08	.04	.10	.04	.02
HEXACO-PI-R Honesty-Humility facets								
Sincerity	.39***	.31***	.21***	.33***	1	.58***	.38***	.50***
Fairness	.36***	.30***	.25***	.31***	.58***	1	.46***	.53***
Modesty	.10+	.05	.36***	.20***	.38***	.46***	1	.49***
Greed-Avoidance	.11+	.16**	.21***	.27***	.50***	.53***	.49***	1

Note. SI = Sincerity, FA = Fairness, MO = Modesty, GR = Greed-Avoidance

[1]For the purpose of estimating correlations of ACH facets with the HAS, we removed from the HAS Honesty-Humility scale the items that were also present in the ACH, depending on the facet. In particular, we removed the

item *avido/avid* to estimate the correlation with Greed-Avoidance, and items *sincero/sincere* and *ipocrita/*hypocritical to estimate the correlation with Sincerity. For correlation with Fairness and Modesty the full HAS Honesty-Humility scale was used.
[2]For the purpose of estimating correlations of HEXACO-PI-R Honesty-Humility facets with the HEXACO-PI-R traits, we removed from the HEXACO-PI-R Honesty-Humility scale the items corresponding to each facet. For example, to estimate the correlation with the Sincerity facet, all the Sincerity items were removed from the Honesty-Humility scale of HEXACO-PI-R.
* $p < .05$, ** $p < .01$, *** $p < .001$

Discussion

In this study we refined the set of the ACH items, developing the final 22-item scale, inspected its factorial structure, and examined its convergence with the HAS and HEXACO-PI-R. Confirmatory factor analysis showed that the attribution of adjectives to the four facets based on expert raters' scores of Study 1 could be recovered in the scale's structure. Whereas the ACH facet converged with Honesty-Humility as expected, sizable correlations with Conscientiousness emerged for facets Sincerity and Fairness. We followed up this result by examining correlations between ACH facets and HEXACO-60 Conscientiousness items (Ashton & Lee, 2009). The results are reported in the OSF Supplement S4 and show that the most substantial correlations were with items of Diligence and Perfectionism, and particularly with items "I do only the minimum amount of work needed to get by", "I often push myself very hard when trying to achieve a goal", and "I always try to be accurate in my work, even at the expense of time". Interestingly, these items include aspects related to workplace ethics, which are conceptually and empirically related to trait Honesty-Humility (e.g., Pletzer et al., 2019).

Although ACH scales showed positive correlations with respect to the HEXACO-PI-R Honesty-Humility facet scales, a clear pattern of convergent and discriminant validity did not emerge, particularly with respect to Greed-Avoidance. We argue that this may be related to the specific content of the HEXACO-PI-R items, which is not always in line with the facet labels, but covers a broader range of content (see Fleeson, 2020). For example, Fairness items mostly

portray situations in which a person is not willing to cheat or break rules for an advantage (e.g., "I would be tempted to buy stolen property if I were financially tight"), whereas our adjective items concern fairness in more general terms, not necessarily connected to getting or giving up a personal advantage. HEXACO-PI-R Sincerity items mostly concern pretending to like someone in order to get something from them (e.g., "*I see nothing wrong with flattering people to get ahead in life*"), whereas ACH Sincerity adjectives assess a tendency to fostering truth in others and be authentic, without implying specific motives for doing so. Greed-Avoidance items are related to not-wanting luxury (e.g. "*I would like to live in an expensive and exclusive area*") and desire for admiration (e.g., "*I would like to be seen driving around in a very expensive car*"), while our adjectives emphasize a dichotomy between being greedy or generous, without including love for luxury or admiration. HEXACO-Modesty items mostly refer to seeing oneself or being seen as better than others (e.g., "*I am a normal person, I am not better than others*", "*I want people to know that I am someone of some importance*"). In this case, the content is closer with ACH than for other facets, and this resulted in higher convergence between ACH and HEXACO-PI-R Modesty scales. Although some ACH facets do not converge perfectly with corresponding HEXACO-PI-R facets, we chose to retain the original facet names because the initial descriptor selection and judgments by expert raters in Study 1 were explicitly guided by the alignment with HEXACO-PI-R facet labels. Hence, the ACH facet scales can be seen as slightly different but complementary operationalizations of the same facets.

Overall, we have shown that the Adjective Checklist of Honesty is able to capture the information contained in HEXACO Honesty-Humility; in doing so, it also captures Honesty-Humility facets on a more general level than what is captured by HEXACO-PI items. The final selection of ACH items is presented in Table 5.

Table 5. *Final selection of ACH adjectives.*

Facet	Selected Adjectives
Fairness	obiettivo (objective/unbiased), giusto (just/fair)
Unfairness	ingiusto (unjust/unfair), immorale (immoral)
Modesty	modesto (modest), semplice (simple)
Immodesty	sbruffone (boaster), vanitoso (vain)
Greed-Avoidance	generoso (generous), non avaro (not stingy)
Greed	venale (mercenary), avido* (greedy)
Sincerity	vero (truthful), diretto (direct), autentico (authentic), sincero* (sincere), amante della verità (truth-lover)
Insincerity	menzognero (lying), falso (false), ipocrita* (hypocritical), insincero (insincere), bugiardo (liar)

Note. *This adjective is also part of the HEXACO Adjective Scales (Romano et al., 2023)

Study 3 - Validity of the ACH

The aim of this study was twofold: first, we wanted to provide converging evidence of the scale structure by administering the final version of the ACH to a new independent sample; second, we wanted to assess the relationship between the ACH and a series of validated measures of constructs conceptually related to each subscale.

Participants

N = 300 participants (*M* age = 31.5; SD = 9.98), all Italian native speakers, participated in the study through the Prolific Academic platform. Of them, 148 identified as men, 145

identified as women, and 7 identified as “other”. Participants received written informed consent and were compensated with £3.75/each for a 30-minute task.

In order to identify highly inattentive responses, the HAS and the HEXACO-PI, and the Modest Behavior Scale each included a directed question (Maniaci & Rogge, 2014), an item that was visually presented in a similar format to other items, but asked for a specific response (e.g., “This is a verification question, answer four and proceed with the questionnaire”). Ten participants who failed two or more directed questions were immediately expelled from the questionnaire and did not receive compensation. To achieve the pre-registered sample size of 300 participants, new participants were invited to replace excluded ones. The final sample size allows detecting a correlation as small as $r = .16$ with 80% power in a two-tailed test, at the conventional α level of .05.

Materials and Procedure

Personality traits were assessed through the same set of questionnaires used in Study 2. Participants completed the ACH (Cronbach’s α between .65 and .89) and the HAS (Cronbach’s α between .81 and .92; Romano et al., 2023). They subsequently completed the HEXACO-60 (Ashton et al., 2009), with the Honesty-Humility trait being investigated with the full 32-item HEXACO-PI-R scale (Lee & Ashton, 2004; α between .77 and .91 for traits and between .79 and .84 for honesty facets). In addition, participants completed a battery of measures exploring constructs that we expected to overlap in content with each facet of Honesty-Humility, which were administered after the personality scales, in a random order.

For *Sincerity*, we employed the Revised Lie Acceptability Scale (Oliveira & Levine, 2008; α = .84) and the Authenticity Scale (A. M. Wood et al., 2008; α = .89). The Revised Lie Acceptability Scale comprised 11 items, wherein participants indicated their level of agreement (on a scale from 1 = *strongly disagree* to 7 = *strongly agree*) with statements such as "Lying is

no big deal." The Authenticity Scale consisted of 12 items, requiring participants to indicate the extent to which each statement (e.g., "I am true to myself in most situations") described them, using a scale from 1 (*it does not describe me at all*) to 7 (*it describes me very well*). The Authenticity Scale encompassed three subscales: Authentic Living (α = .73; "I think it is better to be yourself than to be popular"), Self-Alienation (α = .87; "I feel as if I don't know myself very well"), and Accepting External Influence (α = .90; "I am strongly influenced by the opinions of others").

For *Fairness*, we employed the Justice Sensitivity Short Scale (Baumert et al., 2014), which captured Justice Sensitivity from different perspectives (Schmitt et al., 2010). Each subscale included two items: Perpetrator ("I feel guilty when I enrich myself at the cost of others", α = .84), Observer (α = .80; "I am upset when someone is undeservingly worse off than others"), Victim (α = .72; "It makes me angry when others are undeservingly better off than me"), and Beneficiary (α = .69; "I feel guilty when I am better off than others for no reason"). Participants indicated their level of agreement with each statement, using a 6-point scale, ranging from 1 (*strongly disagree*) to 6 (s*trongly agree*), reflecting how they would respond in unfair situations.

For *Modesty*, we employed the Modest Behavior Scale (Xiaohua Chen et al., 2009; α = .83), which consisted of 39 items capturing a range of modest behaviors. Examples of these behaviors included attributing success to luck rather than personal ability in front of others and admitting and rectifying mistakes after recognizing them. Participants responded to these items using a 5-point scale, anchored from 1 (*strongly disagree*) to 5 (*strongly agree*). The Modest Behavior Scale encompassed three distinct subscales: Self Effacement (α = .70; "Talking myself down to downplay my talent"), Other Enhancement (α = .83; "Usually praise other people"), and Avoid Attention Seeking (α = .81; "Not praise myself in an attention-getting

way"). Additionally, we assessed Entitlement via the Psychological Entitlement Scale (Campbell et al., 2004; α = .89), in which participants indicated their level of agreement with nine statements such as "I honestly feel I'm just more deserving than others." Participants rated their agreement on a scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Lastly, for *Greed-Avoidance*, we included two scales: the Vices and Virtue Scale (VAVS) - Greed subscale (Veselka et al., 2014; α = .78) and the Dispositional Greed Scale (Krekels & Pandelaere, 2015; α = .84). In the VAVS, participants rated their agreement level with 6 items, such as "I enjoy being a part of exclusive clubs or groups that are not open to everyone," using a 5-point scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Similarly, the Dispositional Greed Scale comprised 6 items, including statements like "No matter how much I have of something, I always want more," where participants indicated their agreement level on a scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Furthermore, we also included in the battery the Dirty-Dozen scale (Jonason & Webster, 2010; Italian version by Chiorri et al., 2019) to capture the dark triad, a cluster of three personality traits associated with exploitative behaviors. The scale consisted of 12 items, organized into three subscales: Machiavellianism (α = .87; "I tend to manipulate others to get my way"), Psychopathy (α = .82; "I tend to be unconcerned with the morality of my actions"), and Narcissism (α = .80; "I tend to want others to admire me"). Participants were asked to indicate their agreement level with a series of statements on a scale ranging from 1 to 7.

To ensure linguistic equivalence, scales for which an Italian version was not available were translated into Italian and back-translated by the authors. The back-translation was cross-checked with the original version by a native English speaker for accuracy and consistency. Scale and facet scores were computed as the mean of the corresponding items.

Results

Replication of the factorial structure

Since in Study 2 we used the same sample for selection of items and for selective analyses (Kriegeskorte et al., 2009), the first aim of Study 3 was to replicate the factorial structure identified in Study 2 in a new pre-registered one. The model confirmed its good fit, $\chi^2(181) = 343.415$, $p < .001$; CFI = .941, RMSEA = .054, SRMR = .047; see the OSF supplement S3 and Figure 2 for individual loadings).

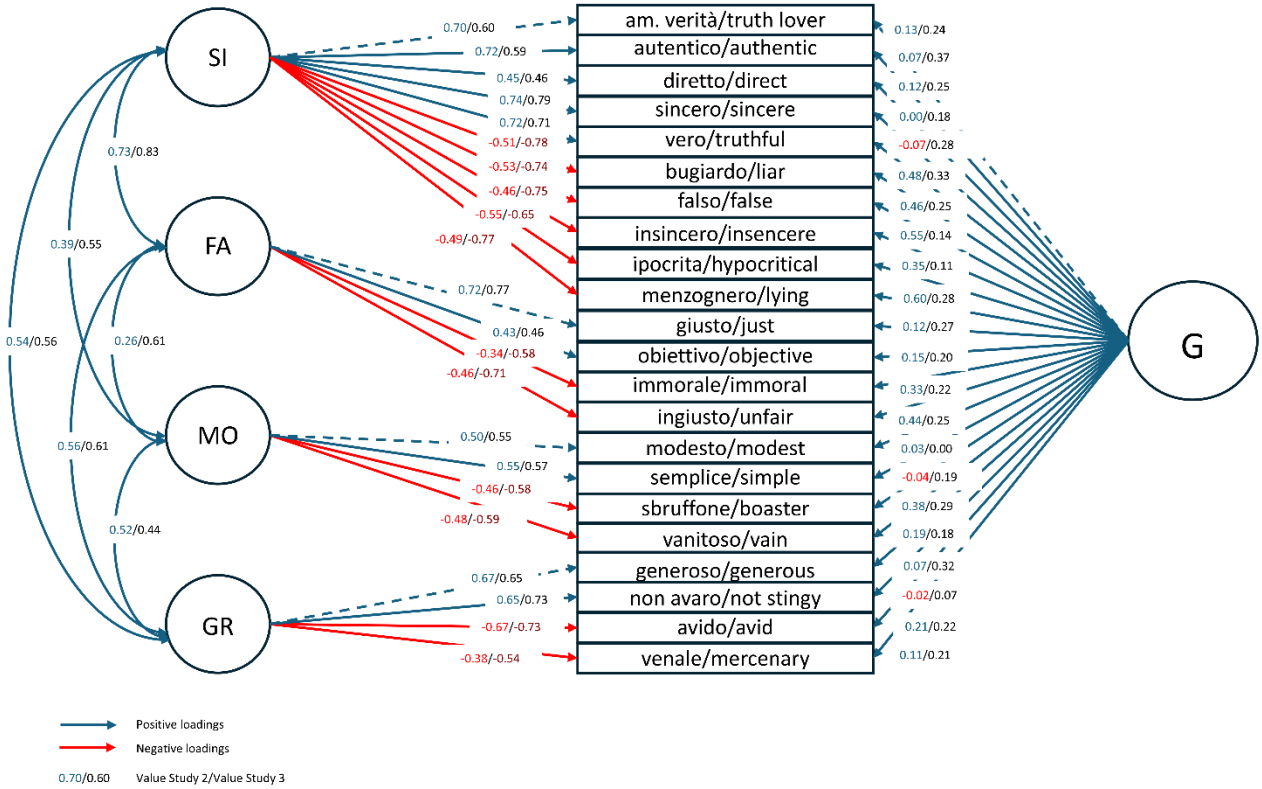


Figure 2. Standardized model parameters from the bifactor models in Studies 2 and Study 3. SI = Truthfulness-Sincerity, FA = Fairness, MO = Modesty, GR = Greed-avoidance, G = General factor.

Correlations with Honesty-Humility Scales

Table 6 reports the correlations between the ACH and the HAS and Honesty-Humility subscales. The overall pattern of correlations mirrored the one found in Study 2. In particular, whereas ACH scales correlated with HAS and HEXACO-PI-R Honesty-Humility scales, substantial correlations between Conscientiousness and facets Sincerity and Fairness emerged, which, in the case of HEXACO-PI-R, were even stronger than those with Honesty-Humility. The pattern of correlations at the item-level was similar to that emerged in Study 2 (see Table S3 in the OSF supplement). Positive correlations were found between all ACH facets and their counterparts assessed with the HEXACO-PI-R. It is worth noting, however, that ACH Greed-Avoidance showed higher correlations with HEXACO-PI-R Sincerity (r = .36, p < .001), Fairness (r = .43, p < .001), and Modesty (r =. 24, p < .001), than with HEXACO-PI-R Greed-Avoidance (r = .15, p < .01).

Table 6. *Study 3. Correlations between ACH scores and HEXACO-PI-R Honesty-Humility facets scores (columns), and HEXACO-PI-R / HAS scales (rows).*

	Adjective Checklist of Honesty (ACH)				HEXACO-PI-R Honesty-Humility facets			
	SI	FA	MO	GR	SI	FA	MO	GR
	HAS traits							
Honesty-Humility[1]	.68***	.70***	.64***	.53***	.40***	.48***	.31***	.21***
Emotionality	-.28***	-.16**	.02	-.07	-.16**	.02	.14*	.05
eXtraversion	.29***	.18**	-.15**	.28***	.05	.06	-.20***	-.16**
Agreeableness vs. Anger	.35***	.45***	.48***	.33***	.31***	.30***	.29***	.18**
Conscientiousness	.51***	.52***	.32***	.27***	.16**	.22***	-.03	.00
Openness to experience	.29***	.23***	-.13*	.24***	.22***	.16**	-.03	.10
	HEXACO-PI-R traits							
Honesty-Humility[2]	.31***	.33***	.34***	.39***	.57***	.52***	.64***	.54***
Emotionality	-.03	.03	.05	.10	-.07	.13*	.00	-.05
eXtraversion	.35***	.22***	-.08	.26***	.14*	.09	-.22***	-.13*
Agreeableness vs. Anger	.15**	.24***	.29***	.29***	.20***	.19***	.28***	.17**
Conscientiousness	.45***	.44***	.27***	.21***	.18**	.25***	.03	.03
Openness to experience	.23***	.16**	-.05	.18**	.28***	.20***	.07	.10
	HEXACO-PI-R Honesty-Humility facets							
Sincerity	.43***	.32***	.20***	.36***	1	.50***	.39***	.44***
Fairness	.34***	.40***	.31***	.43***	.50***	1	.42***	.32***
Modesty	.09	.18**	.36***	.24***	.39***	.42***	1	.53***
Greed-Avoidance	.08	.11*	.16**	.15**	.44***	.32***	.53***	1

Note. SI = Sincerity, FA = Fairness, MO = Modesty, GR = Greed-Avoidance

[1]For the purpose of estimating correlations of ACH facets with the HAS, we removed from the HAS Honesty-Humility scale the items that were also present in the ACH, depending on the facet. In particular, we removed the item *avido/avid* to estimate the correlation with Greed-Avoidance, and items *sincero/sincere* and *ipocrita/*hypocritical to estimate the correlation with Sincerity. For correlation with Fairness and Modesty the full HAS Honesty-Humility scale was used.

[2]For the purpose of estimating correlations of HEXACO-PI-R Honesty-Humility facets with the HEXACO-PI-R traits, we removed from the HEXACO-PI-R Honesty-Humility scale the items corresponding to each facet. For example, to estimate the correlation with the Sincerity facet, all the Sincerity items were removed from the Honesty-Humility scale of HEXACO-PI-R.

* $p < .05$, ** $p < .01$, *** $p < .001$

Correlations with other constructs

The main aim of Study 3 was to assess the relationship between the ACH subscales and other validated measures capturing similar constructs, with the pre-registered hypothesis that each measure would correlate with the corresponding ACH scale more than with the other scales. Correlations corresponding to preregistered hypotheses are bolded in Table 7. In addition, we explored whether this pattern was similar or different from the one found with the HEXACO-PI Honesty-Humility facets. Finally, we examined the connections between Honesty-Humility facets and the Dark Triad traits.

As hypothesized, the Authenticity scale, its facets, and the Lie Acceptability scales showed the highest correlations with ACH Sincerity. The same pattern was observed for HEXACO-PI-R facets, although correlations were generally larger in size for Sincerity assessed with ACH than HEXACO-PI-R.

Perpetrator sensitivity, the aspect of Justice Sensitivity more clearly related to Honesty-Humility in previous studies (Baumert & Schmitt, 2016), showed positive correlations with ACH Fairness, as did Observer sensitivity. In both cases, similarly sized or even larger correlations emerged with other ACH facets, particularly Modesty and Greed-Avoidance. No significant correlations emerged between ACH Fairness and Victim or Beneficiary sensitivities. Victim sensitivity was higher for greedy individuals (with low Greed-Avoidance), whereas Beneficiary sensitivity showed only weak correlations with Modesty and Greed-Avoidance. HEXACO-PI-R Fairness showed stronger correlations with Perpetrator, Observer, and Victim sensitivity, whereas consistent with the ACH, no correlation emerged between Fairness and Beneficiary sensitivity.

The Modest Behavior scale correlated most with ACH Modesty. Considering its subscales, however, it is noteworthy that this relationship was much stronger for the Avoid Attention Seeking scale than for the other ones. A similar pattern emerged for HEXACO-PI-R Modesty. As hypothesized, Entitlement correlated most with Modesty, but this correlation was

weak and similar in size to that observed for Greed-Avoidance. Conversely, Entitlement seemed to reflect more central elements of HEXACO-PI-R Modesty and of HEXACO-PI-R Honesty-Humility in general.

As expected, the highest correlations with both measures of Greed (VAVS and Dispositional Greed) emerged with ACH Greed-Avoidance. A similar pattern emerged for the HEXACO-PI-R facets, albeit the correlations were generally more sizable with HEXACO-PI-R than with ACH scales.

The Dark Triad traits showed substantial negative correlations with all ACH subscales. However, each ACH facet showed a slightly different pattern of correlations with these traits. Sincerity correlated mostly with Machiavellianism, the tendency to manipulate others and to prefer a pragmatic approach to situations. Fairness and Greed-Avoidance showed a stronger association with Psychopathy, which is characterized by impulsive behaviors, lack of empathy and cooperation, and a general absence of moral qualities and remorse. Modesty showed the highest correlation with Narcissism, which is characterized by entitlement and an enhanced sense of self. A similar pattern of correlations emerged also with the HEXACO-PI-R subscales, with a noteworthy difference: HEXACO-PI-R Greed-Avoidance seemed to be more clearly characterized by (lack of) Narcissism and not by (lack of) Psychopathy, whereas the opposite was true for ACH Greed-Avoidance.

Table 7. *Correlations between ACH and HEXACO-PI-R Honesty-Humility facets and other constructs.*

	Adjective Checklist of Honesty (ACH)				HEXACO-PI-R			
	SI	FA	MO	GR	SI	FA	MO	GR
Authenticity	**.53*****	.46***	.14*	.24***	.39***	.24***	.06	.18**
Authentic Living	**.53*****	.49***	.26***	.30***	.37***	.28***	.17**	.21***
Self Alienation	**-.41*****	-.38***	-.05	-.17**	-.26***	-.19***	.03	-.05
Accepting external influence	**-.40*****	-.30***	-.09	-.18**	-.34***	-.16**	-.06	-.21***
Lies Acceptability	**-.49*****	-.38***	-.20***	-.32***	-.44***	-.50***	-.25***	-.16**
JS -Perpetrator	.14*	**.21*****	.22***	.19***	.35***	.49***	.30***	.24***
JS -Observer	.17**	**.16****	.15**	.23***	.09	.19***	.09	.06
JS - Victim	-.15**	**-.11**	.02	-.20***	-.28***	-.27***	-.22***	-.25***
JS -Beneficiary	-.08	**-.07**	.13*	.11*	.08	.11	.12*	.00
Modest Behavior	.23***	.30***	**.44*****	.24***	.25***	.32***	.50***	.30***
Self Effacement	-.12*	-.06	**.16****	.02	-.11	-.02	.19**	.06
Other Enhancement	.45***	.48***	**.24*****	.30***	.30***	.36***	.23***	.09
Avoid Attention Seeking	.20***	.26***	**.51*****	.22***	.37***	.36***	.61***	.46***
Entitlement	.06	-.04	**-.14***	-.13*	-.34***	-.38***	-.69***	-.43***
VAVS-Greed	-.21***	-.24***	-.18**	**-.31*****	-.51***	-.49***	-.57***	-.68***
Dispositional Greed	-.13*	-.15**	-.16**	**-.19****	-.33***	-.23***	-.36***	-.52***
Machiavellianism	-.51***	-.40***	-.29***	-.38***	-.68***	-.46***	-.46***	-.39***
Psychopathy	-.37***	-.48***	-.32***	-.50***	-.41***	-.53***	-.36***	-.24***
Narcissism	-.20***	-.22***	-.40***	-.27***	-.41***	-.30***	-.54***	-.55***

Note. SI = Sincerity, FA = Fairness, MO = Modesty, GR = Greed-Avoidance, JS = Justice Sensitivity, VAVS = Vices And Virtue Scale.
Correlations corresponding to preregistered hypotheses are marked in bold.
* $p < .05$, ** $p < .01$, *** $p < .001$

Discussion

The aim of the current study was twofold: First, we wanted to confirm the factorial structure of the ACH in a preregistered study, by administering the scale to a new sample; second, we intended to demonstrate the convergent and discriminant validity of the scale, showing that its four subscales could be consistently associated with other validated measures of similar constructs. Study 3 corroborated the good fit of the bifactor model identified in Study 2. All four ACH scales correlated with their respective counterparts in the HEXACO-PI-R, and ACH subscales generally aligned with the hypothesized related constructs. However, the strength of these associations varied.

ACH Sincerity aligned with its HEXACO-PI-R counterparts. Interestingly, this scale showed generally stronger correlations with Authenticity and Lies Acceptability than the HEXACO-PI-R, suggesting that truthfulness and authenticity might be more central in the operationalization of Sincerity within the ACH than within the HEXACO-PI-R (Fleeson, 2020; Fleeson et al., 2022; Miller et al., 2021).

Fairness aligned with the corresponding HEXACO-PI-R scale, but, unexpectedly, correlated with Justice Sensitivity only from the perspectives of the Perpetrator and of the Observer, but not of the Beneficiary or the Victim. These findings suggest that ACH Fairness can capture some nuances of Justice Sensitivity, but not the whole construct. In particular, individuals with high ACH Fairness could be particularly concerned with preventing others from being treated unfairly, especially when they themselves are in the position of perpetrating such injustice, in line with the idea that Honesty-Humility reflects proactive cooperation (Ashton & Lee, 2007; Hilbig et al., 2013). A similar pattern emerged for HEXACO-PI-R Fairness, which, however, correlated significantly with Victim sensitivity, reflecting the sensitivity to injustice when suffering a personal disadvantage, corroborating the idea that

HEXACO-PI-R Fairness might be different from ACH Fairness because the former makes a clearer reference to personal advantages or disadvantages than the latter.

In line with our predictions, within the ACH, Modesty was the facet most correlated with the Modest Behavior scale, particularly so with the Avoid Attention Seeking subscale. A correlation between ACH Modesty and entitlement emerged, but this relationship was more sizeable for HEXACO-PI-R Modesty. In fact, Entitlement was quite strongly correlated with all HEXACO-PI-R facets, whereas ACH facets showed much weaker relationships with this construct. This suggests that entitlement might be particularly central for the HEXACO-PI-R operationalization of Honesty-Humility, but it might not be as central for other measures of the same construct (see also Thielmann et al., 2017).

In line with Study 2, among ACH facets, Greed-Avoidance showed the weakest convergent and discriminant validity with respect to HEXACO-PI-R facet scales. This distinction was further reflected in how the ACH and HEXACO Greed-Avoidance were related to other constructs. Whereas ACH Greed-Avoidance aligned with other measures of Greed, HEXACO-PI-R Greed-Avoidance was more clearly related to these constructs. In addition, HEXACO-PI-R Greed-Avoidance was more strongly connected to Narcissism, consistent with previous studies (Aghababaei et al., 2014), further corroborating the idea that this scale may reflect desire for luxury and admiration, rather than just a dichotomy between generosity and greed. Conversely, ACH Greed-Avoidance was more closely associated with psychopathy, a trait that reflects a lack of empathy and a lack of cooperative behaviors, particularly in situations where this could lead to an immediate payoff (Balafoutas et al., 2021; Gunschera et al., 2022). Adjectives in the negative pole of the ACH Greed-Avoidance subscale ("greedy" and "mercenary") may indeed reflect a broader antagonistic tendency (see also Sekhar et al., 2020).

General discussion

Honesty is crucial for the functioning of society, as it is closely related to prosocial and moral behaviors and attitudes (Ashton & Lee, 2008). As such, studying honesty as both a moral virtue and a personality trait is relevant for understanding and predicting socially important behaviors (Miller et al., 2021). However, assessing honesty is complex as it requires considering its multiple facets beyond its narrowed understanding as the mere absence of lying and cheating behavior (Miller et al., 2021). The first aim of this work was to create and validate an adjective checklist able to measure both the Honesty-Humility domain and its facets, as was done in previous studies for other traits (Costantini et al., 2015; Costantini & Perugini, 2016). Romano et al. (2023) have proposed a comprehensive adjective scale to measure, through lexical descriptors, the main personality traits reflected in the HEXACO model. The resulting HEXACO Adjective Scales (HAS) offer an efficient and effective way to assess broad personality traits. Nevertheless, its current version does not account for the full complexity of the Honesty-Humility HEXACO domain and does not include adjectives that differentiate its facets. For this reason, we have developed a new adjective checklist, which captures the four main Honesty-Humility facets: Sincerity, Fairness, Modesty, and Greed-Avoidance. The ACH offers a comprehensive and efficient assessment of Honesty-Humility through adjectives, complementing existing proposals (Ashton, Lee, & Goldberg, 2004; Romano et al., 2023; Thielmann et al., 2017) and emphasizing Truthfulness-related aspects (Fleeson, 2020; Fleeson et al., 2022). The full content of the scale is reported in the Appendix.

Adjectives-based measures rely on the lexical hypothesis (Goldberg, 1993, 2002; D. Wood, 2015) and, when compared to phrased items, present both advantages and disadvantages. Measures based on phrased items are more contextualized than adjectives, and for this reason, their interpretation can be clearer (Condon et al., 2020). On the other hand, adjectives-based measures are faster to administer, can be more easily included in intensive

longitudinal studies (e.g., Di Sarno et al., 2023; Fleeson, 2001; Ringwald et al., 2022), and they can be used as the basis to implement reaction-time-based procedures (Back et al., 2009; Costantini et al., 2015, 2016; but see Yovel & Friedman, 2013). For these reasons, we see the ACH as a complement to existing measures of Honesty-Humility that can be useful in situations where brief and decontextualized items are desirable.

The factorial validity of the ACH was confirmed in two independent studies. Furthermore, the four ACH scales generally converged with the hypothesized constructs and often showed a stronger relationship with the constructs than other facets, as pre-registered. The nomological network of Honesty-Humility as a trait is theoretically robust (Ashton & Lee, 2007, 2020), and it has been corroborated by a large body of empirical studies (e.g., Zettler et al., 2020). Our studies further contribute to clarifying the nomological network of Honesty-Humility at a level that has been underinvestigated, the level of its facets. The results observed suggest that some connections with conceptually related constructs may depend on the specific operationalization of Honesty-Humility facets. In fact, the HEXACO-PI-R and the ACH facet scales emphasize different aspects of the Honesty-Humility, with constructs such as psychological entitlement and narcissistic tendencies, encompassing, for example, love for luxury, being more central for HEXACO-PI items (see Ashton & Lee, 2020) and Truthfulness and theoretically related constructs, such as authenticity, being closer to the core of the ACH. We additionally explored the possibility that the tendency to respond at the extremes of the scales in adjective vs. phrased items could be responsible for some of the differences emerged in the patterns of correlations characterizing the ACH and the HEXACO-PI-R[3]. The results of this analysis are reported in the OFS Supplement S5 and did not suggest meaningful differences between the ACH and the HEXACO-PI-R in terms of response extremity. Given these distinctions and potential measurement considerations, future research would benefit from

[3] We would like to thank an anonymous reviewer for this suggestion.

further corroborating the construct validity of the scale, for example, by examining self-observer agreement for those instruments. Past research focusing on the HEXACO-PI found somewhat attenuated self-observer convergence for the Honesty-Humility and Agreeableness vs. Anger dimensions relative to the remaining four factors (Ashton & Lee, 2010). In addition, previous research on the HAS has documented notably lower self-observer agreement specifically for its Honesty-Humility scale compared to other dimensional scales within the same instrument (Romano et al., 2023). These discrepancies suggest the need to conduct a larger study with self- and observer reports for the ACH, HAS, and HEXACO instruments. Looking at the relationships between the ACH and the HEXACO-PI-R traits, we observed correlations between the ACH facet scales and HEXACO-PI-R Conscientiousness, which were in some cases even stronger than the correlations observed with HEXACO-PI-R Honesty-Humility. Notably, this pattern of relationships was much less marked when Conscientiousness was assessed with an adjective measure, the HAS. The nature of the overlap between traits Honesty-Humility and Conscientiousness is currently debated, with some ascribing them to the presence of common superordinate factors (e.g., Catano et al., 2018) and others assuming that some indicators result from a blend of different traits (Ashton et al., 2009). In previous studies, such a correlation involved mainly facet Fairness (Ashton & Lee, 2020), whereas in our data, a similar pattern extended to facet Sincerity. These results are in line with past research showing that some aspects of moral behavior fall within Conscientiousness, particularly those reflecting obligations towards other individuals and society in general (Roberts et al., 2005). This is reflected in conceptual overlaps between the Conscientiousness and the Honesty-Humility domains, with the former encompassing task-related conscience and the latter moral conscience. These overlaps become evident in workplace moral behavior, which both traits substantially contribute to predict (de Vries & van Gelder, 2015; Lee et al., 2005; Pletzer et al., 2019). A follow-up analysis at the item-level (reported in Table S3) showed that the correlation

with HEXACO-PI Conscientiousness was, in fact, mainly driven by HEXACO-PI-R Conscientiousness items related to workplace behavior. These results corroborate the idea that an overlap between these traits may be simply explained by the presence of blended indicators, without the need to postulate second-order factors (Ashton & Lee, 2020).

The second aim of this work was to further deepen the connections between Honesty-Humility and Truthfulness. In developing the ACH, we used a systematic approach to capture the studied domain comprehensively. Recent discussions in personality psychology highlight the need for systematic bottom-up procedures in measurement construction that focus on personality characteristics below traditional traits—such as aspects, facets, and nuances (Mõttus et al., 2019; Revelle et al., 2021; Seeboth & Mõttus, 2018; Wilt & Revelle, 2015). The procedure outlined in Study 1 is not specific to the development of the ACH, but it can be more generally employed as a tool for screening candidate items when developing psychometric scales, aligning the item's content with the intended constructs. By using a bottom-up method, involving expert raters, and following a transparent and pre-registered procedure, we successfully minimized the risk of jingle-jangle fallacy (Flake & Fried, 2020; Wulff & Mata, 2025). This strategy allowed disconfirming the initial idea to assess five separate facets, pointing to the fact that Sincerity and Truthfulness were too overlapped to be assessed separately. However, in developing the ACH, we decided to assess the Sincerity facet (encompassing Truthfulness) with a larger number of items, to obtain a measure of this facet that could assess this part of the trait more closely.

More generally, our findings corroborate the idea that the observed mismatch between Honesty-Humility and Truthfulness noted in some works (Fleeson, 2020; Fleeson et al., 2022) might stem from the specificity of some of the HEXACO-PI-R items, with Truthfulness being a crucial part of the core of Honesty-Humility as a trait (Thielmann et al., 2024).

References

Aghababaei, N., Mohammadtabar, S., & Saffarinia, M. (2014). Dirty Dozen vs. the H factor: Comparison of the Dark Triad and Honesty–Humility in prosociality, religiosity, and happiness. *Personality and Individual Differences*, *67*, 6–10. https://doi.org/10.1016/j.paid.2014.03.026

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycho-lexical study. *Psychological Monographs*, *47*(1), i–171. https://doi.org/10.1037/h0093360

Anglim, J., Morse, G., De Vries, R. E., MacCann, C., & Marty, A. (2017). Comparing Job Applicants to Non–Applicants Using An Item–Level Bifactor Model on the Hexaco Personality Inventory. *European Journal of Personality*, *31*(6), 669–684. https://doi.org/10.1002/per.2120

Ashton, M. C., & Lee, K. (2001). A theoretical basis for the major dimensions of personality. *European Journal of Personality*, *15*(5), 327–353. https://doi.org/10.1002/per.417

Ashton, M. C., & Lee, K. (2005). The Lexical Approach to the Study of Personality Structure: Toward the Identification of Cross–Culturally Replicable Dimensions of Personality Variation. *Journal of Personality Disorders*, *19*(3), 303–308. https://doi.org/10.1521/pedi.2005.19.3.303

Ashton, M. C., & Lee, K. (2007). Empirical, Theoretical, and Practical Advantages of the HEXACO Model of Personality Structure. *Personality and Social Psychology Review*, *11*(2), 150–166. https://doi.org/10.1177/1088868306294907

Ashton, M. C., & Lee, K. (2008). The HEXACO Model of Personality Structure and the Importance of the H Factor. *Social and Personality Psychology Compass*, *2*(5), 1952–1962. https://doi.org/10.1111/j.1751-9004.2008.00134.x

Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A Short Measure of the Major Dimensions of Personality. *Journal of Personality Assessment*. https://www.tandfonline.com/doi/full/10.1080/00223890902935878

Ashton, M. C., & Lee, K. (2010). Trait and source factors in HEXACO-PI-R self- and observer reports. *European Journal of Personality*, *24*(3), 278–289. https://doi.org/10.1002/per.759

Ashton, M. C., & Lee, K. (2020). Objections to the HEXACO Model of Personality Structure—And why those Objections Fail. *European Journal of Personality*, *34*(4), 492–510. https://doi.org/10.1002/per.2242

Ashton, M. C., Lee, K., & De Vries, R. E. (2014). The HEXACO Honesty-Humility, Agreeableness, and Emotionality Factors: A Review of Research and Theory. *Personality and Social Psychology Review*, *18*(2), 139–152. https://doi.org/10.1177/1088868314523838

Ashton, M. C., Lee, K., de Vries, R. E., Perugini, M., Gnisci, A., & Sergi, I. (2006). The HEXACO model of personality structure and indigenous lexical personality dimensions in Italian, Dutch, and English. *Journal of Research in Personality*, *40*(6), 851–875. https://doi.org/10.1016/j.jrp.2005.06.003

Ashton, M. C., Lee, K., & Goldberg, L. R. (2004). A Hierarchical Analysis of 1,710 English Personality-Descriptive Adjectives. *Journal of Personality and Social Psychology*, *87*(5), 707–721. https://doi.org/10.1037/0022-3514.87.5.707

Ashton, M. C., Lee, K., Goldberg, L. R., & de Vries, R. E. (2009). Higher Order Factors of Personality: Do They Exist? *Personality and Social Psychology Review*, *13*(2), 79–91. https://doi.org/10.1177/1088868309338467

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., De Vries, R. E., Di Blas, L., Boies, K., & De Raad, B. (2004). A Six-Factor Structure of Personality-Descriptive Adjectives: Solutions From Psycholexical Studies in Seven Languages. *Journal of Personality and Social Psychology*, *86*(2), 356–366. https://doi.org/10.1037/0022-3514.86.2.356

Ashton, M. C., Lee, K., & Son, C. (2000). Honesty as the sixth factor of personality: Correlations with machiavellianism, primary psychopathy, and social adroitness. *European Journal of Personality*, *14*(4), 359–368. https://doi.org/10.1002/1099-0984(200007/08)14:4<359::aid-per382>3.0.co;2-y

Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology*, *97*(3), 533–548. https://doi.org/10.1037/a0016229

Balafoutas, L., García-Gallego, A., Georgantzis, N., Jaber-Lopez, T., & Mitrokostas, E. (2021). Psychopathy and Economic Behavior Among Prison Inmates: An Experiment. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.732184

Baumert, A., Beierlein, C., Schmitt, M., Kemper, C. J., Kovaleva, A., Liebig, S., & Rammstedt, B. (2014). Measuring Four Perspectives of Justice Sensitivity With Two Items Each. *Journal of Personality Assessment*, *96*(3), 380–390. https://doi.org/10.1080/00223891.2013.836526

Baumert, A., & Schmitt, M. (2016). Justice Sensitivity. In *Handbook of Social Justice Theory and Research* (pp. 161–180). Springer, New York, NY. https://doi.org/10.1007/978-1-4939-3216-0_9

Campbell, W. K., Bonacci, A. M., Shelton, J., Exline, J. J., & Bushman, B. J. (2004). Psychological Entitlement: Interpersonal Consequences and Validation of a Self-Report Measure. *Journal of Personality Assessment*, *83*(1), 29–45. https://doi.org/10.1207/s15327752jpa8301_04

Caprara, G. V., & Perugini, M. (1994). Personality described by adjectives: The generalizability of the Big Five to the Italian lexical context. *European Journal of Personality*, *8*(5), 357–369. https://doi.org/10.1002/per.2410080502

Catano, V. M., O'Keefe, D. F., Francis, R. E., & Owens, S. M. (2018). Construct-based approach to developing a short, personality-based measure of integrity. *International Journal of Selection and Assessment*, *26*(1), 75–92. https://doi.org/10.1111/ijsa.12197

Chiorri, C., Garofalo, C., & Velotti, P. (2019). Does the Dark Triad Manifest Similarly in men and Women? Measurement Invariance of the Dirty Dozen across sex. *Current Psychology*, *38*(3), 659–675. https://doi.org/10.1007/s12144-017-9641-5

Choshen-Hillel, S., Shaw, A., & Caruso, E. M. (2020). Lying to appear honest. *Journal of Experimental Psychology: General*, *149*(9), 1719–1735. https://doi.org/10.1037/xge0000737

Comrey A. L., & Lee H. B. (1992). *A first course in factor analysis*. Hillsdale, NJ: Erlbaum.

Condon, D. M., Wood, D., Mõttus, R., Booth, T., Costantini, G., Greiff, S., Johnson, W., Lukaszewski, A., Murray, A., Revelle, W., Wright, A. G. C., Ziegler, M., & Zimmermann, J. (2020). Bottom Up Construction of a Personality Taxonomy. *European Journal of Psychological Assessment*, *36*(6), 923–934. https://doi.org/10.1027/1015-5759/a000626

Cooper, B., Cohen, T. R., Huppert, E., Levine, E. E., & Fleeson, W. (2023). Honest Behavior: Truth-Seeking, Belief-Speaking, and Fostering Understanding of the Truth in Others. *Academy of Management Annals*, *17*(2), 655–683. https://doi.org/10.5465/annals.2021.0209

Costa Jr., P. T., & McCrae, R. R. (1988). From catalog to classification: Murray's needs and the five-factor model. *Journal of Personality and Social Psychology*, *55*(2), 258–265. https://doi.org/10.1037/0022-3514.55.2.258

Costantini, G., & Perugini, M. (2016). The network of conscientiousness. *Journal of Research in Personality*, *65*, 68–88. https://doi.org/10.1016/j.jrp.2016.10.003

Costantini, G., & Perugini, M. (2018). A Framework for Testing Causality in Personality Research. *European Journal of Personality*, *32*(3), 254–268. https://doi.org/10.1002/per.2150

Costantini, G., Perugini, M., Dentale, F., Barbaranelli, C., Alessandri, G., Vecchione, M., & Caprara, G. V. (2016). Assessing Positive Orientation with the Implicit Association Test. *European Journal of Psychological Assessment*, *2010*, 1–8. https://doi.org/10.1027/1015-5759/a000362

Costantini, G., Richetin, J., Borsboom, D., Fried, E. I., Rhemtulla, M., & Perugini, M. (2015). Development of Indirect Measures of Conscientiousness: Combining a Facets Approach and Network Analysis. *European Journal of Personality*, *29*(5), 548–567. https://doi.org/10.1002/per.2014

De Raad, B., Perugini, M., Hrebícková, M., & Szarota, P. (1998). Lingua Franca of Personality: Taxonomies and Structures Based on the Psycholexical Approach. *Journal of Cross-Cultural Psychology*, *29*(1), 212–232. https://doi.org/10.1177/0022022198291011

de Vries, R. E., & van Gelder, J.-L. (2015). Explaining workplace delinquency: The role of Honesty–Humility, ethical culture, and employee surveillance. *Personality and Individual Differences*, *86*, 112–116. https://doi.org/10.1016/j.paid.2015.06.008

Di Blas, L., & Forzi, M. (1998). An alternative taxonomic study of personality-descriptive adjectives in the Italian language. *European Journal of Personality*, *12*, 75–101.

Di Sarno, M., Costantini, G., Richetin, J., Preti, E., & Perugini, M. (2023). Why are you (un)conscientious? The dynamic interplay of goals, states, and traits in everyday life. *Journal of Personality*, *91*(4), 977–991. https://doi.org/10.1111/jopy.12701

Flake, J. K., & Fried, E. I. (2020). Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *Advances in Methods and Practices in Psychological Science*, *3*(4), 456–465. https://doi.org/10.1177/2515245920952393

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*(6), 1011–1027. https://doi.org/10.1037/0022-3514.80.6.1011

Fleeson, W. (2020). Live simply so that others may simply live: Trying to get clarity on the meaning of honesty/humility. *European Journal of Personality*, *34*(4), 524–526.

Fleeson, W., Furr, R. M., Jayawickreme, E., & Hardin, B. (2022). Honesty as a trait. *Current Opinion in Psychology*, *47*, 101418. https://doi.org/10.1016/j.copsyc.2022.101418

Gallucci, M. (2026). *PAMLj. Power Analysis for linear models* (Versione 0.9.0) [Jamovi]. https://github.com/pamlj/pamlj

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist*, *48*(1), 26–34. https://doi.org/10.1037/0003-066X.48.1.26

Goldberg, L. R. (2002). An Alternative "Description of Personality": The Big-Five Factor Structure. In *Personality and Personality Disorders*. Routledge.

Gunschera, L. J., Brazil, I. A., & Driessen, J. M. A. (2022). Social economic decision-making and psychopathy: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *143*, 104966. https://doi.org/10.1016/j.neubiorev.2022.104966

Hilbig, B. E., Zettler, I., Leist, F., & Heydasch, T. (2013). It takes two: Honesty–Humility and Agreeableness differentially predict active versus reactive cooperation. *Personality and Individual Differences*, *54*(5), 598–603. https://doi.org/10.1016/j.paid.2012.11.008

Hörz-Sagstetter, S., Volkert, J., Rentrop, M., Benecke, C., Gremaud-Heitz, D. J., Unterrainer, H.-F., Schauenburg, H., Seidler, D., Buchheim, A., Doering, S., Feil, M. G., Clarkin, J. F., Dammann, G., & Zimmermann, J. (2021). A Bifactor Model of Personality Organization. *Journal of Personality Assessment*, *103*(2), 149–160. https://doi.org/10.1080/00223891.2019.1705463

Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment*, *22*(2), 420–432. https://doi.org/10.1037/a0019265

Kline, R. B. (2023). *Principles and practice of structural equation modeling*. Guilford publications.

Krekels, G., & Pandelaere, M. (2015). Dispositional greed. *Personality and Individual Differences*, *74*, 225–230. https://doi.org/10.1016/j.paid.2014.10.036

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5), 535–540. https://doi.org/10.1038/nn.2303

Lee, K., & Ashton, M. C. (2004). Psychometric Properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research*, *39*(2), 329–358. https://doi.org/10.1207/s15327906mbr3902_8

Lee, K., & Ashton, M. C. (2006). Further assessment of the HEXACO Personality Inventory: Two new facet scales and an observer report form. *Psychological Assessment*, *18*(2), 182–191. https://doi.org/10.1037/1040-3590.18.2.182

Lee, K., & Ashton, M. C. (2008). The HEXACO Personality Factors in the Indigenous Personality Lexicons of English and 11 Other Languages. *Journal of Personality*, *76*(5), 1001–1054. https://doi.org/10.1111/j.1467-6494.2008.00512.x

Lee, K., & Ashton, M. C. (2018). Psychometric Properties of the HEXACO-100. *Assessment*, *25*(5), 543–556. https://doi.org/10.1177/1073191116659134

Lee, K., Ashton, M. C., & De Vries, R. E. (2005). Predicting Workplace Delinquency and Integrity with the HEXACO and Five-Factor Models of Personality Structure. *Human Performance*, *18*(2), 179–197. https://doi.org/10.1207/s15327043hup1802_4

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*, 61–83. https://doi.org/10.1016/j.jrp.2013.09.008

Marelli, M. (2017). Word-Embeddings Italian Semantic Spaces: A semantic model for psycholinguistic research. *Psihologija*, *50*(4), 503–520.

Markon, K. E. (2019). Bifactor and Hierarchical Models: Specification, Inference, and Interpretation. *Annual Review of Clinical Psychology*, *15*, 51–69. https://doi-org.mu.idm.oclc.org/10.1146/annurev-clinpsy-050718-095522

McArthur, J., Jarvis, R., Bourgeois, C., & Ternes, M. (2022). Lying motivations: Exploring personality correlates of lying and motivations to lie. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *54*(4), 335–340. https://doi.org/10.1037/cbs0000328

McArthur, J., McPhee, P., & Ternes, M. (2024). Examining the acceptability and likelihood of self-centered and other-centered lying using the HEXACO. *Personality and Individual Differences*, *220*, 112526. https://doi.org/10.1016/j.paid.2023.112526

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*(1), 81–90. https://doi.org/10.1037/0022-3514.52.1.81

McCrae, R. R., & Costa, P. T. Jr. (2008). The five-factor theory of personality. In *Handbook of personality theory and research. O. P. John, R. W. Robins, & L. A. Pervin (Eds.)* (3rd ed., pp. 159–181). Guilford.

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Miller, C. B., Fleeson, W., Cohen, T. R., Furr, R. M., & Jayawickreme, E. (2021). *Honesty: Big questions, philosophical issues, and empirical research (white paper for the Honesty Project grant)*.

Mõttus, R., Sinick, J., Terracciano, A., Hřebíčková, M., Kandler, C., Ando, J., Mortensen, E. L., Colodro-Conde, L., & Jang, K. L. (2019). Personality characteristics below facets:

A replication and meta-analysis of cross-rater agreement, rank-order stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, *117*(4), e35–e50. https://doi.org/10.1037/pspp0000202

Oliveira, C. M., & Levine, T. R. (2008). Lie Acceptability: A Construct and Measure. *Communication Research Reports*, *25*(4), 282–288. https://doi.org/10.1080/08824090802440170

Paul, A. A., Lee, K., & Ashton, M. C. (2022). Who tells prosocial lies? A HEXACO model investigation. *Journal of Research in Personality*, *98*, 104232. https://doi.org/10.1016/j.jrp.2022.104232

Perugini, M., Gallucci, M., & Costantini, G. (2018). A Practical Primer To Power Analysis for Simple Experimental Designs. *International Review of Social Psychology*, *31*(1). https://doi.org/10.5334/irsp.181

Perugini, M., & Leone, L. (2009). Implicit self-concept and moral action. *Journal of Research in Personality*, *43*(5), 747–754. https://doi.org/10.1016/j.jrp.2009.03.015

Pletzer, J. L., Bentvelzen, M., Oostrom, J. K., & de Vries, R. E. (2019). A meta-analysis of the relations between personality and workplace deviance: Big Five versus HEXACO. *Journal of Vocational Behavior*, *112*, 369–383. https://doi.org/10.1016/j.jvb.2019.04.004

R Core Team. (2022). *R: A Language and Environment for Statistical Computing*. [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Revelle, W., Dworak, E. M., & Condon, D. M. (2021). Exploring the persome: The power of the item in understanding personality structure. *Personality and Individual Differences*, *169*, 109905. https://doi.org/10.1016/j.paid.2020.109905

Ringwald, W. R., Manuck, S. B., Marsland, A. L., & Wright, A. G. C. (2022). Psychometric Evaluation of a Big Five Personality State Scale for Intensive Longitudinal Studies. *Assessment*, *29*(6), 1301–1319. https://doi.org/10.1177/10731911211008254

Roberts, B. W., Chernyshenko, O. S., Stark, S., & Goldberg, L. R. (2005). The Structure of Conscientiousness: An Empirical Investigation Based on Seven Major Personality Questionnaires. *Personnel Psychology*, *58*(1), 103–139. https://doi.org/10.1111/j.1744-6570.2005.00301.x

Romano, D., Costantini, G., Richetin, J., & Perugini, M. (2023). The HEXACO Adjective Scales and Its Psychometric Properties. *Assessment*, *30*(8), 2510–2532. https://doi.org/10.1177/10731911231153833

Rosseel, Y. (2012). lavaan: An *R* Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Schmitt, M., Baumert, A., Gollwitzer, M., & Maes, J. (2010). The Justice Sensitivity Inventory: Factorial Validity, Location in the Personality Facet Space, Demographic Pattern, and Normative Data. *Social Justice Research*, *23*(2), 211–238. https://doi.org/10.1007/s11211-010-0115-2

Ścigała, K. A., Schild, C., & Zettler, I. (2020). Dishonesty as a signal of trustworthiness: Honesty-Humility and trustworthy dishonesty. *Royal Society Open Science*, *7*(10), 200685. https://doi.org/10.1098/rsos.200685

Seeboth, A., & Mõttus, R. (2018). Successful Explanations Start with Accurate Descriptions: Questionnaire Items as Personality Markers for More Accurate Predictions. *European Journal of Personality*, *32*(3), 186–201. https://doi.org/10.1002/per.2147

Sekhar, S., Uppal, N., & Shukla, A. (2020). Dispositional greed and its dark allies: An investigation among prospective managers. *Personality and Individual Differences*, *162*, 110005. https://doi.org/10.1016/j.paid.2020.110005

Sharp, C., Wright, A. G. C., Fowler, J. C., Frueh, B. C., Allen, J. G., Oldham, J., & Clark, L. A. (2015). The structure of personality pathology: Both general ('g') and specific ('s') factors? *Journal of Abnormal Psychology*, *124*(2), 387–398. https://doi.org/10.1037/abn0000033

The jamovi project (2025). *jamovi* (Version 2.6) [Computer Software]. Retrieved from https://www.jamovi.org

Thielmann, I., Hilbig, B. E., Klein, S. A., Seidl, A., & Heck, D. W. (2024). Cheating to benefit others? On the relation between Honesty-Humility and prosocial lies. *Journal of Personality*, *92*(3), 870–882. https://doi.org/10.1111/jopy.12835

Thielmann, I., Hilbig, B. E., Zettler, I., & Moshagen, M. (2017). On Measuring the Sixth Basic Personality Dimension: A Comparison Between HEXACO Honesty-Humility and Big Six Honesty-Propriety. *Assessment*, *24*(8), 1024–1036. https://doi.org/10.1177/1073191116638411

Veselka, L., Giammarco, E. A., & Vernon, P. A. (2014). The Dark Triad and the seven deadly sins. *Personality and Individual Differences*, *67*, 75–80. https://doi.org/10.1016/j.paid.2014.01.055

Wilt, J., & Revelle, W. (2015). Affect, Behaviour, Cognition and Desire in the Big Five: An Analysis of Item Content and Structure. *European Journal of Personality*, *29*(4), 478–497. https://doi.org/10.1002/per.2002

Wood, A. M., Linley, P. A., Maltby, J., Baliousis, M., & Joseph, S. (2008). The authentic personality: A theoretical and empirical conceptualization and the development of the Authenticity Scale. *Journal of Counseling Psychology*, *55*(3), 385–399. https://doi.org/10.1037/0022-0167.55.3.385

Wood, D. (2015). Testing the lexical hypothesis: Are socially important traits more densely reflected in the English lexicon? *Journal of Personality and Social Psychology*, *108*(2), 317–335. https://doi.org/10.1037/a0038343

Wulff, D. U., & Mata, R. (2025). Semantic embeddings reveal and address taxonomic incommensurability in psychological measurement. *Nature Human Behaviour*, *9*(5), 944–954. https://doi.org/10.1038/s41562-024-02089-y

Xiaohua Chen, S., Bond, M. H., Chan, B., Tang, D., & Buchtel, E. E. (2009). Behavioral Manifestations of Modesty. *Journal of Cross-Cultural Psychology*, *40*(4), 603–626. https://doi.org/10.1177/0022022108330992

Yovel, I., & Friedman, A. (2013). Bridging the gap between explicit and implicit measurement of personality: The questionnaire-based implicit association test. *Personality and Individual Differences*, *54*(1), 76–80. https://doi.org/10.1016/j.paid.2012.08.015

Zettler, I., Thielmann, I., Hilbig, B. E., & Moshagen, M. (2020). The Nomological Net of the HEXACO Model of Personality: A Large-Scale Meta-Analytic Investigation. *Perspectives on Psychological Science*, *15*(3), 723–760. https://doi.org/10.1177/1745691619895036

Appendix. ACH scale with instructions

Italian (original): Di seguito, le chiediamo di indicare quanto ciascuno dei seguenti aggettivi (o breve espressione) risulti appropriato a descriverla, usando la seguente scala di risposta. Indichi 1 se l'aggettivo non la descrive per niente, 7 se l'aggettivo la descrive completamente. Usi le risposte 2, 3, 4, 5 e 6 per indicare gradi intermedi di importanza. Sappiamo che in molti casi la risposta dipende dalla situazione specifica. Le chiediamo però di rispondere pensando a quanto ciascun aggettivo la descrive "in media", attraverso diverse situazioni.

English (translated): Below, we ask you to indicate how appropriate each of the following adjectives (or brief expressions) is in describing you, using the following response scale. Indicate 1 if the adjective does not describe you at all, 7 if the adjective describes you completely. Use responses 2, 3, 4, 5, and 6 to indicate intermediate degrees of importance. We understand that in many cases, the answer depends on the specific situation. However, we ask you to respond thinking about how each adjective describes you "on average," across different situations.

1. amante della verità (truth-lover)
2. autentico/a (authentic)
3. avido/a (avid)
4. bugiardo/a (liar)
5. diretto/a (direct)
6. falso/a (false)
7. generoso/a (generous)
8. giusto/a (just)
9. immorale (immoral)
10. ingiusto/a (unfair)
11. insincero/a (insincere)
12. ipocrita (hypocritical)
13. menzognero/a (lying)
14. modesto/a (modest)
15. non avaro/a (not stingy)
16. obiettivo/a (objective)
17. sbruffone/a (boaster)
18. semplice (simple)
19. sincero/a (sincere)
20. vanitoso/a (vain)
21. venale (mercenary)
22. vero/a (truthful)

Scoring:

- Sincerity: 1, 2, 4*, 5, 6*, 11*, 12*, 13*, 19, 22
- Fairness: 8, 9*, 10*, 16
- Modesty: 14, 17*, 18, 20*
- Greed-Avoidance: 3*, 7, 15, 21*

Items marked with a star(*) must be reverse-coded with the formula 8-X, where X is the participants' response. Then, a score for each facet can be obtained by averaging the items.

Note. The scale can be administered together with the HEXACO Adjective Scales (Romano et al., 2023). In this case, items 3, 12, and 19, which are already included in the HAS, should not be administered twice.

Supplementary Materials

Supplementary material S1. Instructions for raters

Original Italian Version (see below for English translation)

PAGE 1

Grazie per aver accettato di partecipare. Ti chiediamo di partecipare a questo studio in qualità di giudice esperto.

Il tratto di "onestà" in psicologia può avere significati e sfaccettature differenti. Noi siamo interessati a comprendere se e in quale misura diversi aggettivi, usati per descrivere le persone e i loro comportamenti, siano rilevanti per diversi aspetti dell'onestà. Gli aspetti dell'onestà a cui siamo interessati sono:

1) **Falsità vs. veridicità.** Una persona vera è una che ricerca informazioni vere, comunica agli altri ciò che ritiene essere vero e cerca di promuovere credenze vere negli altri. Una persona falsa ha invece la tendenza a ingannare gli altri o mentire (incluso mentire a se stessi).

2) **Insincerità vs. sincerità.** Una persona sincera è autentica nelle relazioni interpersonali, non ama manipolare gli altri. Una persona insincera usa lusinghe e a finge che gli altri gli piacciano per farsi strada.

3) **Iniquità vs. equità**. Una persona equa evita frodi e corruzione, non ama sfruttare gli altri o approfittarsi di loro. Una persona iniqua è disposta a mentire o rubare, pur di ottenere un guadagno.

4) **Avidità vs. evitamento dell'avidità.** Una persona che evita l'avidità non è interessata a ricchezze e sfarzi, a beni di lusso o ad un alto status sociale. Una persona avida vuole avere e mostrare grandi ricchezze e un alto status sociale.

5) **Immodestia vs. modestia.** Una persona modesta è senza pretese, non richiede di essere trattata in modo diverso dagli altri. Una persona immodesta si considera superiore e pensa di avere privilegi che gli altri non hanno.

Nelle prossime schermate, ti proporremo una serie di aggettivi usati per descrivere persone o i loro comportamenti. Ti chiediamo di valutare in che misura ciascun aggettivo sia rappresentativo di ciascun tratto. Un aggettivo si intende rappresentativo di un tratto se sapere che una persona si comporta nel modo descritto dall'aggettivo ci dice qualcosa sul livello del tratto posseduto da quella persona (es. aggettivi quali "socievole" e "asociale" si potrebbero considerare rappresentativi rispettivamente di "estroversione" e "introversione", ma non di un tratto come "apertura all'esperienza").

PAGE 2

Ti preghiamo di indicare quanto giudichi che ciascun aggettivo è rappresentativo di ciascun tratto. Indica 0 se l'aggettivo non è rappresentativo né del polo positivo, né di quello negativo del tratto. Indica -5 o +5 se l'aggettivo è estremamente rappresentativo rispettivamente del polo negativo o di quello negativo. Usa valori intermedi per indicare gradazioni nella misura in cui l'aggettivo rappresenta il tratto

"Onesto"

-5 = *falsità* ----------0 = non rappresentativo ---------- 5 = *veridicità*.
-5 = *insincerità* ----------0 = non rappresentativo ---------- 5 = *sincerità*.
-5 = *iniquità* ----------0 = non rappresentativo ---------- 5 = *equità*.
-5 = avidità----------0 = non rappresentativo ---------- 5 = *evitamento dell'avidità*.
-5 = *immodestia* ----------0 = non rappresentativo ---------- 5 = *modestia*.

[Vengono mostrati tutti i 120+30 aggettivi, in pagine diverse. In fondo alla pagina sono sempre disponibili le definizioni dei tratti fornite a pagina 1]

English translation

PAGE 1

Thank you for agreeing to participate. We ask you to take part in this study as an expert rater. The trait of "honesty" in psychology can have various meanings and facets. We are interested in understanding whether, and to what extent, different adjectives used to describe people and their behaviors are relevant to different aspects of honesty. The aspects of honesty we are interested in are:

1. **Falsehood vs. Truthfulness**: A truthful person seeks true information, communicates to others what they believe to be true, and attempts to foster true beliefs in others. In contrast, a false person tends to deceive others or lie (including self-deception).
2. **Insincerity vs. Sincerity**: A sincere person is genuine in interpersonal relations, is unwilling to manipulate others. An insincere person uses flattery and pretends to like others as a way to get ahead.
3. **Unfairness vs. Fairness**: A fair person avoids fraud and corruption, is unwilling to exploit or take advantage of others. An unfair person is willing to gain by cheating or stealing.
4. **Greed vs. Greed-Avoidance**: A person who avoids greed has little interest in lavish wealth, luxury goods, or high social status. A greedy person wants to enjoy and display great wealth and status.
5. **Immodesty vs. Modesty**: A modest person is modest and unassuming, makes no claim to special treatment. An immodest person considers self superior and entitled to privileges that others do not have.

In the following screens, we will present a series of adjectives used to describe people or their behaviors. We ask you to assess the extent to which each adjective is representative of each trait. An adjective is considered representative of a trait if knowing that a person behaves in the way described by the adjective tells us something about the level of that trait possessed by that person (e.g., adjectives like "sociable" and "withdrawn" could be considered representative of "extraversion" and "introversion," respectively, but not of a trait like "openness to experience").

PAGE 2

Please indicate how representative you believe each adjective is for each trait. Mark 0 if the adjective is neither representative of the positive nor the negative end of the trait. Mark -5 or +5 if the adjective is extremely representative of the negative or positive end of the trait, respectively. Use intermediate values to indicate the degree to which the adjective represents the trait.

For "Honest":
-5 = Falsehood ---------- 0 = Not representative ---------- 5 = Truthfulness.
-5 = Insincerity ---------- 0 = Not representative ---------- 5 = Sincerity.
-5 = Unfairness ---------- 0 = Not representative ---------- 5 = Fairness.
-5 = Greed ---------- 0 = Not representative ---------- 5 = Greed-Avoidance.
-5 = Immodesty ---------- 0 = Not representative ---------- 5 = Modesty.

[All 120+30 adjectives are shown on different pages. Definitions of traits provided at PAGE 1 are always available at the bottom of the page.]

Supplementary material S2. Characteristics of the 150 adjectives included in Study 1.

Table S1 reports the features of the 150 adjectives administered in Study 1. For each adjective, the following properties are indicated:

- **Closest Facet** is the facet pole to which the item was closest, with respect to which the F-distance was computed. Control adjectives are simply marked as "control".
- **Fdist** is the F-distance, the Euclidean distance of the adjective from the closest facet pole in the four-dimensional space defined by the Honesty-Humility facets.
- **Cdist** is the C-distance, the Euclidean distance of the adjective from the central point of the space in the four-dimensional space defined by the Honesty-Humility facets. The F and C-distance are described in detail in the manuscript text and visualized in Figure 1.
- **Final selection**: For the 30 items selected, this column indicates the facet pole each adjective best represented. For control adjectives, this column indicates the facet pole to which they belong in the HAS (Romano et al., 2023). An additional set of 18 items were selected as representative of "honesty" or "dishonesty", because they had a relatively large C-distance and/or because they were already part of the Honesty-Humility scale of the HAS (Romano et al., 2023). We took care to balance this additional set of items in terms of trait pole, nine of them represent honesty and nine represent dishonesty. Although these adjectives are not directly relevant for developing the ACH, they can be used in future studies in which a broader set of items relevant for honesty and dishonesty in general is needed.
- Greed Avoid., Modesty, Fairness, and Sinc are mean scores of rater's judgments, normalized between -1 and 1. These were the scores used for computing the F- and C-distances. Sincerity is the average of Truthfulness and Sincerity judgments.

Table S1. *Features of 150 adjectives administered in Study 1.*

Adjective	Closest Facet	F-dist	C-dist	Final selection	Greed Avoid	Modesty	Fairness	Sincerity
Adjectives selected as a starting pool for developing the ACH								
vero	true	0.19	1.00	sincere	.12	.12	.08	.98
diretto	true	0.32	0.68	sincere	.00	.00	.00	.68
franco	true	0.34	0.72	sincere	.04	.08	.12	.70
autentico	true	0.42	0.86	sincere	.20	.28	.08	.78
sincero	true	0.42	1.03	sincere	.12	.16	.36	.94
amante della verità	true	0.48	0.98	sincere	.28	.08	.36	.86
menzognero	untrue	0.26	0.97	insincere	-.08	.00	-.24	-.94
in malafede	untrue	0.32	0.84	insincere	-.08	.00	-.24	-.80
falso	untrue	0.40	1.04	insincere	-.04	.00	-.40	-.96
ipocrita	untrue	0.41	0.99	insincere	-.16	-.08	-.36	-.90
insincero	untrue	0.46	0.70	insincere	-.08	.00	-.28	-.64
bugiardo	untrue	0.50	1.00	insincere	-.04	-.04	-.48	-.88
ingordo	greed	0.30	0.83	greed	-.80	-.16	-.16	.00
venale	greed	0.36	0.90	greed	-.84	-.32	-.04	-.02
non avaro*	greed	0.46	0.80	greed	-.72	-.16	-.32	-.04
avido	greed	0.47	1.10	greed_av	-1.00	-.32	-.28	-.20
senza pretese	greed_av	0.70	1.08	greed_av	.84	.64	.24	.00
generoso	greed_av	0.83	1.14	greed_av	.80	.24	.72	.28
sobrio	modesty	0.52	0.84	modesty	.44	.72	.04	.00
modesto	modesty	0.55	1.14	modesty	.40	1.00	.36	.12
semplice	modesty	0.68	0.86	modesty	.48	.64	.20	.26
sbruffone	immodesty	0.25	0.99	immodesty	-.16	-.96	-.12	-.14
immodesto	immodesty	0.29	1.04	immodesty	-.28	-1.00	-.08	-.02
vanitoso	immodesty	0.32	0.84	immodesty	-.24	-.80	-.08	-.02
imparziale	fairness	0.47	1.03	fairness	.16	.12	.92	.42
obiettivo	fairness	0.62	0.71	fairness	.12	.24	.56	.34
giusto	fairness	0.68	1.18	fairness	.48	.20	.96	.44
ingiusto	unfairness	0.36	0.80	unfairness	-.20	-.16	-.76	-.06
prepotente	unfairness	0.70	0.92	unfairness	-.56	-.24	-.68	-.10
immorale	unfairness	0.82	1.09	unfairness	-.48	-.20	-.76	-.58
HAS items and adjectives identified as relevant for Honesty-Humility based on C-distance, but not specific to a facet								
retto	fairness	1.11	1.44	honesty	.68	.44	.92	.76
morale	fairness	0.95	1.32	honesty	.64	.32	.92	.62
equo	fairness	0.83	1.30	honesty	.68	.36	1.00	.32
onesto	true	0.88	1.24	honesty	.40	.16	.76	.88
come una persona di buon cuore	fairness	0.96	1.20	honesty	.60	.44	.76	.56
umile	modesty	0.76	1.19	honesty	.68	.92	.28	.18
maligno	untrue	1.09	1.26	dishonesty	-.56	-.56	-.68	-.70
ingannatore	untrue	0.79	1.24	dishonesty	-.32	-.24	-.68	-.96
raggiratore	untrue	0.83	1.18	dishonesty	-.40	-.20	-.68	-.86
imbroglione	untrue	0.79	1.18	dishonesty	-.36	-.12	-.68	-.88

Adjective	Closest Facet	F-dist	C-dist	Final selection	Greed Avoid	Modesty	Fairness	Sincerity
sleale	untrue	0.75	1.13	dishonesty	-.28	-.08	-.68	-.86
egoista	greed	0.72	1.13	dishonesty	-.88	-.16	-.68	-.12
leale	true	0.75	0.98	honesty	.28	.20	.60	.70
fedele	true	0.61	0.67	honesty	.04	.00	.40	.54
snob	immodesty	0.68	1.07	dishonesty	-.64	-.84	-.16	-.02
disonesto	untrue	0.59	1.05	dishonesty	-.24	-.04	-.52	-.88
altezzoso	immodesty	0.39	1.03	dishonesty	-.32	-.96	-.20	-.08
integro	true	0.91	1.14	honesty	.44	.20	.72	.74
Selected Control adjectives								
calmo	control		0.00	A+	0.00	0.00	0.00	0.00
tranquillo	control		0.15	A+	.12	.08	.04	.00
litigioso	control		0.06	A-	-0.04	-0.04	0.00	-0.02
collerico	control		0.00	A-	0.00	0.00	0.00	0.00
organizzato	control		0.00	C+	0.00	0.00	0.00	0.00
attento	control		0.23	C+	.08	.00	.20	.08
disorganizzato	control		0.00	C-	0.00	0.00	0.00	0.00
distratto	control		0.00	C-	0.00	0.00	0.00	0.00
vulnerabile	control		0.04	E+	0.00	0.00	0.00	0.04
emotivo	control		0.11	E+	0.08	0.04	0.00	0.06
coraggioso	control		0.12	E-	.12	.00	.00	.00
stabile emotivamente	control		0.06	E-	0.00	0.00	0.00	0.06
innovativo	control		0.08	O+	0.00	-0.08	0.00	0.00
originale	control		0.12	O+	.00	-.12	.00	.02
tradizionalista	control		0.07	O-	-0.04	-0.04	-0.04	0.00
non creativo	control		0.07	O-	0.00	0.04	0.04	0.04
estroverso	control		0.04	X+	0.00	-0.04	0.00	0.00
vivace	control		0.04	X+	0.00	0.00	0.00	0.04
timido	control		0.09	X-	0.04	0.08	0.00	0.00
introverso	control		0.00	X-	0.00	0.00	0.00	0.00
Honesty adjectives discarded from the final pool								
trasparente	true	0.51	0.99	discarded	.16	.24	.40	.86
traditore	untrue	0.62	0.99	discarded	-.24	-.12	-.52	-.80
ingannevole	untrue	0.51	0.99	discarded	-.08	-.04	-.48	-.86
mendace	untrue	0.42	0.99	discarded	-.04	-.08	-.40	-.90
attendibile	true	0.46	0.99	discarded	.12	.16	.40	.88
supponente	immodesty	0.46	0.99	discarded	-.32	-.88	-.24	-.20
genuino	true	0.63	0.98	discarded	.40	.40	.16	.78
fraudolento	untrue	0.74	0.98	discarded	-.32	.00	-.60	-.70
impostore	untrue	0.56	0.97	discarded	-.20	-.08	-.48	-.82
centrato su se stesso	immodesty	0.64	0.97	discarded	-.44	-.76	-.36	-.18
arrogante	immodesty	0.36	0.94	discarded	-.28	-.88	-.20	.00
magnanimo	fairness	0.70	0.88	discarded	.52	.20	.64	.22
ciarlatano	untrue	0.45	0.87	discarded	-.16	.00	-.36	-.78
vanesio	immodesty	0.37	0.86	discarded	-.28	-.80	-.04	-.14

Adjective	Closest Facet	F-dist	C-dist	Final selection	Greed Avoid	Modesty	Fairness	Sincerity
adulatore	untrue	0.62	0.86	discarded	-.44	-.16	-.24	-.68
verace	true	0.29	0.85	discarded	.08	.08	.20	.82
calunniatore	untrue	0.53	0.85	discarded	-.16	-.12	-.40	-.72
probo	true	0.63	0.84	discarded	.32	.12	.40	.66
invidioso	greed	0.54	0.81	discarded	-.68	-.24	-.20	-.30
virtuoso	true	0.84	0.79	discarded	.44	.16	.44	.46
esibizionista	immodesty	0.58	0.78	discarded	-.44	-.64	-.04	-.08
borioso	immodesty	0.45	0.75	discarded	-.24	-.68	-.20	-.02
lusinghiero	untrue	0.64	0.73	discarded	-.40	-.20	-.12	-.56
schietto	true	0.31	0.70	discarded	.00	.04	.04	.70
dispotico	greed	0.72	0.69	discarded	-.48	-.24	-.44	.00
dissoluto	greed	0.66	0.69	discarded	-.52	-.16	-.36	-.22
pretenzioso	immodesty	0.59	0.69	discarded	-.40	-.56	.00	-.02
furbo	untrue	0.71	0.66	discarded	-.24	.00	-.40	-.46
naturale	true	0.62	0.65	discarded	.12	.36	.12	.52
sornione	untrue	0.51	0.65	discarded	-.12	-.12	-.24	-.58
dissimulatore	untrue	0.38	0.65	discarded	.00	-.08	-.08	-.64
pieno di orgoglio	immodesty	0.50	0.61	discarded	-.20	-.56	-.12	-.08
nobile	fairness	0.63	0.60	discarded	.28	.12	.48	.18
egocentrico	immodesty	0.56	0.60	discarded	-.28	-.52	-.08	.00
disponibile	greed_av	0.77	0.55	discarded	.36	.28	.28	.14
sprezzante	immodesty	0.71	0.55	discarded	-.24	-.40	-.28	-.08
compassionevole	fairness	0.64	0.53	discarded	.12	.24	.44	.14
soverchiante	unfairness	0.69	0.53	discarded	-.28	-.20	-.40	.00
inattendibile	untrue	0.50	0.50	discarded	.00	.00	-.04	-.50
spavaldo	immodesty	0.54	0.50	discarded	-.12	-.48	-.08	.00
gentile	greed_av	0.82	0.48	discarded	.28	.24	.28	.12
moderato	modesty	0.75	0.45	discarded	.16	.32	.28	.02
compiaciuto	immodesty	0.56	0.44	discarded	-.04	-.44	.00	.00
ospitale	fairness	0.80	0.44	discarded	.24	.24	.28	.04
discreto	modesty	0.76	0.38	discarded	.16	.28	.20	.00
pettegolo	untrue	0.69	0.34	discarded	.00	-.04	-.12	-.32
mite	modesty	0.74	0.32	discarded	.16	.28	.00	.00
decoroso	true	0.81	0.30	discarded	.04	.20	.04	.22
composto	modesty	0.77	0.26	discarded	.08	.24	.04	-.02
desideroso	greed	0.76	0.25	discarded	-.24	-.04	-.04	.00
non pretenzioso	modesty	0.75	1.19	discarded	.64	.92	.32	.22
corretto	fairness	0.75	1.12	discarded	.36	.16	.84	.62
inesorabile	greed	0.93	0.13	discarded	-.08	.00	-.08	.06
millantatore	immodesty	0.80	1.11	discarded	-.36	-.80	-.24	-.64
calcolatore	untrue	0.91	1.10	discarded	-.56	-.32	-.56	-.70
corrotto	greed	0.96	1.10	discarded	-.64	-.08	-.64	-.62
superbo	immodesty	0.45	1.10	discarded	-.40	-1.00	-.20	-.06
degno di fiducia	true	0.80	1.10	discarded	.32	.36	.60	.78
inflessibile	fairness	0.92	0.11	discarded	0.00	0.00	0.08	0.08
immaturo	greed	0.96	0.08	discarded	-0.04	-0.04	-0.04	-0.04

Adjective	Closest Facet	F-dist	C-dist	Final selection	Greed Avoid	Modesty	Fairness	Sincerity
doppiogiochista	untrue	0.54	1.06	discarded	-.20	-.12	-.48	-.92
incorrutibile	fairness	0.71	1.05	discarded	.56	.20	.80	.34
altruista	fairness	0.80	1.04	discarded	.68	.24	.72	.22
dimesso	modesty	0.69	1.04	discarded	.64	.80	.16	.02
presuntuoso	immodesty	0.46	1.03	discarded	-.32	-.92	-.32	-.06
affidabile	true	0.64	1.03	discarded	.20	.16	.56	.82
providente	fairness	0.96	0.04	discarded	0.00	0.00	0.04	0.00
ostentatore	immodesty	0.44	1.02	discarded	-.36	-.92	-.16	-.18
megalomane	immodesty	0.64	1.01	discarded	-.52	-.80	-.12	-.30
vanaglorioso	immodesty	0.50	1.00	discarded	-.40	-.88	-.08	-.26
subdolo	untrue	0.64	1.00	discarded	-.16	-.16	-.56	-.80
superficiale	greed_av	1.00	0.00	discarded	0.00	0.00	0.00	0.00
Control adjectives discarded from the final pool								
coscienzioso	control	C+	0.60	discarded	.36	.16	.36	.28
aggressivo	control	A-	0.27	discarded	-.16	-.08	-.20	-.06
pacifico	control	A+	0.18	discarded	.12	.12	.04	-.02
sicuro	control	E-	0.18	discarded	.00	.08	.00	.16
amante dell'arte	control	O+	0.14	discarded	-.08	-.12	.00	.00
non amante dell'arte	control	O-	0.14	discarded	.08	.12	.00	.00
incostante	control	C-	0.14	discarded	.00	.00	.00	-.14
silenzioso	control	X-	0.13	discarded	.04	.12	.00	-.02
pauroso	control	E+	0.06	discarded	0.04	0.04	0.00	-0.02
espansivo	control	X+	0.02	discarded	0.00	0.00	0.00	0.02

Note. * Avaro was adapted as "non avaro" to reflect greed avoidance instead of greed

Supplementary material S3. CFA loadings (Study 2 and Study 3)

Table S2. *Standardized CFA loadings of the final HAS items in Studies 2 and 3.*

Factor	Adjective/item	English translation	Study 2	Study 3
Sincerity	Amante della verità	Truthlover	.705	.601
Sincerity	Autentico	Authentic	.716	.592
Sincerity	Diretto	Direct	.450	.457
Sincerity	Sincero	Sincere	.742	.789
Sincerity	Vero	True	.716	.709
Sincerity	Bugiardo	Liar	-.506	-.784
Sincerity	Falso	False	-.532	-.744
Sincerity	Insincero	Insincere	-.463	-.750
Sincerity	Ipocrita	Hypocritical	-.547	-.652
Sincerity	Menzognero	Liar	-.486	-.770
Fairness	Giusto	Just	.725	.768
Fairness	Obiettivo	Objective	.434	.460
Fairness	Immorale	Immoral	-.342	-.576
Fairness	Ingiusto	Unfair	-.455	-.706
Modesty	Modesto	Modest	.504	.546
Modesty	Semplice	Simple	.550	.571
Modesty	Sbruffone	Boaster	-.459	-.580
Modesty	Vanitoso	Vain	-.476	-.585
Greed-avoidance	Generoso	Generous	.669	.653
Greed-avoidance	Non avaro	Not stingy	.646	.734
Greed-avoidance	Avido	Greedy	-.666	-.728
Greed-avoidance	Venale	Mercenary	-.385	-.542
General factor	Amante della verità	Truthlover	.127	.239
General factor	Autentico	Authentic	.073	.373
General factor	Diretto	Direct	.123	.248
General factor	Sincero	Sincere	.003	.178
General factor	Vero	True	-.069	.279
General factor	Bugiardo	Liar	.485	.327
General factor	Falso	False	.456	.255
General factor	Insincero	Insincere	.549	.141
General factor	Ipocrita	Hypocritical	.352	.107
General factor	Menzognero	Liar	.596	.281
General factor	Giusto	Just	.116	.273
General factor	Obiettivo	Objective	.150	.199
General factor	Immorale	Immoral	.328	.225
General factor	Ingiusto	Unfair	.438	.248
General factor	Modesto	Modest	.026	.000
General factor	Semplice	Simple	-.042	.186
General factor	Sbruffone	Boaster	.380	.289
General factor	Vanitoso	Vain	.195	.180
General factor	Generoso	Generous	.067	.325
General factor	Non avaro	Not stingy	-.025	.073
General factor	Avido	Greedy	.213	.215
General factor	Venale	Mercenary	.110	.210

Supplementary material S4. Supplementary analysis Study 3.

Table S3. *Correlations between ACH scales and HEXACO-60 conscientiousness items.*

		Study 2				Study 3			
#	Item (facet)	SI	FA	MO	GR	SI	FA	MO	GR
2	I plan ahead and organize things, to avoid scrambling at the last minute (Organization)	.14*	.10	.03	-.03	.21***	.25***	.29***	.09
8	I often push myself very hard when trying to achieve a goal (Diligence)	.40***	.30***	.13*	.19**	.43***	.30***	.10	.21***
14	When working on something, I don't pay much attention to small details (Perfectionism, reverse-coded)	-.30***	-.22***	-.11	-.16**	-.23***	-.23***	-.15**	-.10
20	I make decisions based on the feeling of the moment rather than on careful thought (Prudence, reverse-coded)	-.08	-.21***	-.10	-.04	-.21***	-.19**	-.12*	.01
26	When working, I sometimes have difficulties due to being disorganized (Organization, reverse-coded)	-.18**	-.14*	-.04	-.05	-.32***	-.29***	-.23***	-.15*
32	I do only the minimum amount of work needed to get by (Diligence, reverse-coded)	-.32***	-.32***	-.18**	-.24***	-.35***	-.30***	-.17**	-.29***
38	I always try to be accurate in my work, even at the expense of time (Perfectionism)	.37***	.29***	.21***	.20***	.37***	.37***	.23***	.24***
44	I make a lot of mistakes because I don't think before I act (Prudence, reverse-coded)	-.13*	-.19**	-.09	-.09	-.25***	-.34***	-.15*	-.08
50	People often call me a perfectionist (Perfectionism)	.17**	.20***	.01	-.02	.23***	.15**	.01	.09
56	I prefer to do whatever comes to mind, rather than stick to a plan (Prudence, reverse-coded)	-.05	-.19**	-.03	.02	-.12*	-.24***	-.18**	-.04

Note. Item numbers (#) refer to Ashton & Lee (2009). SI = Sincerity, FA = Fairness, MO = Modesty, GR = Greed-Avoidance.
* $p < .05$, ** $p < .01$, *** $p < .001$

Supplementary material S5. Extreme responses and difficulty of ACH and HEXACO-PI-R.

Honesty-Humility is generally considered a socially desirable trait, and its measures show positive correlations with impression management scales—though the interpretation of these associations remains debated (Ashton & Lee, 2020; de Vries et al., 2014, 2018, 2019; Feldman, 2019; Feldman et al., 2017). If phrased vs. adjective items of Honesty-Humility were differentially vulnerable to impression management, for example if adjective items were easier to endorse by individuals just to manage their social impression, this would imply a different pattern of extremity in responses in adjective-based vs. phrased measures, and it would also be reflected in systematically lower difficulty for adjective measures as compared to phrased ones. In turn, this might threaten the validity of adjective measures. We thus analyzed pooled data from Studies 2 and 3 (N = 566) and examined whether participants used extreme response options differently across the two formats and whether these formats differ in difficulty.

First, all items were coded so that higher scores indicated higher levels of Honesty-Humility. We then inspected the frequency of extreme responses and conducted an IRT analysis using the generalized partial credit model (Muraki, 1992). Item difficulty was quantified using the LI_{IRF} index (Ali et al., 2015), implemented via the *gen.difficulty* function in the R package *mirt* (Chalmers, 2012). LI_{IRF} expresses the latent trait level (in Z-scores) that results in the expectation that the participant will respond using the midpoint of a Likert-type response scale, generalizing the concept of item difficulty for polytomous items. Despite differences in response scales—ACH used 7 points (1 = "Does not describe me at all", 7 = "Describes me completely") and HEXACO-PI-R used 5 points (1 = "Strongly disagree", 5 = "Strongly agree")—LI_{IRF} enables a rough comparison of item difficulty centered on the midpoints (4 for ACH, 3 for HEXACO). Results (see Table S5) showed relatively similar

patterns of extreme responding across the two formats. Very few participants endorsed the lowest response category (< 5%), while a substantial proportion endorsed the highest category, especially for the HEXACO-PI-R (30.6–44.7%) and to a slightly lesser extent for the ACH (23.6–32.7%). The higher extreme-positivity for the HEXACO-PI-R may partly reflect the lower number of response options. The LI_{IRF} values were also comparable. Across facets, the estimated trait levels needed to endorse the midpoint corresponded to below-average Honesty-Humility for both formats (−0.80 to −1.67 for HEXACO-PI-R; −1.27 to −1.58 for ACH). In conclusion, these findings do not suggest meaningful differences in response extremity or item difficulty between adjective- and phrase-based measures.

Table S4. *IRT analysis and extreme point usage for ACH and HEXACO-PI-R facet scales*

	LI_{IRF}				Extreme point usage	
	Mean	SD	Min	Max	Prop Min	Prop Max
Sincerity (HEXACO-PI-R)	-1.29	0.33	-1.92	-0.98	.015	.306
Sincerity (ACH)	-1.42	0.18	-1.72	-1.12	.003	.327
Fairness (HEXACO-PI-R)	-1.17	0.54	-1.97	-0.54	.047	.344
Fairness (ACH)	-1.58	0.32	-2.01	-1.30	.000	.303
Modesty (HEXACO-PI-R)	-1.67	0.77	-2.91	-0.99	.013	.447
Modesty (ACH)	-1.27	0.38	-1.81	-0.92	.013	.272
Greed-Avoidance (HEXACO-PI-R)	-0.80	0.75	-1.94	-0.30	.045	.327
Greed-Avoidance (ACH)	-1.39	0.37	-1.72	-0.86	.007	.236

Note. LI_{IRF} = generalized item difficulty index (Ali et al., 2015). Prop Min is the proportion of participants using the value representing minimal levels of Honesty-Humility (1 for both scales) and prop Max is the proportion of participants using response option (5 for the HEXACO-PI-R and 7 for the ACH) corresponding to the maximum possible level of Honesty-Humility.

References

Ali, U. S., Chang, H.-H., & Anderson, C. J. (2015). Location Indices for Ordinal Polytomous Items Based on Item Response Theory. *ETS Research Report Series*, *2015*(2), 1–13. https://doi.org/10.1002/ets2.12065

Ashton, M. C., & Lee, K. (2020). Objections to the HEXACO model of personality structure—And why those objections fail. *European Journal of Personality*, *34*(4), 492–510. https://doi.org/10.1002/per.2242

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6), 1–29. https://doi.org/10.18637/jss.v048.i06

de Vries, R. E., Hilbig, B. E., Zettler, I., Dunlop, P. D., Holtrop, D., Lee, K., & Ashton, M. (2019). *Still misinterpreting lie scales: Reply to Feldman's rejoinder*.

de Vries, R. E., Hilbig, B. E., Zettler, I., Dunlop, P. D., Holtrop, D., Lee, K., & Ashton, M. C. (2018). Honest people tend to use less—Not more—Profanity: Comment on Feldman et al.'s (2017) Study 1. *Social Psychological and Personality Science*, *9*(5), 516–520. https://doi.org/10.1177/1948550617714586

de Vries, R. E., Zettler, I., & Hilbig, B. E. (2014). Rethinking Trait Conceptions of Social Desirability Scales: Impression Management as an Expression of Honesty-Humility. *Assessment*, *21*(3), 286–299. https://doi.org/10.1177/1073191113504619

Feldman, G. (2019). What Is Honesty? Laypersons Interpret High Lie Scale Scores as Reflecting Intentional Dishonesty: Rejoinder to de Vries et al.'s (2017) Comment on Feldman et al. (2017). *Social Psychological and Personality Science*, *10*(2), 220–226. https://doi.org/10.1177/1948550617737141

Feldman, G., Lian, H., Kosinski, M., & Stillwell, D. (2017). Frankly, We Do Give a Damn: The Relationship Between Profanity and Honesty. *Social Psychological and Personality Science*, *8*(7), 816–826. https://doi.org/10.1177/1948550616681055

Muraki, E. (1992). A Generalized Partial Credit Model: Application of an Em Algorithm. *ETS Research Report Series*, *1992*(1), i–30. https://doi.org/10.1002/j.2333-8504.1992.tb01436.x

3

Uncovering the motivational core of Honesty-Humility

Galkina, A.*, Amenta, S.*, Zarbo, C., Perugini, M., Costantini, G. (2025), Uncovering the motivational core of Honesty-Humility. (Under review).
Preprint: https://doi.org/10.31234/osf.io/u93zp_v1

*The two authors contributed equally to this publication.

Abstract

Personal goals represent the motivational foundation underlying personality traits, yet honesty-humility motivational underpinnings remain poorly understood. We developed and validated the Goals for Honesty and Goals for Dishonesty (GHGD) questionnaire - the first instrument measuring honesty-humility through its motivational core. Using a multi-method, bottom-up approach, we identified and classified personal goals associated with honesty-humility. Study 1 (N = 250) employed participant-driven goal elicitation using personality-descriptive adjectives. Responses were classified using natural language processing and raters' coding, yielding a final set of goals. Study 2 (N = 300) refined goals, selecting those specifically associated with honesty-humility, developing the GHGD questionnaire. Study 3 (N = 400 participants and N = 216 informants) provided validation using self-reports, informant-reports, and observed behavior in an incentivized cheating task. Results revealed a two-factor structure distinguishing honesty-oriented(GH) from dishonesty-oriented goals (GD). These dimensions were moderately negatively correlated, indicating distinct rather than bipolar motivational orientations. GD predicted dishonest behavior beyond trait measures, while GH related more strongly to authentic living and self-concept integrity. Overall, goal-based assessments improved the prediction of honest behavior compared to trait self-reports. This work integrates trait-based and process-oriented approaches to personality, delivering theoretical and empirical contributions for future research on moral personality and behavior.

Introduction

Personal goals are cognitive representations of desired end states that a person is committed to attaining (e.g., Milyavskaya & Werner, 2018). Research shows that personality traits are related to personal goals (e.g., Reisz et al., 2013; Roberts & Robins, 2000) and that goals and traits co-evolve over the lifespan (Atherton et al., 2020; Buchinger et al., 2024; Lüdtke et al., 2009; Roberts et al., 2004). Traits have been described as characteristics that distinguish individuals and manifest consistently through their behaviors, emotions, and thoughts (McCrae & Costa Jr., 1990). The correspondence perspective (Baumert et al., 2017) posits that each trait results from a distinct causal generator with biological foundation, and their changes over time can be fully explained by biological developmental processes (McCrae & Costa, 2008; McCrae & Sutin, 2018). In this view, goals are seen as characteristic adaptations, reflecting trait adaptation to environmental demands, without affecting trait levels (Bleidorn et al., 2010; McCrae & Sutin, 2018). However, the more recent emergence perspective views personality traits as emerging from a network of causal interactions involving behaviors, thoughts, feelings, situations, and goals (Baumert et al., 2017; Costantini & Perugini, 2018). This latter perspective is supported by evidence that traits can undergo modifications over time, in response to life events (Bühler et al., 2023), or experimental interventions (Roberts et al., 2017; Stieger et al., 2021, 2024), and that people can change aspects of their personality in the desired direction (Hudson et al., 2020). It has also been suggested that the two views are not necessarily incompatible: traits might result from a combination of both types of processes (Mõttus & Allerhand, 2018).

Conceptually closer to a correspondence view, cybernetic models of personality assume that traits are stable parameters of a cybernetic motivational system regulating goal pursuit (Allen & DeYoung, 2017; DeYoung, 2015; Quirin et al., 2020, 2023). The whole trait theory (Fleeson & Jayawickreme, 2021; Jayawickreme et al., 2019), closer to the emergence

perspective, observes that whereas broad trait levels show stability, the corresponding state levels vary widely within individuals. The whole trait theory seeks to explain both phenomena by postulating that traits are parameters of the distributions of persons' state levels over time. Whereas states change often, the parameters describing their distribution, such as the mean, are relatively stable and correlate with scores in standard personality questionnaires, thus indicating a strong relationship between traits and their momentary manifestations (Fleeson & Gallagher, 2009). Seeking to explain the processes governing within-person variation in states, McCabe and Fleeson (2012, 2016) found that individuals modulate their states across situations to pursue their goals, viewing traits as goal-directed distributions of states.

The implication of this research is that goals are crucial for understanding the interplay between personality structure and processes (Costantini & Perugini, 2018). However, trait-related goals remain underinvestigated. A few studies have identified the motivational core of traits using a bottom-up procedure, starting from conscientiousness, and using a three-step procedure (Costantini et al., 2020; Costantini & Perugini, 2018). This procedure started from lexical markers of the trait and used such markers to elicit a corpus of goals from individuals, which was then used to create questionnaire items. Capitalizing on this early work, the aim of the current study is to explore the motivational core of another multifaceted personality trait: honesty-humility, one of the six broad dimensions of personality in the HEXACO model(Ashton et al., 2004; Ashton & Lee, 2020). From a behavioral standpoint, it is characterized by avoiding manipulating others for personal gain, being sincere and modest, and refraining from exploiting others, even when it goes unpunished (Ashton & Lee, 2005, 2007; Lee et al., 2013). As such, honesty-humility predicts prosocial behaviors related to active cooperation and lack of exploitation, with important societal implications (Hilbig et al., 2013; Thielmann et al., 2020; Zettler et al., 2020).

Despite its theoretical and practical significance, the motivational underpinnings of honesty-humility remain poorly understood. Few studies have attempted to identify the personal goals tied to the expression of honesty or dishonesty. This gap is surprising given the increasing recognition that honesty, both as a psychological trait and as a moral virtue, cannot be fully captured by specific behaviors (e.g., telling the truth or refusing to cheat). Rather, it must be understood through underlying goals and motives (Miller, 2020; Miller et al., 2021; Wilson, 2018). Research shows that observers judge morality based on inferred goals and motives, rather than actions (Carlson et al., 2022). For instance, some individuals may tell the truth not merely to comply with social norms, but because they are committed to a deeper goal of being a trustworthy person. Conversely, others may lie for altruistic purposes, such as protecting someone's feelings, which complicates the relationship between honesty, behavior, and underlying motivation. Empirical findings reveal complex interplay between honesty-humility and goals. On the one hand, individuals high in honesty-humility behave truthfully regardless of whether lying could serve prosocial purposes (Thielmann et al., 2023), pointing to an internalized moral compass. Yet, other studies show that such individuals may lie if they perceive dishonesty as serving a moral or benevolent end (Choshen-Hillel et al., 2020; McArthur et al., 2024; Ścigała et al., 2020). These findings suggest honesty-humility may reflect not just a tendency toward specific behaviors, but a motivational orientation toward honesty as a terminal or instrumental goal—raising the question of what kinds of goals truly define this trait.

More broadly, this focus aligns with emerging views conceptualizing personality traits not as fixed causal generators, but as dynamic networks of interrelated psychological components, including thoughts, feelings, behaviors, and goals (Baumert et al., 2017; Costantini & Perugini, 2018; Cramer et al., 2012; Fleeson & Jayawickreme, 2015). Identifying the goal structures associated with honesty-humility is crucial both for deepening our

understanding of the trait itself and exploring its potential malleability and responsiveness to environmental or intervention-based influences (e.g., Stieger et al., 2021, 2024). In summary, uncovering the motivational architecture of honesty-humility may offer a richer and more process-oriented account of this key moral trait. Whereas several studies investigated trait-goal connections, most focused on the Big Five traits or their subset, neglecting honesty-humility and considering either very specific goals (McCabe & Fleeson, 2012, 2016) or very broad motives (Bleidorn et al., 2010; Buchinger et al., 2024; Lüdtke et al., 2009; Roberts et al., 2004; Roberts & Robins, 2000). A few studies on goals of deceptive behaviors (DePaulo et al., 1996; T. R. Levine et al., 2016) have not fully examined the relationships between motives and honesty-humility, nor developed a psychometric assessment of such goals.

The present study, therefore, aims to empirically identify a comprehensive questionnaire to assess the subjective importance of goals related to honesty (GH) and to dishonesty (GD). This was achieved through three studies. Study 1 identified and classified a set of goals associated with honesty-humility. Based on these classes, a preliminary version of the questionnaire was developed. Study 2 refined the questionnaire by selecting goals and items specifically associated with trait honesty-humility. Study 3 collected validity evidence for the final questionnaire by predicting behavioral tendencies related to honesty-humility, through self-reports, informant-reports, and observed dishonest behavior in a computerized task.

All studies were pre-registered (preregistration files, data, and data-analyses scripts are available at https://osf.io/m4bhg/overview?view_only=0ec9ba70639d4275b3970b7cd8aec344). All studies were judged by the Research Evaluation Committee of the Psychology Department of the University of Milan-Bicocca as involving minimal risk for participants (protocol nrs. RM-2021-472, RM-2022-552, and RM-2022-572).

Study 1. Eliciting and classifying goals related to honesty and dishonesty

The aims of the study were to empirically identify a comprehensive set of goals related to honesty and to dishonesty, and to identify a set of items to assess them. Mirroring previous work (Costantini et al., 2020; Costantini & Perugini, 2018), we used personality-descriptive adjectives as the basis for goal elicitation (Amenta et al., 2025). We considered adjectives related to the positive and negative poles of four different facets of honesty-humility: (im-)modesty, greed (avoidance), (un-)fairness, and (in-)sincerity (Ashton & Lee, 2020).

Methods

Participants

Two hundred fifty participants took part in the study (*M* age = 27.6, *SD* = 8.59) through the Prolific Academic platform (https://www.prolific.co/). Of them, 123 were male, 120 female, and 7 identified as "other". The only inclusion criterion, specified in the Prolific Academic platform, was being native Italian speakers. Two participants indicated insufficient proficiency in the Italian language and were thus not allowed in the study. Five additional participants were expelled during the study, because they repeatedly failed the instruction checks (see below, Procedure)[1]. Participants received £3.75 for their participation in the study. Study 1 was fully exploratory: No *a-priori* hypothesis was tested. The sample size was thus determined in advance with the aim of eliciting sufficient breadth in participants' responses (see also Costantini & Perugini, 2018).

Procedure

The study was administered via the Qualtrics platform (www.qualtrics.com). We adapted the two-step procedure that Costantini and Perugini (2018) used to elicit goals for trait conscientiousness, which was in turn inspired by the procedures used by McCabe and Fleeson

[1] An additional pre-registered data exclusion criterion was completing the survey too quickly, less than 3 *SD* below the mean. However, the mean completion time was 30.7 minutes, with an SD of about 14.6 minutes, thereby the threshold for fast completion was negative and it could not be applied.

(McCabe & Fleeson, 2016) and by Bagozzi and colleagues (Bagozzi et al., 2003). This procedure seeks to elicit goals at different abstraction levels, including more concrete goals in the first step, and more abstract ones in the second step. Goals were elicited starting from a list of 48 person-descriptive adjectives identified in previous studies, as broadly associated with specific facets of honesty-humility or with general aspects of the trait[2] (Amenta et al., 2025). Twenty additional *control* adjectives assessing the positive and negative poles of other HEXACO traits were also included in the list (see Table 1) (Romano et al., 2023).

Before starting the procedure, participants were instructed about what was and was not to be considered a goal (see Supplementary Material S1). In the first step, participants answered the question "Why do you or would you behave in an X way?", where X was an adjective, e.g., "honest". Each participant was presented with a total of 18 adjectives randomly selected from the pool of 68: two adjectives for each of the facets fairness, modesty, and greed-avoidance, four adjectives for the facet sincerity and for general honesty, and four control adjectives, balanced by positive and negative poles. In the second step, participants were reminded of each of their answers (e.g., "in the previous questionnaire, you answered that you do or would behave in an *X* way to *Y*" - where *Y* is their previous answer) and were asked to write down a superordinate goal for each by answering to the question: "why is or could this be important to you?". Overall, each participant wrote down 36 goals.

[2] It has been argued that sincerity and truthfulness are underrepresented within popular assessments of this trait, to the point that some have proposed to rename the trait Benevolence / Simplicity (Fleeson, 2020). Studies contrasting truthful and prosocial tendencies have found mixed evidence, with some studies showing that honesty-humility is connected to truthfulness irrespective of its prosociality (Thielmann et al., 2023), and other studies finding that people with high levels of honesty-humility tend to lie for prosocial reasons (Choshen-Hillel et al., 2020; McArthur et al., 2024; Ścigała et al., 2020). For this reason, in the goal-eliciting procedure, we decided to include additional adjectives specifically related to sincerity and truthfulness (Amenta et al., 2025).

Table 1. Full list of 68 adjectives used for the goal eliciting procedure.

Trait	Selected Adjectives
Fairness	imparziale (impartial), obiettivo (objective/unbiased), giusto (just/fair), ingiusto* (unfair), prepotente* (bully), immorale* (immoral)
Greed-Avoidance	senza pretese (unpretentious), generoso (generous), non avaro (not stingy), ingordo* (greedy), venale* (mercenary), avido* (greedy)
Modesty	sobrio (sober/moderate), modesto (modest), semplice (simple), sbruffone* (boaster), immodesto* (immodest), vanitoso* (vain)
Sincerity	vero (truthful), diretto (direct), franco (frank/candid), autentico (authentic/genuine), sincero (sincere), amante della verità (truth-lover), menzognero* (lying), in malafede* (in bad faith), falso* (false), ipocrita* (hypocritical), insincero* (insincere), bugiardo* (liar)
General honesty	retto (righteous), morale (moral), equo (fair), di buon cuore (good-harted), onesto (honest), umile (humble), leale (loyal), fedele (faithful), integro (incorruptible), maligno* (evil), ingannatore* (deceiving), raggiratore* (fleecer), imbroglione* (swindler), sleale* (disloyal), egoista* (egoistic), snob* (poser), disonesto* (dishonest), altezzoso* (haughty).
Control-Emotionality	vulnerabile (vulnerable), emotivo (emotional), coraggioso* (corageous), emotivamente stabile* (stable - emotionally)
Control-Extraversion	estroverso (extraverted), vivace (vivacious), timido* (shy), introverso* (introverted)
Control-Agreeableness vs. anger	calmo (calm), tranquillo (tranquil), litigioso* (quarrelsome), collerico* (choleric)
Control-Conscientiousness	organizzato (organized), attento (attentive), disorganizzato* (disorganized), distratto* (inattentive)
Control-Openness to experience	innovativo (innovative), originale (original), tradizionalista* (traditional), non creativo* (uncreative)

Note. The 30 adjectives related to Honesty-Humility were identified in the development of the Adjective Checklist of Honesty (Study 1; Amenta et al., 2025). Control adjectives were items of the Hexaco Adjective Scales (Romano et al., 2023).
*The adjective assesses the negative pole of the trait.

To make sure that participants understood the task and paid attention while completing the questionnaire, two types of attention checks were included in the study. After reading the instructions for the first and second step, participants were asked whether the general goal of the study was about "goals of behaviors" (correct option), "causes of behaviors", "ways of behaving", or "situations in which one behaves in certain ways". Those who did not indicate the correct option were allowed to read the instructions again and answer the question a second

time. Five participants who gave the wrong answer twice were not allowed to take part in the remainder of the study. Additionally, two directed questions (Maniaci & Rogge, 2014) were included among goal answers, asking participants to write a specific word (e.g., "attento", Eng. "attentive") instead of a goal: All participants passed at least one of these attention checks, and were retained for subsequent analyses.

Results

We obtained 9000 textual answers. We analyzed the textual corpus relying on a combination of independent raters' judgment, natural language processing algorithms, network analysis, and clustering techniques that were developed in previous studies (Costantini et al., 2020; Costantini & Perugini, 2018). This analysis had two main aims: identifying goals and non-goals, and grouping goals into a limited set of classes. Relying on knowledge of which adjective elicited which response, we established an initial link between goal classes and honesty and its facets.

Identification of goals and non-goals

The first step consisted of separating goals from non-goals. Despite the instructions, some answers did not correspond to the definition of goals provided. To facilitate the task for raters, we grouped sentences based on their similarity in content and structure using Natural Language Processing (NLP), which provided quantitative representations of each sentence for classification analyses. Each of the 9000 sentences was encoded in the embeddings[3] created by the MUSE[4] (Multilingual Universal Sentence Encoder) NLP model (Yang et al., 2019), which

[3] A word-embedding is a representation of the word characteristics (meaning, syntactic properties, etc.) in the form of real-number vectors. Word embeddings can be induced using different methods which allow to map words (or sentences) from natural language to a vectorial space. The most used methods to induce embeddings (or vectorial representation, or distributed representations) are dimensionality reduction on co-occurences matrixes (e.g., Latent Semantic Analysis; Landauer & Dumais, 1997) or neural networks (e.g., word2vec; Mikolov et al., 2013) and, more recently, transformer models (e.g., USE, Cer et al., 2018; BERT, Devlin et al., 2019; see Vaswani et al., 2017).

[4] MUSE (Yang et al., 2019) is a transformer-based model trained on 16 languages (among which, Italian, the language of the current study).

encodes any short text (sentence or paragraph) in natural language into a vector of 512 dimensions[5]. The encoding process is based on distributional semantics (Turney & Pantel, 2010), guided by the distributional hypothesis (Harris, 1954), which states that words with similar distributions (i.e., which co-occur in similar contexts) have similar meanings. More recent developments in distributional semantics models allow not only to encode single words but also whole phrases and sentences. MUSE belongs to this next generation of NLP models, which is based on transformer sentence embeddings to induce contextualized vector representations of sentences. Meaning similarity can be approximated by measuring the distances of vectors in the 512-dimensional vectorial space through cosine similarity: The more similar the meaning of two sentences, the smaller the angle between them, and the higher their cosine. In the current studies, after obtaining embeddings for each one of the 9000 sentences, we computed a cosine dissimilarity matrix among them. Cosine dissimilarity was then used as input for a hierarchical clustering using the R package *fastcluster* (Müllner, 2013). The dendrogram was cut automatically using the *dynamicTreeCut* method (Langfelder et al., 2008). The algorithm identified 1664 clusters.

Five independent raters, MSc Psychology students, were instructed to classify each response as a goal or a non-goal. Answers were grouped by cluster to facilitate their task. For example, all answers in the form “I don’t know” were in the same cluster and thus presented close to each other: This reduced raters’ effort to identify them as non-goals. The five raters were trained by classifying the first 500 sentences in our corpus. According to the intraclass correlations (ICC; Shrout & Fleiss, 1979), the five raters showed an excellent agreement, ICC(2, 5) = .96. The reliability of the ratings of threerater subsetswere also very consistent, with ICCs(2,3) ranging between .92 and .95. Thus, subsets of three raters evaluated the remaining 8500 sentences This way, each sentence was evaluated by three raters. The overall

[5] For a more detailed explanation of the algorithm, please see Hussain et al. (2024)

reliability of the final ratings ranged between ICC(2, 3) = .87 and .93, indicating a very good agreement. Responses were marked as goals vs. non-goals, according to the majority judgment. Overall, raters identified 7592 goals and 1408 non-goals.

Classification of goals

The 7592 valid goals were further analyzed by asking the same five raters to group sentences in classes. Using the same NLP algorithm detailed above, we first calculated the cosine distance between each of the 7592 responses and the centroids[6] of 86 goal-classes identified in previous studies (Costantini et al., 2020, 2025; Costantini & Perugini, 2018). A sample class is *be safe*, including items such as "Safeguard Myself" and "Do not get in trouble" (Costantini et al., 2020; Costantini & Perugini, 2018). We considered the three goal classes with centroids closest to a given answer, and five raters rated whether each goal could be classified within one of these three classes or into a residual category. According to the Fleiss' K statistic (Fleiss, 1971), the raters showed moderate agreement (K = .60). Of the initial 7592 goals, 3102 (41%) were classified by the majority of raters into the first class identified by the NLP algorithm; 1010 (13%) into the second class; 480 (6%) into the third class; 2711 (36%) into the residual class. These results indicate that the raters agreed that the algorithm correctly identified the class of the responses among the three closest ones in the majority of cases (4592 cases, 60%). On the remaining 289 responses (4%) raters did not find a clear agreement: These responses were further inspected by the authors, who either assigned them to one of the classes indicated by the raters (70 goals), or classified them as not belonging to the proposed classes (203 goals; these goals were added to the 2711 goals which were classified in the residual class; see below), or classified them as being ambiguous (16 goals; ambiguous goals were discarded).

[6] Centroids are computed through a procedure of vector averaging. Therefore, for each of the 86 goal classes, we computed the centroid by averaging the vectors of the items it contained. Centroid can hence be thought as representation of the overall content of each one of the goal classes.

The 2914 goals indicated as not being clearly part of any existing class(residual class) were further inspected by the authors, who either classified them into an existing class (2064 goals), or as idiosyncratic responses not deserving further investigation (281 goals), or as novel content (569 goals). Based on this novel content we created nine new classes: “evitare di faticare” (take the easy way), “giustizia e imparzialità” (justice and fairness), “gratificazione fisica” (physical satisfaction), “mentire e nascondere” (hide and lie), “mettere alla prova” (test someone), “mettere se stesso al primo posto” (put oneself first), “sobrietà” (sobriety), “dire la verità” (tell the truth), “ottenere un vantaggio a qualsiasi costo” (get what you want at any cost). At the end, all responses were either discarded or assigned to one of the 95 (existing or newly created) classes. Thus classified, the linguistic material could be analysed to link each goal to the facets of honesty/dishonesty.

Association between goal classes and HEXACO traits

We assessed whether goals elicited by adjectives of honesty/dishonesty fell into different classes compared to goals elicited by control adjectives. The relative frequency with which specific adjectives elicited goals belonging to each of the 95 classes (the existing 86 plus the nine newly generated ones) was used for establishing a quantitative link between goal classes and honesty-humility, using a chi-square test for independence (with bootstrapped p-values). We tested the preregistered hypothesis that goals elicited by adjectives belonging to the positive and negative poles of each honesty-humility facet (including general honesty-humility, see Table 1) and five control traits fell into different classes. The results rejected the null hypothesis of no association, χ^2 (1786) = 18670, $p < .001$, suggesting that some goal classes could be associated with honesty-humility or its facets.

To identify goal classes that were sufficiently associated with any of the honesty-humility facets, including general honesty, we calculated the standardized residuals of the chi-square test. Residuals in a contingency table are computed as the difference between the

observed frequency in a cell and the frequency that would be expected in the cell if the two variables (in this case, the goal class and the poles of honesty) were independent. A positive (vs. negative) residual in a cell indicates that the observed frequency is larger (vs. smaller) than expected. Standardized residuals can be computed by dividing residuals by their standard error. Standardized residuals follow a standard normal distribution, therefore, absolute values larger than 3 indicate a highly significant lack of fit of the independence hypothesis in that cell (Agresti, 2007; see also Costantini & Perugini, 2018). We identified 46 (out of the initial 95) goal classes whose standardized residuals with respect to at least one pole of one of the facets of honesty were larger than 3, and that did not show an even larger standardized residual with any of the remaining HEXACO traits (see Table S1). In addition, we decided to take into further consideration two classes ("avoid being excluded or ignored" and "persuasion") that showed indications of a possible association with honesty-humility in a previous study (Costantini et al., 2025), but not in this one. Therefore, overall, we considered 48 goal classes instead of just 46.

To create a questionnaire assessing the subjective importance of these goal classes, we used 143 available items generated in previous studies (Costantini et al., 2020, 2025; Costantini & Perugini, 2018), related to 39 of these 48 classes. For the remaining nine newly-generated classes, items still needed to be created.

New item generation

We inspected the individual responses for each newly-generated class and generated questionnaire items representing as accurately as possible their content while removing personal and idiosyncratic elements. For example, for class *get what you want at any cost* responses like "per trarre profitto dalla paura degli altri" ("to profit from other people's fear") and "allo scopo di ottenere ciò che voglio e che non posso ottenere onestamente" ("to get what I want and I cannot obtain in an honest way") were represented by items such as "Ottenere

qualcosa illecitamente" ("Obtaining something illicitly") and "Ottenere un vantaggio a discapito di altri" ("Gaining an advantage at the expense of others"). This way, we generated 38 novel items.

For the remaining 39 goal classes, we had available sets of items from previous studies (Costantini et al., 2020, 2025). However, we inspected individual responses categorized in those classes to identify underrepresented content, generating 10 additional items (e.g., "avere la coscienza pulita", "having a clear conscience" for class "be true to one's ideals and principles"). The preliminary set of items (Table S2) included 191 items (143 from previous studies and 48 newly generated).

Discussion

The overarching aims of Study 1 were to identify a set of goals potentially relevant for honesty and dishonesty and to assemble an initial set of questionnaire items to assess them. Using a bottom-up procedure starting from personality descriptive adjectives, we developed a broad preliminary set of 191 items meant to assess 48 goals potentially related to honesty-humility. The full set of items is reported in Table S2.

Study 2. Development of the Goals for Honesty / Goals for Dishonesty (GHGD) questionnaire

The aim of Study 2 was to refine the set of goals identified in Study 1, relying on the empirical association between goals and self-reported personality, to develop the Goals for Honesty and Goals for Dishonesty (GHGD) questionnaire. Regarding the assessment of personality, sentence-based questionnaires, such as the HEXACO-PI (Lee & Ashton, 2004), offer a reliable and valid tool. However, sentence-items (e.g., "If I want something from a person I dislike, I will act very nicely toward that person in order to get it", HEAXCO-PI sincerity scale) often include content related to desires, motivations, and goals ("I want something", "in order to get it"; see also Wilt & Revelle, 2015). When assessing personality -

goal relationships, this might lead to observing associations driven by content overlap (Costantini et al., 2020). To limit this possibility, in Study 2 we decided to use an adjective-based assessment of personality (Amenta et al., 2025; Romano et al., 2023).

Methods

Participants and procedure

Three hundred participants (*M* age = 28.5, *SD* = 8.6) recruited through Prolific Academic (https://www.prolific.co) took part in the study. Of them, 145 were female, 151 male, and 4 qualified as "other". Participants were all Italian speakers and gave informed written consent before entering the study. Participants filled in the GHGD questionnaire and then the HEXACO Adjective scales. To limit order effects, the order of items within questionnaires was randomized. None of the participants were excluded based on our preregistered criterion of failing at least two attention checks (see below). The study was administered via the Qualtrics platform (www.qualtrics.com). Three hundred participants allow detecting correlations as small as $r = .161$ with 80% power at the conventional α level of .05 in a two-tailed test.

Materials

Preliminary GHGD questionnaire. Participants rated the subjective importance of each of the 191 candidate goal items identified in Study 1 on a scale from 1 (*not important to me at all*) to 5 (*very important to me*). Items were meant to assess 48 goals - corresponding to the 48 classes identified in Study 1 (see Table S2).

Adjective assessment of personality. Participants indicated the extent to which each of 96 adjectives (e.g., "honest") described them on a scale from *1* (*it does not describe me at all*) to *7* (*it describes me completely*). The adjectives were obtained from the *HEXACO Adjective Scales* (HAS; Romano et al., 2023), the *Adjective Checklist of Honesty* (ACH; Amenta et al., 2025), plus additional adjectives that were used for the goal-eliciting procedure in Study 1 (see Table 1) and that were not already part of the aforementioned measures (see also the

Supplementary Materials S3 for a minor deviation from the preregistration). These adjectives were used to assess honesty-humility (46 items, α = .93), emotionality (10 items, α = .83), extraversion (10 items, α = .91), agreeableness vs. anger (10 items, α = .86), conscientiousness (10 items, α = .87), and openness to experience (10 items, α = .78). The items of the Adjective Checklist of Honesty were used to assesses four honesty-humility facets: sincerity (10 items, α = .87), fairness (4 items, α = .64), modesty (4 items, α = .54) and greed-avoidance (4 items, α = .68; see Supplementary Material S2 for further information regarding the factorial structure of this scale).

Attention checks. To serve as an attention check, four directed questions were included, three in the GHGD questionnaire and one in the HAS, with a similar structure to the main questions of the survey but asking to select a specific value on the scale (e.g., "In this question, answer 4 and proceed with the questionnaire").

Results

Preregistered hypothesis: General associations between honesty and goals

The main preregistered hypothesis was that the first principal component scores extracted from the 46 honesty/dishonesty adjectives would correlate with the first principal component scores extracted from the 191 goal items, reflecting a general connection between honesty/dishonesty and honest/dishonest goals. The correlation between the first extracted components was r = -.46, p < .001, indicating a clear relationship between trait honesty and goals. The negative sign was due to markers of the positive pole of honesty showing a negative loading on the first principal component.

Goal structure and reliability

Once having established the association between goal items and adjectives of honesty/dishonesty, we performed additional analyses to assess the main psychometric properties (e.g., dimensionality, internal consistency) of each of the GHGD and their connections with HEXACO traits.

We first examined the dimensionality of the items within each of the 48 goals of the GHGD questionnaire by performing 48 individual parallel analyses (Horn, 1965). Forty-six goals resulted in a unidimensional structure, whereas only two of them (Hide & Lie and Justice & Fairness) showed a two-component solution. We further explored two-component solutions for these two, which suggested splitting each of them into two more specific goals (see Tables S3-S4), hence resulting in 50 goals (including Justice, Fairness, Hide, and Lie separately). In this process, two items with sizeable cross-loadings were removed from the final set (Table S3). The 50 resulting goals showed overall acceptable reliability (mean Cronbach's alpha= 0.71, SD=0.12; for all values, see Table S5).

Correlation between goals and HEXACO traits and final item selection

In order to obtain a tool easily and reliably deployable to investigate the goals related to honesty-humility, we aimed at retaining only items belonging to goals showing a sufficient association with this trait or with one of its facets. We thus estimated correlations between goals and traits (Table S6). Of the 50 goals considered, 39 showed a correlation with honesty-humility or with one of its facets that was larger than the largest correlation with any of the remaining HEXACO traits.

We inspected correlations for each goal to decide which should be retained. In the absence of an objective criterion for selecting goals, we reasoned that this task was conceptually similar to retaining the best indicators in principal component analysis (which is similarly characterized by the absence of objective criteria; Schreiber, 2021). We thus decided to mirror typical procedures adopted for variable selection in principal component analysis (e.g., Guadagnoli & Velicer, 1988; Schreiber, 2021), and retained goals showed a correlation of at least $r = .30$ with honesty-humility or one of its facets, in absolute value, if this correlation was at least 130% of the largest correlation with other HEAXCO traits. The items belonging to these goals were included in the final goals of honesty (GH) and of dishonesty (GD) scales.

Discussion

The main aims of Study 2 were to refine the set of goals that emerged as potentially relevant for honesty-humility in Study 1 and to develop the final GHGD questionnaire. Through a correlational analysis, we retained 21 goals that were related to trait honesty-humility or one of its facets, and whose relations to other HEXACO traits were weaker. Seventeen of them were positively related to honesty-humility and were denoted as Goals of Honesty (GH), whereas four were negatively related to the trait and were denoted as Goals of Dishonesty (GD). We then generated the GHGD questionnaire by including the 78 items generated to assess those goals, 64 assessing GH, and 14 assessing GD. These items can be considered as a sufficiently specific assessment of the goals of honesty-humility and should be used when one is interested in studying the motivational underpinnings of honesty-humility, without including goals that also have a strong motivational overlap with other traits. The final questionnaire is presented in Appendix 1.

Study 3. Behavioral validation of the questionnaire in self- and informant-report form

The main aim of Study 3 was to collect validity evidence for the GHGD Questionnaire and, particularly, its ability to predict participants' actual behavior. To this aim, we employed an adapted version of a computerized task called *Spot the difference*, devised by Speer and colleagues (2020). In this task, participants are told that they will receive a monetary reward based on the number of differences they report to have found in a series of image pairs, without controlling for the actual differences they find. The experiment is designed to create a temptation to lie and increase one's reward, thus creating a favorable situation for dishonest behavior. We tested whether the GHGD questionnaire would predict behavior in the expected direction, that is, a mitigating effect of GH and an enhancing effect of GD on the number of lies. In addition to observed behavior, we considered behavioral tendencies and attitudes theoretically connected to honesty-humility, in particular, lying in everyday situations,

acceptability of lying, and different aspects of authenticity (Hart et al., 2019; Oliveira & Levine, 2008; Wood et al., 2008).

We decided to extend our investigation to observers' reports. Research shows that people tend to be honest about their (dis)honesty (de Vries et al., 2014; DeVries et al., 2018; Halevy et al., 2014; Hilbig et al., 2024). Nonetheless, honesty-humility as a trait can be judged with sufficient accuracy by observers, with informant-reports of this trait being able to contribute to the prediction of actual behavior (Cohen et al., 2013). This suggests that honesty-humility, despite being an evaluative trait, might have sufficient observability (Vazire, 2010). In addition, recent research shows that people make spontaneous inferences not only regarding other people's traits, but also about their goals (Günsoy & Okten, 2022; Moskowitz & Olcaysoy Okten, 2016). However, to the best of our knowledge, no previous study has investigated whether observers can also accurately report on other people's (dis)honest goals. We planned to do so by inspecting whether informant-reports of goals align with self-reports and can predict behavior, and whether self- and informant-reports of goals align with everyday lying, as perceived by informants.

After inspecting the structure of the GHGD questionnaire with principal component analysis, we tested whether goals informant-reports of goals would converge with self-reports. We also tested whether goals would converge with trait honesty-humility and its facets, as well as with self- and informant-reported behavioral tendencies. In addition, we inspected whether goals related to honesty-humility would predict the frequency of cheating behavior observed in informant-reports, and if they would do so above and beyond personality traits, honesty-humility facets. Furthermore, to ensure that the connection with cheating behavior was sufficiently specific to honesty-humility and not related to generic motivational aspects, we inspected whether GH and GD would predict cheating behavior above and beyond goals

related to conscientiousness, which present some conceptual overlap with honesty-humility (e.g., “be trustworthy”, “comply with rules”, etc.; see Costantini and colleagues, (2020).

Methods

Participants and informants

Four hundred participants (*M* age = 29.9, *SD* = 9.1) enrolled in the study via Prolific Academic (https://www.prolific.co/). Of them, 191 were female, 198 male, and 11 non-binary. Participants were all Italian speakers and gave informed written consent before entering the study. Ten additional participants who didn’t comply with our preregistered criterion of successfully passing at least two attention checks (see Procedure) were not allowed to proceed with the behavioral task and were replaced by other participants to keep the sample to the preregistered size of N = 400. An additional preregistered exclusion criterion concerned participants who found out the rule behind the spot-the-difference task from the analyses involving that task. This entailed the exclusion of N = 239 participants from some of the analyses, which were performed on a subsample of N = 161 individuals.

Participants were invited to involve up to two informants. Overall, 216 informants took part in the study. Of them, six were excluded from the analysis because they indicated that their responses should not have been considered through the Self-Report Single Item (SRSI; Meade & Craig, 2012). One further informant was excluded because they indicated no past relationship with the person rated. Overall, we considered the responses of 209 informants (*M* age = 33.53, *SD* = 13.47), of whom 128 were females, 78 males, and 3 non-binary. Informants reported on N = 116 target participants (*M* age = 28.89, *SD* = 7.88), of whom 63 were females, 49 males, and 4 non-binary. Of them, 29 had one, 81 participants had two, and six participants had three informant reports. We retained all informant reports available, including those of participants who invited more than two informants. For participants who received more than one informant report, we averaged the informant reports received.

The full sample of N = 400 allows detecting correlations as small as $r = .139$ with 80% power at the conventional $\alpha = .05$, using a two-tailed test. The sample used for analysing the spot-the-difference task (N = 161), the full informant-report sample (N = 209), and the sample of target participants with at least one informant (N = 116) allowed detecting correlations as small as $r = .219$, $r = .192$, and $r = .257$ in the same conditions. N = 65 participants qualified for the spot-the-difference task and involved at least one informant. This sample provides sufficient power for detecting correlations of only $r = .339$ in the same conditions: when this small sample size of N = 65 was used for the analyses, we clearly marked the results as needing further empirical corroboration.

Materials

HEXACO Adjective Scales (HAS; Romano et al., 2023) and Adjective Checklist of Honesty (Amenta et al., 2025). Participants indicated the extent to which each of 79 adjectives[7] (e.g., "honest") described them on a scale from *1* (*it does not describe me at all*) to *7* (*it describes me completely*). Informants indicated the extent to which each adjective described the person who sent them the questionnaire, from *1* (*it does not describe them at all*) to *7* (*it describes them completely*). The adjectives assessed honesty-humility (29 items, $\alpha_{self} = .92$ and $\alpha_{informant} = .91$), emotionality (10 items, $\alpha_{self} = .86$ and $\alpha_{informant} = .82$), extraversion (10 items, $\alpha_{self} = .92$ and $\alpha_{informant} = .92$), agreeableness vs. anger (10 items, $\alpha_{self} = .88$ and $\alpha_{informant} = .88$), conscientiousness (10 items, $\alpha_{self} = .88$ and $\alpha_{informant} = .87$), and openness to experience (10 items, $\alpha_{self} = .82$ and $\alpha_{informant} = .74$), as well as four honesty-humility facets: sincerity (10 items, $\alpha_{self} = .90$ and $\alpha_{informant} = .84$), fairness (4 items, $\alpha_{self} = .68$ and $\alpha_{informant} = .57$), modesty (4 items, $\alpha_{self} = .60$ and $\alpha_{informant} = .63$) and greed-avoidance (4 items, $\alpha_{self} = .71$ and $\alpha_{informant} = .66$; see

[7] Notice that, unlike in Study 1, we decided to limit the assessment of honesty-humility by including adjectives that were part of the HAS or the ACH, without administering additional items.

the Supplementary Material S2 for further information regarding the factorial structure of this scale).

HEXACO-60 (Ashton & Lee, 2009). The short version of the HEXACO-PI assesses six major personality traits with 10 items each: honesty–humility (α_{self} = .90 and $\alpha_{informant}$ = .91), emotionality (α_{self} = .81 and $\alpha_{informant}$ = .67), extraversion (α_{self} = 86 and $\alpha_{informant}$ = .86), agreeableness versus anger (α_{self} = .73 and $\alpha_{informant}$ = .82), conscientiousness (α_{self} = .78 and $\alpha_{informant}$ = .79), and openness to experience (α_{self} = .76 and $\alpha_{informant}$ = .77). In our study trait honesty was investigated with a more detailed version, including 32 items from HEXACO-100 (Lee & Ashton, 2004). A sample item is "*I wouldn't use flattery to get a raise or promotion at work, even if I thought it would succeed*" (honesty-humility). Participants rated each item on a 5-point scale, from 1 (*completely disagree*) to 5 (*completely agree*).

Goals. The GHGD questionnaire developed in Study 2 was administered to obtain self and informant reports of goals for honesty and dishonesty. Participants indicated the extent to which each of 78 goal items (e.g., "obtaining something illicitly") was important to them on a scale from 1 (*not important to me at all*) to 5 (*very important to me*). Informants indicated the extent to which the same goals described their target, on a scale from 1 (*not important to him/her at all*) to 5 (*very important to him/her*). Participants and informants also rated 72 additional goal-items relevant to conscientiousness (e.g., "Completing a project" Costantini et al., 2020). To limit the length of each questionnaire page, items were divided into blocks, three including GHGD goals, and two including conscientiousness goals. Blocks and items within blocks were administered in a random order. We thus obtained self- and informant-report scores for GH (α_{self} = .94, α_{peer} = .95), GD (α_{self} = .85, α_{peer} = .87). We also obtained ratings of goals for conscientiousness (GC, α_{self} = .96, α_{peer} = .97) and goals for unconscientiousness (GU; α_{self} = 84, α_{peer} = .83).

Lies in Everyday Situations Scale (LIES, Hart et al., 2019) is a questionnaire that assesses the use of various forms of lying. Participants and informants reported their agreement with 14 items (e.g., "I lie in order to be friendly and cordial with others" – "S/he lies in order to be friendly and cordial with others") on a 7-point scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). In our study, the scale showed good reliability with Cronbach's alpha = 0.90 for self-report and α = 0.93 for informant-report.

Lies Acceptability Scale (Oliveira & Levine, 2008) represents a self-report measure of an individual's attitude towards deceptive communication (α = 0.84). The scale consists of 11 items (e.g., "The best way to handle people is to tell them what they want to hear"). Participants rated their agreement with each item on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*).

Authenticity Scale (Wood et al., 2008) is a measure of dispositional authenticity, which is designed to assess a tripartite conception of authenticity, comprising self-alienation, authentic living, and accepting external influence. For the purposes of our study, this and the following scales were back-translated into Italian by the authors, and the back-translations were checked for correctness by a native English speaker. The scale consists of 12 items (e.g., "I am true to myself in most situations"), each belonging to one of the three subscales of 4 items each: self-alienation subscale (α = 0.82), authentic living (α = 0.77), and accepting external influence (α = 0.87). Participants expressed their agreement with the items on a scale from 1 (*does not describe me at all*) to 7 (*describes me very well*).

Adapted spot-the-difference task. The task was adapted from the study by Speer and colleagues (Speer et al., 2020) and involved participants misreporting personal scores for monetary gains. The task was implemented with the PsychoPy software (Peirce et al., 2019). Participants were informed that the task aimed to investigate visual search abilities. They were presented with 72 pairs of cartoon landscape images in a "spot the difference" game, where each pair contained subtle differences such as changes in color, shape, size, or position of

objects. Participants were told that each pair had exactly three differences, and they would receive a small bonus payment if they claimed to have found all three. During each trial, participants had six seconds to observe the images and three seconds to respond by pressing a "Yes" or "No" button on the screen, with a countdown timer visible. Critically, although participants had to report finding all three differences to receive payment, they were not required to specify the locations of these differences. Unbeknownst to them, the actual number of differences varied between one and three.

The task had three conditions: normal, difficult, and very difficult. In the normal condition (50% of the trials), images had three differences that were relatively easy to spot. The difficult and very difficult conditions each constituted 25% of the trials, with these labels randomly assigned to trials containing one or two differences. Participants were informed that the trials varied in difficulty, with corresponding bonus payments: £0.01 for normal trials, £0.03 for difficult trials, and £0.05 for very difficult trials. Without cheating, the maximum reward was £0.36. By always cheating, the reward could total up to £1.80. However, the actual reward was not dependent on their performance: All participants were awarded a net bonus payment of £1.98 after the study, which was slightly more than the maximum reward that they could have obtained.

To familiarize participants with the task, five practice trials were included in which the images had three differences. In addition, following Speer and colleagues (2020), to enhance the task's credibility, participants were informed that there would be “special trials”, where participants had to click on the differing elements in the right-hand image. They had 10 seconds to make their selections. These trials ended after three clicks, regardless of accuracy. Participants knew beforehand whether a trial was an easy, normal, difficult, or a point-and-click trial.

Procedure

Through the Qualtrics platform, participants first completed the HAS and HEXACO-60, and the goals measures. Three directed questions were included in the goal measure, with a similar structure to the main questions of the survey but asking to select a specific value on the scale (e.g., "In this question, answer 4 and proceed with the questionnaire"). Participants failing two or more attention checks were not allowed to proceed further with the study and were automatically replaced by other participants through Prolific Academic.

Participants were then redirected to the Pavlovia platform (https://pavlovia.org/) to complete the spot-the-difference task, and back again to Qualtrics to complete the Authenticity Scale, LiES, and Lie Acceptability measures. These scales were administered after the spot-the-difference task, because they explicitly dealt with the concepts of lying and dishonesty, whereas we didn't want to give to the participants any clue of how the experiment could be interpreted. After a short demographic questionnaire, participants were given the possibility of inviting informants to complete the observer-report questionnaires. Participants were instructed to copy a personalized link provided in the questionnaire and send it to two potential informants, who knew them well and who were proficient in Italian.

At the end of the informant invitation part, participants were informed that in the spot-the-difference task, there was one disguising aspect. They were asked whether they thought that they knew what this aspect was and, if so, to briefly indicate it in an open-ended question. N = 274 participants responded that they found the deception. After the study, two independent raters (AG and GC) determined that 239 participants understood the deception and 35 did not (Cohen's K = .84, 264 agreements and 10 disagreements, which were resolved by discussion). Some of those who did not understand deception noticed the lack of control over performance (e.g., "You could have answered yes even without finding all the differences"), but this was a feature of the task, and it was not used as an exclusion criterion. At the end of the study,

participants could download a debriefing document, approved by the ethics committee, that clearly explained the goals of the study, including the spot-the-difference task.

Informants completed peer-report versions of the HAS, HEXACO-60, goal questionnaire, and LiES. They were instructed to provide responses about the person who sent them the questionnaire, in the absence of this person. They were also asked how well they believed they knew this person on a scale from 0 (*not at all*) to 10 (*extremely well*). Informants responding "not at all" to this question were excluded from the analyses. At the end of the questionnaire, the SRSI question (Meade & Craig, 2012) was presented, asking to self-evaluate if the responses given could be used for the research. Informants responding "no" to this question were also excluded from the analyses.

Participants were compensated with £4.2, plus a bonus of £1.98 connected to the spot-the-difference task. In addition, they received a bonus payment of £1 if they identified at least two informants who completed the questionnaire.

Results

Component structure of goals

We first inspected the structure of the self-reported (N = 400) and informant-reported (N = 209) goals related to honesty, using principal component analysis, as implemented in the R package *psych* (Revelle, 2016). The first eigenvalues were 9.18, 2.54, 1.69, and 0.98 for self-reports and 10.12, 2.88, 1.36, and 0.93 for informant reports. Parallel analysis (Horn, 1965) indicated that the 95th percentile of the same eigenvalues obtained from random data were 1.43, 1.36, 1.30, and 1.25 for self-reports and 1.61, 1.50, 1.42, and 1.35 for informant-reports, thus suggesting that three components in self-reports and two in informant-reports explained more variance than their random counterparts. Adopting a conservative approach, we inspected a two-component solution for both self- and informant-report data. The results, after oblimin rotation, indicated that the two factors could be readily interpreted as GH and GD (see Table S8 for component loadings), showing only a modest correlation with each other (*r*

= -.27 in self-reports and -.24 in informant-reports). The Tucker's congruence coefficients for the self- and informant-reports were 1.00 and .96, respectively, for GH and GD, indicating that the two components that emerged in self- and informant-reports could be considered equal (Lorenzo-Seva & ten Berge, 2006).

Convergence between self- and informant-reports of goals

Table 2 shows the convergence between self-reports and informant-reports of goals, averaged across informants when more than one informant rated a single individual (N = 116). Participants and informants showed a moderate agreement regarding GH (r = .30, p = .001) and GD (r = .48, p < .001). Similar levels of agreement emerged for goals related to conscientiousness. For each goal type, such correlations were larger than those among different goals rated by participants and informants. Nonetheless, large correlations emerged in self-reports and in informant-reports among GH and GC, and among GH and GU, indicating that, according to both participants and informants, goals related to honesty and conscientiousness are related.

We further inspected whether the self-informant agreement in judging goals of honesty and dishonesty improved with acquaintance. To this aim, we fitted two linear mixed models, one for GH and one for GD, in which the absolute value of the difference between self- and informant-reported goals was predicted by the level of acquaintance, including a random intercept per target to account for the fact that some participants were rated by multiple informants. The results indicated that acquaintance was a marginally significant predictor of more accuracy in judging GH (β = -.073, p = .055) but not GD (β = .005, p = .833). A similar analysis for goals related to conscientiousness revealed that acquaintance did not predict accuracy (ps > .477).

Table 2. *Study 3. Pearson's correlations among goals rated by participants and informants.*

	GH	GD	GC	GU	GH (i)	GD (i)	GC (i)
GH	1						
GD	-.27**	1					
GC	.71***	-.04	1				
GU	.01	.44***	.04	1			
GH (i)	.30**	-.23*	.26**	-.11	1		
GD (i)	-.08	.48***	.08	.21*	-.25**	1	
GC (i)	.20*	-.10	.34***	-.17	.79***	-.17	1
GU (i)	-.07	.32***	-.06	.47***	-.04	.58***	-.01

Note. Correlations were estimated on N = 116 individuals who were rated by at least one informant. GH = Goals of Honesty, GD = Goals of Dishonesty, GC = Goals of Conscientiousness, GU = Goals of Unconscientiousness, (i) = informant report.
* $p < .05$, ** $p < .01$, *** $p < .001$

Convergence with personality traits and facets

We inspected whether goals, both reported by participants and informants, converged with personality traits and facets, both in self-reports (Table S9) and informant-reports (Table S10). Irrespective of the source (participant or informant) and of the measure (HAS or HEXACO-PI-R), GH showed positive and significant correlations with honesty-humility (r = .61 with HAS and r = .35 with HEXACO-PI-R for self-reports; r = .50 with HAS and r = .31 with HEXACO-PI-R for informant-reports, all ps < .001), and these correlations were always larger than between GC and the corresponding honesty-humility measures (r = .48, p < .001 with HAS and r = .18, p < .001 with HEXACO-PI-R for self-reports; r = .33, p < .001 with HAS and r = .13, p = .18 with HEXACO-PI-R for informant-reports). Similarly, GD showed negative correlations with honesty-humility (r = -.45 with HAS and r = -.68 with HEXACO-PI-R for self-reports; r = -.63 with HAS and r = -.70 with HEXACO-PI-R for informant-reports, all ps < .001), and these correlations were larger than those observed for GU (r = -.15, p = .003 with HAS and r = -.25, p < .001 with HEXACO-PI-R for self-reports; r = -.39, p < .001 with HAS and r = -.49, p < .001 with HEXACO-PI-R for informant-reports). In addition, correlations of GH and GD with honesty-humility were generally larger than for other traits,

with the exception of informant-reported GH, which, only for the HEXACO-PI-R, showed larger correlations with extraversion (r = .42) and conscientiousness (r = .41) than with honesty-humility (r = .31, all ps < .001). GH and GD correlated with all honesty-humility facets in the expected direction, suggesting that GH and GD involved motivations that were not specific to a single facet, but represented well the entire spectrum of motivations underlying honesty-humility. The only exception was greed-avoidance assessed via the HEXACO-PI-R, which aligned only with GD (r = -.36, p < .001 for self-reports and r = -.41, p < .001 for informant-reports) but not with GH (r = .06, p = .24 for self-reports and r = .12, p = for informant-reports).

Convergence with self-reported and informant-reported behavioral tendencies

As shown in Table 3, self-reported GH were positively associated with Authentic Living and negatively associated with the acceptability of lies and the self-reported LIES scale, indicating that individuals who endorse honesty-related goals also report behaving in more authentic and honest ways. Informant-reported GH showed a weaker but significant association with self-reported Authentic Living, in the expected directions. In addition, informant-reported GH showed significant negative correlations with both self-reported and informant-reported LIES scores, as well as with acceptability of lies.

In contrast, GD, when self-reported, were associated with all facets of authenticity in the expected direction, particularly with Authentic-Living and Self-Alienation, as well as with stronger endorsement of lies acceptability and more frequent self- and informant-reported lying behavior. Informant-reported GD also correlated positively with lie acceptability and with both self- and informant-reported LIES scores, indicating that dishonest goals are observable to others and relate to informants' perceptions of dishonesty. Interestingly, self-reported goals GU showed even stronger associations than GD with negative facets of authenticity, while informant-reported GU showed strong correlations with self-reported lying-related behaviors,

exceeding even those of informant-reported GD. These patterns suggest that unconscientiousness goals may serve as salient cues for dishonesty in the eyes of informants. Importantly, the LIES scale was the only criterion for which both self- and informant-reports were available, and these two versions correlated substantially (r = .49, p < .001), supporting the validity of informant-based assessments.

Table 3. *Study 3. Pearson's correlations between self- and informant-reports of goals, and self- and informant-reports of behavioral tendencies related to honesty and lying.*

	GH	GD	GC	GU	GH (i)	GD (i)	GC (i)	GU (i)
Authentic Living	.52***	-.23***	.44***	-.07	.19*	.05	.18	.00
Self-alientation	-.04	.25***	-.04	.39***	-.17	.18	-.18	.25**
Accepting external influence	.04	.10*	.04	.28***	-.07	.09	-.13	.15
Lies Acceptability	-.33***	.51***	-.19***	.31***	-.27**	.34***	-.19*	.37***
LIES	-.26***	.46***	-.17***	.30***	-.27**	.28**	-.23*	.35***
LIES (i)	-.11	.41***	-.14	.13	-.33***	.50***	-.30**	.36***

Note. Correlations involving self-report data only were estimated on the full sample of N = 400 participants. Correlations involving informant-reports were estimated on a sample of N = 116 individuals who were rated by at least one informant.
GH = Goals of Honesty, GD = Goals of Dishonesty, GC = Goals of Conscientiousness, GU = Goals of Unconscientiousness. LIES = Lying in Everyday Situations scale. (i) = informant-report.
* $p < .05$, ** $p < .01$, *** $p < .001$

Prediction of cheating behavior in the spot-the-difference task

We tested the main preregistered hypothesis that self-reported GH and GD would predict observed cheating behavior in the spot-the-difference task, using a Poisson regression model (N = 161), predicting the number of cheating instances from self-reported GH and GD. Predictors were always standardized (z-scored) before entering the regression model to improve the interpretability. We reported Nagelkerke's pseudo R-squared (R_N^2) for each model (Nagelkerke, 1991). When included in the model as individual predictors, GH predicted less frequent cheating (b = -.072, 95% CI = [-0.112; -0.032], $p < .001$, $R_N^2 = .073$), and GD predicted more frequent cheating (b = .193, 95% CI = [0.156, 0.229], $p < .001$, $= .447$), as expected. When GD and GH were included together as predictors, GD (b = .191, 95% CI = [0.152, 0.229], $p < .001$) but not GH (b = -.005, 95% CI = [-0.047, 0.037], $p = .803$) predicted

cheating ($R^2_N = .448$). These results suggest that the portion of GH that predicts cheating was the one shared with GD, despite the negative correlation between GH and GD being moderate (r = -.34, $p < .001$ in the subsample of N = 161 considered for this analysis and r = -.32, $p < .001$ in the entire sample of N = 400 individuals).

We also carried out a series of additional Poisson regression models (Table 4)[8] to inspect whether cheating frequency would be predicted by personality traits and goals of conscientiousness (Step 1) and whether goals of honesty-humility would predict cheating frequency above and beyond personality traits, honesty-humility facets, and goals of conscientiousness (Step 2). According to both the HAS (Model 1 in Table 4) and the HEXACO-PI (Model 2 in Table 4), honesty-humility and emotionality predicted less cheating, and extraversion predicted more cheating. Agreeableness vs. anger, according to the HAS, and openness to experience, according to the HEXACO-PI, also predicted less cheating. GD always explained cheating frequency significantly above and beyond HEXACO traits. Interestingly, the effects of honesty-humility assessed with the HAS waned (from $b = -.090, p = .004$ to $b = .00009, p = .998$) after controlling for Goals; The effects of honesty-humility assessed with the HEXACO-PI remained significantly different from zero but decreased (from $b = -.212, p < .001$ to $b = -.171, p < .001$), indicating that GD capture a portion of honesty-humility variance related to actual dishonest behavior.

Models 3 and 4 mirrored Models 1 and 2, but honesty-humility was assessed at the facet level. According to the ACH, modesty predicted less cheating. According to the HEXACO-PI, greed avoidance and sincerity predicted less cheating, but fairness, somehow surprisingly, predicted more cheating. This counterintuitive effect likely reflects multicollinearity among facets: When HEXACO-PI Fairness was entered as the sole predictor of cheating, a relationship

[8] A version of Table 4 with model fitted on the entire sample of N = 400 participants can be found in Table S11.

with cheating was in the expected negative direction (b = -.100, p < .001). Crucially, GD always predicted more cheating, even after controlling for honesty-humility at the facet level.

Model 5 revealed that goals related to conscientiousness (GC) predicted less cheating and goals related to unconscientiousness (GU) predicted more cheating. Interestingly, GD outpredicted goals relevant for conscientiousness, and the effects of GC and GU waned after controlling for GH and GD, indicating that the effect of GC/GU was due to a portion of variance overlapping with GH and GD.

As a final follow-up analysis, we examined whether informant-reports of GD would predict observed cheating behavior. The results suggested that, when included as the sole predictor, informant-reported GD predicted more cheating, as expected (b = 0.095, p = .006). When included in a multiple Poisson regression model together with self-reported GD, only self-reported (b = 0.219, p < .001) but not informant-reported (b = -.059, p = .200) GD predicted cheating frequency. However, this analysis could be carried out considering only N = 65 participants who fulfilled the inclusion criteria for the spot-the-difference task and involved at least one informant. Given the limited sample size, these results should be interpreted with prudence.

Overall, current analyses revealed that goal measures provided substantial incremental validity beyond trait assessments. GD remained a significant predictor of actual cheating behavior after controlling for all trait-level and facet-level personality measures, whereas the predictive effects of trait Honesty-Humility were either eliminated or substantially reduced when motivational constructs were included in the models. This pattern was consistent across different operationalizations of trait Honesty-Humility and remained robust when put against GC. The hierarchical regression approach demonstrated that goals captured variance in observed behavior that stable trait descriptions did not account for. These findings indicate that comprehensive behavioral prediction requires assessment across multiple levels of analysis,

with motivational processes appearing to operate as more proximal predictors of specific behavioral outcomes than broader dispositional traits.

Table 4. *Study 3. Multiple Poisson regression models predicting observed cheating behavior in the Spot-the-difference task.*

	B coefficient Step 1	*B coefficient Step 2*
		Model 1
Intercept	2.546***[2.501, 2.590]	2.511***[2.464, 2.557]
H (HAS)	-0.090**[-0.152, -0.029]	0.000 [-0.074, 0.074]
E (HAS)	-0.146***[-0.195, -0.097]	-0.113***[-0.163, -0.062]
X (HAS)	-0.088***[0.043, 0.133]	0.117***[0.071, 0.163]
A (HAS)	-0.059*[-0.115, -0.002]	0.004 [-0.054, 0.064]
C (HAS)	0.033 [-0.023, 0.088]	0.024 [-0.032, 0.081]
O (HAS)	-0.008 [-0.053, 0.037]	-0.005 [-0.052, 0.042]
GH		-0.026 [-0.082, 0.030]
GD		0.212***[0.168, 0.256]
R_N^2	.417	.656
		Model 2
Intercept	2.520***[2.474, 2.565]	2.514***[2.468, 2.559]
H (HEXACO-PI-R)	-0.212***[-0.259, -0.164]	-0.171***[-0.229, -0.112]
E (HEXACO-60)	-0.071**[-0.118, -0.023]	-0.066**[-0.116, -0.016]
X (HEXACO-60)	0.096***[0.049, 0.144]	0.097***[0.047, 0.147]
A (HEXACO-60)	-0.020 [-0.066, 0.027]	-0.011 [-0.057, 0.036]
C (HEXACO-60))	-0.004 [-0.054, 0.045]	0.016 [-0.036, 0.068]
O (HEXACO-60)	-0.070**[-0.114, -0.025]	-0.074**[-0.120, -0.028]
GH		0.045 [-0.005, 0.095]
GD		0.101***[0.048, 0.153]
R_N^2	.648	.683
		Model 3
Intercept	2.546***[2.501, 2.590]	2.515***[2.468, 2.561]
SI (ACH)	-0.093 [-0.224, 0.038]	-0.043 [-0.173, 0.087]
FA (ACH)	0.077 [-0.067, 0.220]	0.082 [-0.065, 0.230]
MO (ACH)	-0.091*[-0.166, -0.015]	-0.078*[-0.152, -0.002]
GR (ACH)	-0.013 [-0.083, 0.058]	-0.032 [-0.104, 0.039]
E (HAS)	-0.148***[-0.199, -0.098]	-0.105***[-0.158, -0.052]
X (HAS)	0.080***[0.034, 0.127]	0.116***[0.068, 0.164]
A (HAS)	-0.060*[-0.114, -0.006]	0.026 [-0.030, 0.083]
C (HAS)	0.030 [-0.027, 0.087]	0.033 [-0.025, 0.092]
O (HAS)	-0.014 [-0.063, 0.035]	-0.011 [-0.061, 0.039]
GH		-0.012 [-0.067, 0.043]
GD		0.211***[0.167, 0.254]
R_N^2	.552	.722
		Model 4
Intercept	2.507***[2.460, 2.552]	2.497***[2.450, 2.543]
SI (HEXACO-PI-R)	-0.141***[-0.197, -0.085]	-0.103***[-0.163, -0.043]
FA (HEXACO-PI-R)	0.092**[0.036, 0.149]	0.120***[0.063, 0.178]
MO (HEXACO-PI-R)	0.023 [-0.031, 0.076]	0.046 [-0.011, 0.104]
GR (HEXACO-PI-R)	-0.236***[-0.286, -0.186]	-0.233***[-0.284, -0.182]
E (HEXACO-60)	-0.152***[-0.203, -0.100]	-0.137***[-0.191, -0.083]
X (HEXACO-60)	0.021 [-0.028, 0.070]	0.032 [-0.019, 0.083]
A (HEXACO-60)	-0.067**[-0.115, -0.018]	-0.064*[-0.112, -0.015]
C (HEXACO-60))	-0.082**[-0.136, -0.029]	-0.058*[-0.114, -0.002]
O (HEXACO-60)	-0.084***[-0.129, -0.038]	-0.085***[-0.132, -0.038]

GH			0.024 [-0.027, 0.076]
GD			0.136***[0.082, 0.188]
	R_N^2	.780	.812
		Model 5	
Intercept		2.561 [2.518, 2.604]	2.547 [2.503, 2.591]
GC		-0.089***[-0.131, -0.046]	-0.070 [-0.140, 0.000]
GU		0.088***[0.043, 0.133]	-0.010 [-0.061, 0.040]
GH			0.052 [-0.018, 0.121]
GD			0.198***[0.153, 0.242]
	R_N^2	.158	.461

Note. The models were fitted to the dataset with N = 161 individuals who fulfilled the inclusion criteria for the spot-the-difference task. In all models, cheating frequency in the Spot-the-difference task is the dependent variable. In Step 1, a different set of predictors is entered in each model. Model 1 = HEXACO-traits assessed with the HAS (Romano et al., 2023); Model 2 = HEXACO-traits assessed with the HEXACO-PI (Lee & Ashton, 2004); Model 3 = HEXACO traits other than Honesty-Humility assessed with the HAS, plus facets of Honesty-Humility assessed with ACH (Amenta et al., 2025); Model 4 = HEXACO traits other than Honesty-Humility, plus Honesty-Humility facets assessed with the HEXACO-PI; Model 5 = Goals of Conscientiousness (GC) and Unconscientiousness (GU; Costantini et al., 2020). In Step 2, Goals of Honesty (GH) and Dishonesty (GD) are included as additional predictors in all models. For each model, we also report R_N^2 = Nagelkerke's (1991) R-squared index.
GH = Goals of Honesty, GD = Goals of Dishonesty, GC = Goals of Conscientiousness, GU = Goals of Unconscientiousness, H = Honesty-Humility, E = Emotionality, X = eXtraversion, A = Agreeableness vs. Anger, C = Conscientiousness, O = Openness to experience, SI = Sincerity, FA = Fairness, MO = Modesty, GR = Greed-Avoidance.
95% Confidence intervals for each parameter are reported in square brackets.
* $p < .05$, ** $p < .01$, *** $p < .001$
Statistical significance of individual regression coefficients was assessed using z-tests.

Discussion

The main goal of Study 3 was to gather comprehensive validity evidence for the GHGD questionnaire by examining self- and informant-reports of goals, their convergence with personality traits and facets, and their associations with honesty-related behavioral tendencies, including objectively observed cheating behavior. Principal component analyses confirmed that Goals of Honesty (GH) and Goals of Dishonesty (GD) are distinguishable and only modestly negatively correlated constructs, emerging consistently across both self- and informant-report data. These findings support the idea that goals for behaving dishonestly are not merely the inverse of goals for honesty-humility, but reflect a distinct motivational tendency (Costantini et al., 2020; Costantini & Perugini, 2018; Richetin et al., 2011).

Moderate agreement between self- and informant-reports of GH and GD further indicated that close others can form reasonably accurate impressions of individuals' honesty-related motivations (Carlson et al., 2022; Moskowitz & Olcaysoy Okten, 2016). This convergence increased marginally with acquaintance for GH but not for GD, suggesting that dishonest motivations may be more difficult to observe or acknowledge. Both GH and GD also showed consistent and theoretically meaningful associations with honesty-humility and its facets, which were stronger than those of goals related to conscientiousness, underscoring the discriminant and convergent validity of the GHGD scales.

Importantly, GH and GD were related not only to self-perceptions but also to behavioral tendencies associated with authenticity and deception, including informant-perceived lying behavior. While both self- and informant-reported GD were significantly associated with lying-related tendencies, informant-reported GU sometimes emerged as an even stronger correlate, suggesting that observers may use a broader range of cues to infer dishonesty.

Critically, both GH and GD predicted observed dishonest behavior in the Spot-the-Difference task when included as sole predictors. When GH and GD were included in the same regression model, only GD remained a significant predictor, outperforming also personality traits, facets, and goals related to conscientiousness. Actual dishonest behavior seems to be uniquely related to the importance of goals related to dishonesty, but not to the low importance of goals related to honesty, which further corroborates the idea that GH and GD do not represent the opposite poles of a continuum, but rather two distinguishable categories. GH were instead more strongly related to authentic living, a facet of authenticity that reflects the tendency of behaving and expressing oneself (including one's beliefs, values, and emotions).

In the prediction of observed cheating behavior, it is also noteworthy that when traits and facets were assessed using the HEXACO-PI-R rather than adjective measures such as the HAS and the ACH, the incremental predictive power of GD was slightly lower. This arguably

reflects the fact that phrased items, such as those included in the HEXACO-PI-R, compared to adjective-based measures, contain more information about the motivations of people's behavior and, in particular, dishonest goals (Amenta et al., 2025; Romano et al., 2023; Wilt & Revelle, 2015), thus presenting a higher degree of content overlap with the items assessing the importance of goals.

A key limitation of Study 3 is that a substantial number of participants (N = 239) had to be excluded from analyses involving the spot-the-difference task because they correctly identified the deception behind it. Although such exclusion was pre-registered and justified to preserve the validity of the task, it nevertheless reduced the effective sample size and statistical power for behavioral analyses. Importantly, key findings regarding the predictive value of goals—particularly GD—remained robust even within this smaller sample. In addition, replicating the analyses on the entire sample (N = 400, see Table S11) confirmed that GD consistently outpredicted personality traits, facets, and other goals. Together, these findings provide strong multi-source, multi-method support for the construct validity of the GHGD questionnaire.

General discussion

The present work set out to uncover the motivational underpinnings of trait honesty-humility, a core dimension in the HEXACO model associated with fairness, sincerity, modesty, and the avoidance of greed and exploitation (Ashton et al., 2004; Ashton & Lee, 2007, 2020). While prior research has linked honesty-humility to a wide range of prosocial and moral behaviors (Hilbig et al., 2013; Thielmann et al., 2020; Zettler et al., 2020), the specific goals that guide or reflect this trait have remained largely uncharted. The overarching aim of this project was thus to develop and validate a comprehensive self-report questionnaire—the GHGD (Goals for Honesty and Goals for Dishonesty)—to assess individual differences in the subjective importance of honesty- and dishonesty-related goals. Across three studies, we

adopted a multi-method, bottom-up approach to identify, refine, and validate the structure and function of these motivational constructs.

In Study 1, we used a bottom-up approach to elicit goal content from participants, grounded in trait-descriptive adjectives covering the breadth of honesty-humility facets (Amenta et al., 2025; Romano et al., 2023). Participants generated thousands of goals in response to trait markers, which were subsequently analyzed using a combination of natural language processing and human ratings to classify the responses into interpretable goal classes. Importantly, a subset of these goals showed robust associations with honesty-humility adjectives, supporting their conceptual relevance. These findings extend earlier work on conscientiousness (Costantini et al., 2020; Costantini & Perugini, 2018) and illustrate the feasibility and richness of using bottom-up procedures to tap into trait-linked motivational structures. **Study 2** focused on refining the GHGD instrument by evaluating the psychometric properties of the elicited items and their empirical associations with personality traits, as measured via adjective-based trait assessments (Romano et al., 2023). We retained a set of 78 items—64 assessing goals related to honesty (GH) and 14 assessing goals related to dishonesty (GD). These items formed internally consistent scales and were associated with honesty-humility, with weaker associations with other HEXACO traits. **Study 3** provided a comprehensive validation of the GHGD using multiple sources and methods. The GH and GD scales converged with multiple measures of honesty-humility, as well as with behavioral tendencies connected to honesty-humility (Hart et al., 2019; Wood et al., 2008), including self-reports, informant-reports, and behavior observed in a computerized task (Speer et al., 2020). Our results align with prior findings that goals are central elements of personality and can even outperform personality in predicting behaviors (Buchinger et al., 2024; McCabe & Fleeson, 2012; Mischel & Shoda, 1995). Furthermore, informant-reports of GD also predicted observed dishonesty, although self-reports remained the stronger predictors—suggesting that while

dishonest goals may be partially observable, individuals retain privileged insight into their own motivational profiles (Thielmann et al., 2017; Vazire, 2010). Altogether, these findings provide robust evidence that the GHGD questionnaire offers a valid and nuanced tool for assessing the motivational underpinnings of honesty-humility. The GHGD thus represents a promising measure for investigating how honesty-humility is instantiated, expressed, and potentially changed through individual goal structures.

While a substantial body of research has examined dishonest behavior and its situational antecedents, relatively few studies have focused on the personal goals that motivate honesty and dishonesty as enduring individual differences. Much of the existing literature has examined deception from a behavioral or communicative standpoint. For instance, DePaulo and colleagues (1996) conducted a foundational investigation into the everyday motivations for lying, identifying goals such as protecting oneself or others. Similarly, Levine and colleagues (2016) explored deceptive goals in interpersonal contexts across cultures, emphasizing functional motives such as covering up personal transgressions, getting personal advantages, avoiding other people, or social utility. While these studies provide a rich taxonomy of immediate or context-bound deceptive goals, they were not designed to capture stable, trait-like motivational tendencies, nor were they linked systematically to honesty-humility. In contrast, the current work offers a trait-anchored, psychometrically validated framework for understanding individual differences in goals related to honesty and dishonesty. The GHGD questionnaire measures the subjective importance of these goals across contexts, distinguishing between people who are chronically motivated to act honestly or dishonestly, rather than examining the situational triggers of specific deceptive acts.

Another relevant line of research has focused on goal inference and moral evaluation. Studies have shown that people often assess others' morality not just based on their behavior, but based on their inferred intentions and goals. For example, Carlson and colleagues (2022)

found that people judge others more harshly when they perceive their dishonest actions as goal-driven by self-interest, rather than by benevolence. Similarly, Wilson (2018) and Miller and colleagues (2021) argue that the moral evaluation of behavior depends crucially on the motives attributed to the actor, with deceptive behaviors sometimes being perceived as secondary to higher-order moral goals (E. E. Levine, 2022; Lupoli et al., 2018). Our work complements these findings by showing that people differ systematically in the goals they prioritize, and that these goals not only shape behavior but are observable by others and predictive of moral conduct.

An interesting finding is that goals related to honesty (GH) and goals related to dishonesty (GD) are **not two ends of a single continuum**, but rather **psychometrically distinct constructs** that reflect separable motivational orientations. Principal component analyses supported a two-factor structure, with GH and GD loading on different components and showing only a modest negative correlation ($r \approx -.25$). This pattern held across self- and informant-ratings, corroborating the idea that honest and dishonest goals represent qualitatively different psychological constructs rather than polar opposites. In line with this distinction, GH and GD predicted **different behavioral outcomes**. GH was mainly associated with authentic living, suggesting the view that honest goals support the enactment of a coherent and self-congruent identity (Maltby et al., 2012). In contrast, GD predicted greater endorsement and enactment of deceptive behavior, including lying in everyday life and actual cheating in a behavioral task. These results point to a dual-motive structure, whereby individuals may simultaneously endorse—or reject—goals associated with honesty and dishonesty to varying degrees. This dual structure is consistent with broader theoretical frameworks that posit coexisting and partly independent motivational systems, such as approach vs. avoidance motivation (Elliot & McGregor, 2001) or agentic vs. communal orientations (Abele & Wojciszke, 2014). Our findings contribute to this literature by showing that honesty-related

motivations also reflect such dual architecture, with implications for understanding moral behavior as the outcome of competing goals rather than the absence of immoral tendencies.

A key methodological strength of our approach lies in the bottom-up procedures used to develop the GHGD questionnaire. Rather than relying on predefined theoretical assumptions about what constitutes a goal related to honesty or dishonesty to generate items, we began with a broad elicitation of participants' own goal statements, prompted by trait-descriptive adjectives, which were in turn identified in a previous study using expert raters (Amenta et al., 2025). This approach helped to minimize researcher bias, as the initial pool of goal content was entirely participant-generated. As such, the resulting items are arguably more likely to capture ecologically valid, personally meaningful goal content (see also Condon et al., 2020; Costantini & Perugini, 2018). For example, goals such as “getting what you want at any cost” or “be true to one's ideals and principles” emerged organically from participants’ responses and reflect psychologically rich nuances in how people conceptualize (dis)honest motivation. This method is highly adaptable: Because it relies on lexical prompts and participant input, it can be extended to different traits, populations, or cultural contexts. This flexibility makes it a promising platform for future research aiming to map the motivational substrates of other personality constructs. An additional innovation of this project lies in the integration of natural language processing (NLP) techniques with human judgment. We combined sentence embeddings (Wulff & Mata, 2025; Yang et al., 2019) with hierarchical clustering to structure and reduce the vast corpus of textual responses. This computational step enabled raters to classify goals efficiently and consistently, while preserving the richness and variability of the original content. The synergy between automated text analysis and expert human interpretation not only ensured high inter-rater agreement and interpretability but also represents a scalable and replicable method for psychometric construction from open-ended data. This hybrid

method advances existing practices in personality scale development and offers a compelling model for future research aiming to bridge data-driven discovery with measurement.

While the current research offers a comprehensive and methodologically innovative examination of goals related to honesty-humility, several limitations should be acknowledged, alongside directions for future inquiry. First, in Study 3, pre-registered exclusion criteria led to the exclusion of a large portion of the sample in the prediction of the adapted spot-the-difference task. Although a similar pattern of results was confirmed in the full sample, this prompts a reflection about the potential limitations of our adaptation of this task, which was originally devised for MRI studies (Speer et al., 2020). Future studies are needed to better adapt this task to the needs of behavioral studies, for example, by reducing the presentation times until the deception becomes more difficult to recognize, or by testing the ability of the GHGD questionnaire to predict cheating behavior using different tasks that do not necessarily involve deception (Hilbig et al., 2024; e.g., Schild et al., 2021). Second, although informant-report and observed data were incorporated, the study design remained cross-sectional. As such, whereas our studies corroborate the centrality of goals for honesty-humility, it remains unclear whether the personal importance of goals fosters greater trait honesty-humility over time (McCabe & Fleeson, 2016), or whether trait levels shape one's motivational architecture (McCrae & Sutin, 2018). Addressing this question will require longitudinal or experience sampling designs, which can examine how personality and goals co-evolve in everyday life (Buchinger et al., 2024; McCabe & Fleeson, 2016).

Conclusions

This work set out to address a central but underexplored question in personality psychology: What are the personal goals that underlie honesty-humility? Across three studies, we developed and validated the GHGD questionnaire, a comprehensive tool to assess individual differences in the subjective importance of honesty- and dishonesty-related goals.

Using a bottom-up, participant-driven methodology and combining NLP and human judgment, we identified a rich set of goals. These goals were shown to be reliably associated with honesty-humility, to display a clear two-factor structure, and to predict a range of self-reported, informant-reported, and observed behaviors. Importantly, our findings suggest that goals for honesty and dishonesty are distinct motivational constructs. Taken together, the present research contributes to bridging the gap between trait structure and motivational process, providing both theoretical insights and practical tools for future research on personality and moral behavior.

References

Abele, A. E., & Wojciszke, B. (2014). Communal and Agentic Content in Social Cognition. In *Advances in Experimental Social Psychology* (Vol. 50, pp. 195–255). Elsevier. https://doi.org/10.1016/B978-0-12-800284-1.00004-7

Agresti, A. (2007). *An introduction to categorical Data Analysis*. John Wiley & Sons.

Allen, T. A., & DeYoung, C. G. (2017). Personality neuroscience and the Five Factor Model. In T. A. Widiger (Ed.), *The Oxford handbook of the Five Factor Model* (pp. 1–63). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199352487.013.26

Amenta, S., Galkina, A., Romano, D., Perugini, M., & Costantini, G. (2025). *A lexical examination of the facets of Honesty-Humility: The Adjective Checklist of Honesty*. OSF. https://doi.org/10.31234/osf.io/yv7un_v1

Ashton, M. C., & Lee, K. (2005). Honesty-Humility, the big five, and the five-factor model. *Journal of Personality*, *73*(5), 1321–1353. https://doi.org/10.1111/j.1467-6494.2005.00351.x

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review*, *11*(2), 150–166. https://doi.org/10.1177/1088868306294907

Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A Short Measure of the Major Dimensions of Personality. *Journal of Personality Assessment*. https://www.tandfonline.com/doi/full/10.1080/00223890902935878

Ashton, M. C., & Lee, K. (2020). Objections to the HEXACO model of personality structure—And why those objections fail. *European Journal of Personality*, *34*(4), 492–510. https://doi.org/10.1002/per.2242

Ashton, M. C., Lee, K., Perugini, M., Szarota, P., Di Blas, L., Boies, K., & De Raad, B. (2004). A Six-Factor Structure of Personality-Descriptive Adjectives: Solutions From Psycholexical Studies in Seven Languages. *Journal of Personality and Social Psychology*, *86*(2), 356–366. https://doi.org/10.1037/0022-3514.86.2.356

Atherton, O. E., Grijalva, E., Roberts, B. W., & Robins, R. W. (2020). Stability and Change in Personality Traits and Major Life Goals from College to Midlife. *Personality and Social Psychology Bulletin*. https://doi.org/10.1177/0146167283093007

Bagozzi, R. P., Bergami, M., & Leone, L. (2003). Hierarchical representation of motives in goal setting. *The Journal of Applied Psychology*, *88*(5), 915–943. https://doi.org/10.1037/0021-9010.88.5.915

Baumert, A., Schmitt, M., Perugini, M., Johnson, W., Blum, G., Borkenau, P., Costantini, G., Denissen, J. J. A., Fleeson, W., Grafton, B., Jayawickreme, E., Kurzius, E., MacLeod, C., Miller, L. C., Read, S. J., Roberts, B. W., Robinson, M. D., Wood, D., & Wrzus, C. (2017). Integrating personality structure, personality process, and personality development. *European Journal of Personality*, *31*(5), 503–528. https://doi.org/10.1002/per.2115

Bleidorn, W., Kandler, C., Hülsheger, U. R., Riemann, R., Angleitner, A., & Spinath, F. M. (2010). Nature and nurture of the interplay between personality traits and major life goals. *Journal of Personality and Social Psychology*, *99*(2), 366–379. https://doi.org/10.1037/a0019982

Buchinger, L., Entringer, T. M., Richter, D., Wagner, G. G., Gerstorf, D., & Bleidorn, W. (2024). Codevelopment of life goals and the Big Five personality traits across adulthood and old age. *Journal of Personality and Social Psychology*, *126*(2), 346–368. https://doi.org/10.1037/pspp0000477

Bühler, J. L., Orth, U., Bleidorn, W., Weber, E., Kretzschmar, A., Scheling, L., & Hopwood, C. J. (2023). Life Events and Personality Change: A Systematic Review and Meta-Analysis. *European Journal of Personality*, 08902070231190219. https://doi.org/10.1177/08902070231190219

Carlson, R. W., Bigman, Y. E., Gray, K., Ferguson, M. J., & Crockett, M. J. (2022). How inferred motives shape moral judgements. *Nature Reviews Psychology*, *1*(8), 468–478. https://doi.org/10.1038/s44159-022-00071-x

Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Strope, B., & Kurzweil, R. (2018). Universal Sentence Encoder for English. In E. Blanco & W. Lu (Eds.), *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 169–174). Association for Computational Linguistics. https://doi.org/10.18653/v1/D18-2029

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Choshen-Hillel, S., Shaw, A., & Caruso, E. M. (2020). Lying to appear honest. *Journal of Experimental Psychology: General*, *149*(9), 1719–1735. https://doi.org/10.1037/xge0000737

Cohen, T. R., Panter, A. T., Turan, N., Morse, L., & Kim, Y. (2013). Agreement and similarity in self-other perceptions of moral character. *Journal of Research in Personality*, *47*(6), 816–830. https://doi.org/10.1016/j.jrp.2013.08.009

Condon, D. M., Wood, D., Mõttus, R., Booth, T., Costantini, G., Greiff, S., Johnson, W., Lukaszewski, A., Murray, A., Revelle, W., Wright, A. G. C., Ziegler, M., & Zimmermann, J. (2020). *Bottom Up Construction of a Personality Taxonomy*. *36*, 923–934.

Costantini, G., Gnisci, A., Sergi, I., Mottola, F., & Perugini, M. (2025). *The Personaltiy Goal Questionnaire (PGQ): Taxonomy and assessment of goals for the main personality dimensions. Manuscript in preparation.*

Costantini, G., & Perugini, M. (2016). The network of conscientiousness. *Journal of Research in Personality*, *65*, 68–88. https://doi.org/10.1016/j.jrp.2016.10.003

Costantini, G., & Perugini, M. (2018). A framework for testing causality in personality research. *European Journal of Personality*, *32*(3), 254–268. https://doi.org/10.1002/per.2150

Costantini, G., Saraulli, D., & Perugini, M. (2020). Uncovering the Motivational Core of Traits: The Case of Conscientiousness. *European Journal of Personality*, *34*(6), 1073–1094. https://doi.org/10.1002/per.2237

Cramer, A. O. J., Van Der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., Kendler, K. S., & Borsboom, D. (2012). Dimensions of Normal Personality as Networks in Search of Equilibrium: You Can't like Parties if you Don't like People. *European Journal of Personality*, *26*(4), 414–431. https://doi.org/10.1002/per.1866

de Vries, R. E., Zettler, I., & Hilbig, B. E. (2014). Rethinking Trait Conceptions of Social Desirability Scales: Impression Management as an Expression of Honesty-Humility. *Assessment*, *21*(3), 286–299. https://doi.org/10.1177/1073191113504619

DePaulo, B. M., Kirkendol, S. E., Kashy, D. A., Wyer, M. M., & Epstein, J. A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology*, *70*(5), 979–995. https://doi.org/10.1037/0022-3514.70.5.979

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)* (pp. 4171–4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423

DeVries, R. E., Hilbig, B. E., Zettler, I., Dunlop, P. D., Holtrop, D., Lee, K., & Ashton, M. C. (2018). *Still misinterpreting lie scales: Reply to Feldman's rejoinder*. https://research.vu.nl/en/publications/still-misinterpreting-lie-scales-reply-to-feldmans-rejoinder

DeYoung, C. G. (2015). Cybernetic Big Five Theory. *Journal of Research in Personality*, *56*, 33–58. https://doi.org/10.1016/j.jrp.2014.07.004

Di Pierro, R., Costantini, G., Fanti, E., Di Sarno, M., Preti, E., Madeddu, F., Clarkin, J. F., Caligor, E., & De Panfilis, C. (2022). Measurement Invariance of the Pathological Narcissism Inventory and Multimethod Examination of Narcissistic Presentations in Community and Clinical Samples. *Assessment*, 107319112211013. https://doi.org/10.1177/10731911221101367

Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, *80*(3), 501–519. https://doi.org/10.1037/0022-3514.80.3.501

Fleeson, W. (2020). Live simply so that others may simply live: Trying to get clarity on the meaning of Honesty/Humility. *European Journal of Personality*, *34*, 524–525. https://doi.org/10.1002/per.2284

Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology*, *97*(6), 1097–1114. https://doi.org/10.1037/a0016786

Fleeson, W., & Jayawickreme, E. (2015). Whole Trait Theory. *Journal of Research in Personality*, *56*, 82–92. https://doi.org/10.1016/j.jrp.2014.10.009

Fleeson, W., & Jayawickreme, E. (2021). Whole traits: Revealing the social-cognitive mechanisms constituting personality's central variable. In B. Gawronski (Ed.), *Advances in Experimental Social Psychology* (Vol. 63, pp. 69–128). Elsevier. https://doi.org/10.1016/bs.aesp.2020.11.002

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*(5), 378–382. https://doi.org/10.1037/h0031619

Funder, D. C. (2016). Taking situations seriously: The situation construal model and the Riverside Situational Q-Sort. *Current Directions in Psychological Science*, *25*(3), 203–208. https://doi.org/10.1177/0963721416635552

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, *103*(2), 265–275. https://doi.org/10.1037/0033-

2909.103.2.265

Günsoy, C., & Okten, I. O. (2022). Inferring Goals and Traits From Behaviors: The Role of Culture, Self-Construal, and Thinking Style. *Social Cognition*, *40*(2), 184–211. https://doi.org/10.1521/soco.2022.40.2.184

Halevy, R., Shalvi, S., & Verschuere, B. (2014). *Being honest about dishonesty: Correlating self-reports and actual lying*. *40*, 54–72. https://doi.org/10.1111/hcre.12019

Harris, Z. S. (1954). Distributional Structure. WORD, *10*(2–3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hart, C. L., Jones, J. M., Terrizzi, J. A., & Curtis, D. A. (2019). Development of the Lying in Everyday Situations Scale. *The American Journal of Psychology*, *132*(3), 343–352. https://doi.org/10.5406/amerjpsyc.132.3.0343

Hilbig, B. E., Thielmann, I., & Heck, D. W. (2024). Filling in the missing pieces: Personality traits (un)related to dishonest behavior. *European Journal of Personality*, 08902070241293621. https://doi.org/10.1177/08902070241293621

Hilbig, B. E., Zettler, I., Leist, F., & Heydasch, T. (2013). It takes two: Honesty–Humility and Agreeableness differentially predict active versus reactive cooperation. *Personality and Individual Differences*, *54*(5), 598–603. https://doi.org/10.1016/j.paid.2012.11.008

Hirschfeld, G., & von Brachel, R. (2014). Multiple-Group confirmatory factor analyses in R - A tutorial in measurement invariance with continuous and ordinal data. *Practical Assessment, Research & Evaluation*, *19*(7), 1–12.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185.

Hudson, N. W., Fraley, R. C., Chopik, W. J., & Briley, D. A. (2020). Change goals robustly predict trait growth: A mega-analysis of a dozen intensive longitudinal studies examining volitional change. *Social Psychological and Personality Science*. https://doi.org/10.1177/1948550619878423

Hussain, Z., Binz, M., Mata, R., & Wulff, D. U. (2024). A tutorial on open-source large language models for behavioral science. *Behavior Research Methods*, *56*(8), 8214–8237. https://doi.org/10.3758/s13428-024-02455-8

Jayawickreme, E., Zachry, C. E., & Fleeson, W. (2019). Whole Trait Theory: An integrative approach to examining personality structure and process. *Personality and Individual Differences*, *136*, 2–11. https://doi.org/10.1016/j.paid.2018.06.045

Kruglanski, A. W., Shah, J. Y., Fishbach, A., Friedman, R., Chun, W. Y., & Sleeth-Keppler, D. (2002). A theory of goal systems. *Advances in Experimental Social Psychology*, *34*,

331–378. https://doi.org/10.1016/S0065-2601(02)80008-9

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Langfelder, P., Zhang, B., & Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. *Bioinformatics*, *24*(5), 719–720. https://doi.org/10.1093/bioinformatics/btm563

Lee, K., & Ashton, M. C. (2004). Psychometric Properties of the HEXACO Personality Inventory. *Multivariate Behavioral Research*, *39*(2), 329–358. https://doi.org/10.1207/s15327906mbr3902_8

Lee, K., Ashton, M. C., Wiltshire, J., Bourdage, J. S., Visser, B. A., & Gallucci, A. (2013). Sex, power, and money: Prediction from the Dark Triad and Honesty-Humility. *European Journal of Personality*, *27*(2), 169–184. https://doi.org/10.1002/per.1860

Levine, E. E. (2022). Community standards of deception: Deception is perceived to be ethical when it prevents unnecessary harm. *Journal of Experimental Psychology: General*, *151*(2), 410–436. https://doi.org/10.1037/xge0001081

Levine, T. R., Ali, M. V., Dean, M., Abdulla, R. A., & Garcia-Ruano, K. (2016). Toward a pan-cultural typology of deception motives. *Journal of Intercultural Communication Research*, *45*(1), 1–12. https://doi.org/10.1080/17475759.2015.1137079

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*(2), 57–64. https://doi.org/10.1027/1614-2241.2.2.57

Lüdtke, O., Trautwein, U., & Husemann, N. (2009). Goal and personality trait development in a transitional period: Assessing change and stability in personality development. *Personality and Social Psychology Bulletin*, *35*(4), 428–441. https://doi.org/10.1177/0146167208329215

Lupoli, M. J., Levine, E. E., & Greenberg, A. E. (2018). Paternalistic lies. *Organizational Behavior and Human Decision Processes*, *146*, 31–50. https://doi.org/10.1016/j.obhdp.2018.01.001

Maltby, J., Wood, A. M., Day, L., & Pinto, D. (2012). The position of authenticity within extant models of personality. *Personality and Individual Differences*, *52*(3), 269–273. https://doi.org/10.1016/j.paid.2011.10.014

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*, 61–83.

https://doi.org/10.1016/j.jrp.2013.09.008

McArthur, J., McPhee, P., & Ternes, M. (2024). Examining the acceptability and likelihood of self-centered and other-centered lying using the HEXACO. *Personality and Individual Differences*, *220*, 112526. https://doi.org/10.1016/j.paid.2023.112526

McCabe, K. O., & Fleeson, W. (2012). What is extraversion for? Integrating trait and motivational perspectives and identifying the purpose of extraversion. *Psychological Science*, *23*(12), 1498–1505. https://doi.org/10.1177/0956797612444904

McCabe, K. O., & Fleeson, W. (2016). Are traits useful? Explaining trait manifestations as tools in the pursuit of goals. *Journal of Personality and Social Psychology*, *110*(2), 287–301. https://doi.org/10.1037/a0039490

McCrae, R. R., & Costa Jr., P. T. (1990). *Personality in Adulthood*. Guilford Press.

McCrae, R. R., & Sutin, A. R. (2018). A Five-Factor Theory perspective on causal analysis. *European Journal of Personality*, *32*, 151–166. https://doi.org/10.1002/per.2134

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, *17*(3), 437–455. https://doi.org/10.1037/a0028085

Mikolov, T., Le, Q. V., & Sutskever, I. (2013). *Exploiting Similarities among Languages for Machine Translation* (No. arXiv:1309.4168). arXiv. https://doi.org/10.48550/arXiv.1309.4168

Miller, C. B. (2020). Motivation and the virtue of honesty: Some conceptual requirements and empirical results. *Ethical Theory and Moral Practice*, *23*(2), 355–371. https://doi.org/10.1007/s10677-019-10055-1

Miller, C. B., Fleeson, W., Cohen, T. R., Furr, R. M., & Jayawickreme, E. (2021). *Honesty: Big questions, philosophical issues, and empirical research (white paper for the Honesty Project grant)*.

Milyavskaya, M., & Werner, K. M. (2018). Goal pursuit: Current state of affairs and directions for future research. *Canadian Psychology / Psychologie Canadienne*, *59*(2), 163–175. https://doi.org/10.1037/cap0000147

Mischel, W., & Shoda, Y. (1995). A cognitive-affective system theory of personality: Reconceptualizing situations, dispositions, dynamics, and invariance in personality structure. *Psychological Review*, *102*(2), 246–268. https://doi.org/10.1037/0033-295X.102.2.246

Moskowitz, G. B., & Olcaysoy Okten, I. (2016). Spontaneous Goal Inference (SGI). *Social and Personality Psychology Compass*, *10*(1), 64–80. https://doi.org/10.1111/spc3.12232

Mõttus, R., & Allerhand, M. (2018). Why do traits come together? The underlying trait and network approaches. In V. Zeigler-Hill & T. K. Shackelford (Eds.), *SAGE handbook of personality and individual differences: Volume 1. The science of personality and individual differences*. Sage.

Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (2017). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*, *112*(3), 474–490. https://doi.org/10.1037/pspp0000100

Mõttus, R., Wood, D., Condon, D. M., Back, M. D., Baumert, A., Costantini, G., Epskamp, S., Greiff, S., Johnson, W., Lukaszewski, A., Murray, A., Revelle, W., Wright, A. G. C., Yarkoni, T., Ziegler, M., & Zimmermann, J. (2020). Descriptive, Predictive and Explanatory Personality Research: Different Goals, Different Approaches, but a Shared Need to Move Beyond the Big Few Traits. *European Journal of Personality*, *34*(6), 1175–1201. https://doi.org/10.1002/per.2311

Müllner, D. (2013). fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for *R* and *Python*. *Journal of Statistical Software*, *53*(9). https://doi.org/10.18637/jss.v053.i09

Nagelkerke, N. J. D. (1991). A Note on a General Definition of the Coefficient of Determination. *Biometrika*, *78*(3), 691–692. https://doi.org/10.2307/2337038

Oliveira, C. M., & Levine, T. R. (2008). Lie Acceptability: A Construct and Measure. *Communication Research Reports*, *25*(4), 282–288. https://doi.org/10.1080/08824090802440170

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Quirin, M., Malekzad, F., Paudel, D., Knoll, A. C., & Mirolli, M. (2023). Dynamics of personality: The Zurich model of motivation revived, extended, and applied to personality. *Journal of Personality*, *91*(4), 928–946. https://doi.org/10.1111/jopy.12805

Quirin, M., Robinson, M. D., Rauthmann, J. F., Kuhl, J., Read, S. J., Tops, M., & DeYoung, C. G. (2020). The Dynamics of Personality Approach (DPA): 20 Tenets for Uncovering

the Causal Mechanisms of Personality. *European Journal of Personality*, *34*, 947–968. https://doi.org/10.1002/per.2295

Reisz, Z., Boudreaux, M. J., & Ozer, D. J. (2013). Personality traits and the prediction of personal goals. *Personality and Individual Differences*, *55*(6), 699–704. https://doi.org/10.1016/j.paid.2013.05.023

Revelle, W. (2016). *Psych: Procedures for personality and psychological research. R package version 1.6.6.* [Computer software].

Richetin, J., Richardson, D. S., & Boykin, D. M. (2011). Role of prevolitional processes in aggressive behavior: The indirect influence of goal. *Aggressive Behavior*, *37*(1), 36–47. https://doi.org/10.1002/ab.20364

Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin*, *143*(2), 117–141. https://doi.org/10.1037/bul0000088

Roberts, B. W., O'Donnell, M., & Robins, R. W. (2004). Goal and personality trait development in emerging adulthood. *Journal of Personality and Social Psychology*, *87*(4), 541–550. https://doi.org/10.1037/0022-3514.87.4.541

Roberts, B. W., & Robins, R. W. (2000). Broad dispositions, broad aspirations: The intersection of personality traits and major life goals. *Personality and Social Psychology Bulletin*, *26*(10), 1284–1296. https://doi.org/10.1177/0146167200262009

Romano, D., Costantini, G., Richetin, J., & Perugini, M. (2023). The HEXACO Adjective Scales and Its Psychometric Properties. *Assessment*, *30*(8), 2510–2532. https://doi.org/10.1177/10731911231153833

Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment*, *29*(4), 347–363. https://doi.org/10.1177/0734282911406661

Schild, C., Lilleholt, L., & Zettler, I. (2021). Behavior in cheating paradigms is linked to overall approval rates of crowdworkers. *Journal of Behavioral Decision Making*, *34*(2), 157–166. https://doi.org/10.1002/bdm.2195

Schreiber, J. B. (2021). Issues and recommendations for exploratory factor analysis and principal component analysis. *Research in Social and Administrative Pharmacy*, *17*(5), 1004–1011. https://doi.org/10.1016/j.sapharm.2020.07.027

Ścigała, K. A., Schild, C., & Zettler, I. (2020). Dishonesty as a signal of trustworthiness: Honesty-Humility and trustworthy dishonesty. *Royal Society Open Science*, *7*(10), 200685. https://doi.org/10.1098/rsos.200685

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.

Soto, C. J., Napolitano, C. M., & Roberts, B. W. (2021). Taking Skills Seriously: Toward an Integrative Model and Agenda for Social, Emotional, and Behavioral Skills. *Current Directions in Psychological Science*, *30*(1), 26–33. https://doi.org/10.1177/0963721420978613

Soto, C. J., Napolitano, C. M., Sewell, M. N., Yoon, H. J., & Roberts, B. W. (2022). An integrative framework for conceptualizing and assessing social, emotional, and behavioral skills: The BESSI. *Journal of Personality and Social Psychology*, *123*(1), 192–222. https://doi.org/10.1037/pspp0000401

Speer, S. P. H., Smidts, A., & Boksem, M. A. S. (2020). Cognitive control increases honesty in cheaters but cheating in those who are honest. *Proceedings of the National Academy of Sciences*, *117*(32), 19080–19091. https://doi.org/10.1073/pnas.2003480117

Stieger, M., Flückiger, C., & Allemand, M. (2024). One year later: Longer-term maintenance effects of a digital intervention to change personality traits. *Journal of Personality*, *92*(5), 1424–1437. https://doi.org/10.1111/jopy.12898

Stieger, M., Flückiger, C., Rüegger, D., Kowatsch, T., Roberts, B. W., & Allemand, M. (2021). Changing personality traits with the help of a digital personality change intervention. *Proceedings of the National Academy of Sciences*, *118*(8), e2017548118. https://doi.org/10.1073/pnas.2017548118

Thielmann, I., Hilbig, B. E., Klein, S. A., Seidl, A., & Heck, D. W. (2023). Cheating to benefit others? On the relation between Honesty-Humility and prosocial lies. *Journal of Personality*. https://doi.org/10.1111/jopy.12835

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological Bulletin*, *146*(1), 30–90. https://doi.org/10.1037/bul0000217

Thielmann, I., Zimmermann, J., Leising, D., Hilbig, B. E., & Back, M. (2017). Seeing is Knowing: On the Predictive Accuracy of Self– and Informant Reports for Prosocial and Moral Behaviours. *European Journal of Personality*, *31*(4), 404–418. https://doi.org/10.1002/per.2112

Turney, P. D., & Pantel, P. (2010). From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188. https://doi.org/10.1613/jair.2934

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., &

Polosukhin, I. (2017). Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 6000–6010.

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*(2), 281–300. https://doi.org/10.1037/a0017908

Wilson, A. T. (2018). Honesty as a Virtue. *Metaphilosophy*, *49*(3), 262–280. https://doi.org/10.1111/meta.12303

Wilt, J., & Revelle, W. (2015). Affect, Behaviour, Cognition and Desire in the Big Five: An Analysis of Item Content and Structure. *European Journal of Personality*, *29*(4), 478–497. https://doi.org/10.1002/per.2002

Wood, A. M., Linley, P. A., Maltby, J., Baliousis, M., & Joseph, S. (2008). The authentic personality: A theoretical and empirical conceptualization and the development of the Authenticity Scale. *Journal of Counseling Psychology*, *55*(3), 385–399. https://doi.org/10.1037/0022-0167.55.3.385

Wulff, D. U., & Mata, R. (2025). Semantic embeddings reveal and address taxonomic incommensurability in psychological measurement. *Nature Human Behaviour*, *9*(5), 944–954. https://doi.org/10.1038/s41562-024-02089-y

Yang, Y., Cer, D., Ahmad, A., Guo, M., Law, J., Constant, N., Abrego, G. H., Yuan, S., Tar, C., Sung, Y., Strope, B., & Kurzweil, R. (2019). Multilingual Universal Sentence Encoder for Semantic Retrieval. *arXiv Preprint arXiv*.

Zettler, I., Thielmann, I., Hilbig, B. E., & Moshagen, M. (2020). The nomological net of the HEXACO model of personality: A large-scale meta-analytic investigation. *Perspectives on Psychological Science*, *15*(3), 723–760. https://doi.org/10.1177/1745691619895036

Appendix. The GHGD questionnaire

Italian (original) version

Di seguito sono elencati una serie di obiettivi. Per ciascun obiettivo, la preghiamo di indicare quanto è importante per lei, usando la seguente scala di risposta: Indichi 1 se l'obiettivo non è per nulla importante per lei, 5 se l'obiettivo è molto importante per lei. Usi le risposte 2, 3 e 4 per indicare gradi intermedi di importanza.

		Per nulla importante per me				Molto importante per me
		1	2	3	4	5
1	Migliorare le mie condizioni di vita	□	□	□	□	□
2	Essere autentico	□	□	□	□	□
3	Non ingannare nessuno	□	□	□	□	□
4	Ottenere qualcosa illecitamente	□	□	□	□	□
5	Mostrare rispetto	□	□	□	□	□
6	Avere fiducia in me stesso/a	□	□	□	□	□
7	Comprendere le altre persone	□	□	□	□	□
8	Avere la coscienza pulita	□	□	□	□	□
9	Fare del bene	□	□	□	□	□
10	Dire la verità senza omissioni	□	□	□	□	□
11	Ottenere qualcosa anche se non la merito	□	□	□	□	□
12	Accontentarsi di quello che si ha	□	□	□	□	□
13	Non far sentire solo qualcuno	□	□	□	□	□
14	Stare bene con me stesso/a	□	□	□	□	□
15	Far capire agli altri ciò che penso	□	□	□	□	□
16	Non arrecare fastidi agli altri	□	□	□	□	□
17	Evitare di far soffrire una persona amata	□	□	□	□	□
18	Vivere una vita semplice	□	□	□	□	□
19	Non farmi prendere in giro	□	□	□	□	□
20	Accertarsi dell'onestà di qualcuno	□	□	□	□	□
21	Avere ciò che voglio a qualsiasi costo	□	□	□	□	□
22	Mantenere alta o aumentare la mia autostima	□	□	□	□	□
23	Far soffrire qualcuno	□	□	□	□	□
24	Smascherare le persone false	□	□	□	□	□
25	Provare la lealtà di qualcuno	□	□	□	□	□
26	Capire se mi posso fidare di qualcuno	□	□	□	□	□
27	Evitare di ferire qualcuno	□	□	□	□	□
28	Non volere troppo	□	□	□	□	□
29	Non mentire a me stesso	□	□	□	□	□
30	Non darmi delle arie con gli altri	□	□	□	□	□
31	Essere sempre corretto con gli altri	□	□	□	□	□
32	Vivere bene	□	□	□	□	□
33	Far sentire agli altri la mia vicinanza	□	□	□	□	□

		Per nulla importante per me 1	2	3	4	Molto importante per me 5
34	Sentirmi appagato/a	□	□	□	□	□
35	Ingraziarmi qualcuno	□	□	□	□	□
36	Raggirare qualcuno	□	□	□	□	□
37	Essere visto dagli altri esattamente per come io sono	□	□	□	□	□
38	Non essere troppo appariscente agli occhi degli altri	□	□	□	□	□
39	Avere stima di me stesso/a	□	□	□	□	□
40	Fare in modo che altri facciano ciò che conviene a me	□	□	□	□	□
41	Dimostrare che posso farcela da solo/a	□	□	□	□	□
42	Avere saldi principi	□	□	□	□	□
43	Essere felice	□	□	□	□	□
44	Avere relazioni solide con gli altri	□	□	□	□	□
45	Nascondere le mie vere intenzioni	□	□	□	□	□
46	Proteggere qualcuno	□	□	□	□	□
47	Ottenere giustizia	□	□	□	□	□
48	Far stare bene gli altri	□	□	□	□	□
49	Sentirmi soddisfatto/a di me stesso/a	□	□	□	□	□
50	Aiutare qualcuno	□	□	□	□	□
51	Manipolare il comportamento di altre persone	□	□	□	□	□
52	Far felice qualcuno	□	□	□	□	□
53	Evitare fraintendimenti	□	□	□	□	□
54	Chiarire i malintesi con gli altri	□	□	□	□	□
55	Essere trasparente con gli altri	□	□	□	□	□
56	Vivere sereno	□	□	□	□	□
57	Ferire qualcuno	□	□	□	□	□
58	Non dipendere dagli altri	□	□	□	□	□
59	Mantenere buoni rapporti con qualcuno	□	□	□	□	□
60	Esprimere il mio punto di vista	□	□	□	□	□
61	Ottenere un vantaggio a discapito di altri	□	□	□	□	□
62	Stare in pace con me stesso	□	□	□	□	□
63	Far sentire qualcuno amato	□	□	□	□	□
64	Avere saggezza, una comprensione matura della vita	□	□	□	□	□
65	Non nascondere niente di me agli altri	□	□	□	□	□
66	Scoprire la verità	□	□	□	□	□
67	Ingannare gli altri	□	□	□	□	□
68	Essere autonomo	□	□	□	□	□
69	Evitare di essere ingannato/a	□	□	□	□	□
70	Far sapere a qualcuno che può contare su di me	□	□	□	□	□
71	Fare in modo che gli altri facciano ciò che voglio io	□	□	□	□	□

		Per nulla importante per me 1	2	3	4	Molto importante per me 5
72	Offrire sostegno a qualcuno	□	□	□	□	□
73	Fare del male a qualcuno	□	□	□	□	□
74	Non creare disuguaglianze	□	□	□	□	□
75	Perseguire i miei ideali	□	□	□	□	□
76	Dire le cose come stanno	□	□	□	□	□
77	Non mettere in ombra altre persone	□	□	□	□	□
78	Rispettare gli altri	□	□	□	□	□

English (back-translated) version.

A number of goals are listed below. For each goal, please indicate how important it is to you, using the following response scale: Indicate 1 if the goal is not at all important to you, 5 if the goal is very important to you. Use responses 2, 3, and 4 to indicate intermediate degrees of importance.

		Per nulla importante per me 1	2	3	4	Molto importante per me 5
1	Improving my living conditions	□	□	□	□	□
2	Being authentic	□	□	□	□	□
3	Not to fool anyone	□	□	□	□	□
4	Obtaining something illicitly	□	□	□	□	□
5	Showing respect	□	□	□	□	□
6	Having confidence in myself	□	□	□	□	□
7	Understanding other people	□	□	□	□	□
8	Having a clear conscience	□	□	□	□	□
9	Doing good	□	□	□	□	□
10	Telling the truth without omissions	□	□	□	□	□
11	Getting something even though I don't deserve it	□	□	□	□	□
12	Making do with what I have	□	□	□	□	□
13	Not to make someone feel alone	□	□	□	□	□
14	Feeling good about myself	□	□	□	□	□
15	Making others understand what I think	□	□	□	□	□
16	Not to inconvenience others	□	□	□	□	□
17	Avoid making a loved one suffer	□	□	□	□	□
18	Living a simple life	□	□	□	□	□
19	Not to be fooled	□	□	□	□	□
20	Ascertaining someone's honesty	□	□	□	□	□
21	Having what I want at any cost	□	□	□	□	□
22	Maintaining high or increasing my self-esteem	□	□	□	□	□

		Per nulla importante per me 1	2	3	4	Molto importante per me 5
23	Making someone suffer	□	□	□	□	□
24	Unmasking false people	□	□	□	□	□
25	Proving someone's loyalty	□	□	□	□	□
26	Understanding whether I can trust someone	□	□	□	□	□
27	Avoid hurting someone	□	□	□	□	□
28	Not to want too much	□	□	□	□	□
29	Not to lie to myself	□	□	□	□	□
30	Not to put on airs with others	□	□	□	□	□
31	Always being fair to others	□	□	□	□	□
32	Living well	□	□	□	□	□
33	Making others feel my closeness	□	□	□	□	□
34	Feeling fulfilled	□	□	□	□	□
35	Currying favor with someone	□	□	□	□	□
36	Scamming someone	□	□	□	□	□
37	Being seen by others exactly as I am	□	□	□	□	□
38	Not to be too flashy in the eyes of others	□	□	□	□	□
39	Have self-esteem	□	□	□	□	□
40	Getting others to do what is convenient for me	□	□	□	□	□
41	Proving that I can do it on my own	□	□	□	□	□
42	Having firm principles	□	□	□	□	□
43	Being happy	□	□	□	□	□
44	Having strong relationships with others	□	□	□	□	□
45	Hiding my true intentions	□	□	□	□	□
46	Protecting someone	□	□	□	□	□
47	Getting justice	□	□	□	□	□
48	Making others feel good	□	□	□	□	□
49	Feeling satisfied with myself	□	□	□	□	□
50	Helping someone	□	□	□	□	□
51	Manipulating other people's behavior	□	□	□	□	□
52	Making someone happy	□	□	□	□	□
53	Avoiding misunderstandings	□	□	□	□	□
54	Clarifying misunderstandings with others	□	□	□	□	□
55	Being transparent with others	□	□	□	□	□
56	Living serenely	□	□	□	□	□
57	Hurting someone	□	□	□	□	□
58	Not to depend on others	□	□	□	□	□
59	Maintaining good relations with someone	□	□	□	□	□
60	Expressing my point of view	□	□	□	□	□
61	Gaining an advantage at the expense of others	□	□	□	□	□

		Per nulla importante per me 1	2	3	4	Molto importante per me 5
62	Being at peace with myself	□	□	□	□	□
63	Making someone feel loved	□	□	□	□	□
64	Having wisdom, a mature understanding of life	□	□	□	□	□
65	Not to hide anything about me from others	□	□	□	□	□
66	Discovering the truth	□	□	□	□	□
67	Deceiving others	□	□	□	□	□
68	Being autonomous	□	□	□	□	□
69	Avoid being deceived	□	□	□	□	□
70	Letting someone know they can count on me.	□	□	□	□	□
71	Making others do what I want them to do	□	□	□	□	□
72	Offering support to someone	□	□	□	□	□
73	Harming someone	□	□	□	□	□
74	Not to create inequality	□	□	□	□	□
75	Pursuing my ideals	□	□	□	□	□
76	Telling things like they are	□	□	□	□	□
77	Not to overshadow other people	□	□	□	□	□
78	Respecting others	□	□	□	□	□

Scoring instructions

The following table lists items for each goal class, distinguishing between honesty (GH) and dishonesty (GD). First, obtain the score of each goal class, by averaging all items within that class. Second, to obtain overall GH and GD scores, average the scores of all goal classes belonging to each type.

Goal class	Type	items
authenticity	GH	2, 29, 37, 55, 65
autonomy	GH	41, 58, 68
avoid being deceived	GH	19, 69
avoid hurting someone	GH	16, 17, 27
be satisfied with oneself	GH	6, 14, 22, 34, 39, 49, 62
be true to one's ideals and principles	GH	8, 42, 64, 75
express one's point of view	GH	15, 53, 60, 76
happiness	GH	1, 32, 43, 56
help others	GH	7, 13, 46, 50, 72
justice	GH	47, 74
make others feel loved	GH	33, 54, 63, 70
make others happy	GH	9, 48, 52
respect others	GH	5, 31, 78
sobriety	GH	12, 18, 28, 30, 38, 77
stability of relationships	GH	44, 59
tell the truth	GH	3, 10, 24, 66
test someone	GH	20, 25, 26
get what you want at any cost	GD	4, 11, 21, 61
lie	GD	36, 45, 67
hurt someone	GD	23, 57, 73
manipulate others	GD	35, 40, 51, 71

Supplementary Materials

Supplementary Material S1. Study 1. Instructions provided to participants

Original Italian version

Un **obiettivo/scopo** è qualcosa che si cerca di ottenere o raggiungere attraverso comportamenti che possono essere iniziati, intrapresi, gestiti, completati proprio al fine di raggiungerlo. Esempi di obiettivi sono laurearsi, fare un viaggio, fare una buona impressione. Un obiettivo può consistere anche nel resistere o nell'evitare di fare qualcosa, per esempio, smettere di fumare o, se si è già smesso, non ricominciare.

Ci sono risposte che **NON sono obiettivi/scopi**, ma che talvolta le persone indicano per errore come se fossero obiettivi. Questi NON-obiettivi includono ad esempio:

Cause e spiegazioni (es. se ti chiedono quale obiettivo vuoi ottenere facendo qualcosa, rispondere "Perché è bene", "Perché è bello", "Perché è importante").

Descrizione di situazioni (es. se ti chiedono a che scopo fai qualcosa, rispondere "Quando bisogna fare qualcosa", "Quando questo è importante").

Risposte circolari (es. se ti chiedono quale obiettivo vuoi ottenere facendo una cosa, rispondere "Per fare quella cosa" o "Perché voglio fare quella cosa")

Risposte vaghe o molto generali (es. rispondere che il tuo obiettivo è "Essere felice" o "Ottenere i miei obiettivi").

Dichiarare apertamente di non sapere cosa rispondere (es. rispondere "dipende", "non so").

Di seguito, le sarà chiesto di rispondere **a domande che riguardano gli obiettivi/scopi** che cerca o potrebbe cercare di raggiungere comportandosi in un certo modo.

Un esempio di domanda potrebbe essere: “A che scopo si comporta o si comporterebbe in modo coscienzioso?”. Esempi di risposte a questa domanda potrebbero essere: “per realizzarmi nel lavoro”, “per superare un esame”.

Un altro esempio di domanda potrebbe essere: “A che scopo si comporta o si comporterebbe in modo pigro?”. Esempi di risposte potrebbero essere: “per non fare qualcosa che non mi interessa fare”, “per distrarmi dagli impegni”.

Anche se si comporta in un certo modo raramente, o addirittura non lo fa mai, le chiediamo di immaginare comunque gli obiettivi che potrebbe cercare di ottenere comportandosi in quel modo.

Un modo semplice per essere sicuri che la risposta che ha dato sia un obiettivo/scopo è quello di **provare se ha senso quando preceduta da "allo scopo di"**.
- Esempio di obiettivo/scopo valido: "allo scopo di... superare l'esame".
- Esempio di obiettivo/scopo non valido: "allo scopo di... quando devo studiare".

Qualora lei non conoscesse il significato di qualcuno degli aggettivi, può cercarli sul dizionario se lo ritiene opportuno.

English translation

A **goal/objective** is something one seeks to obtain or achieve through behaviors that can be initiated, undertaken, managed, and completed for the very purpose of achieving it. Examples of goals are graduating from college, taking a trip, making a good impression. A goal can also consist of resisting or avoiding doing something, for example, quitting smoking or, if one has already quit, not starting again.

There are responses that **are NOT goals/goals**, but which people sometimes mistakenly point to as if they were goals. These NON-goals include, for example:

Causes and explanations (e.g., if asked what goal you want to achieve by doing something, answer "Because it is good," "Because it is beautiful," "Because it is important").

Description of situations (e.g., if you are asked what purpose you are doing something for, answer "When something needs to be done," "When this is important").

Circular answers (e.g., if asked what goal you want to achieve by doing something, answer "To do that thing" or "Because I want to do that thing")

Vague or very general answers (e.g., answer that your goal is "To be happy" or "To achieve my goals").

Openly state that you do not know what to answer (e.g., answer "It depends," "I don't know").

Next, you will be asked to answer **questions that relate to the goals/objectives** you seek or might seek to achieve by behaving in a certain way.

An example of a question might be, "For what purpose do you behave or would you behave conscientiously?" Examples of answers to this question might be, "to fulfill myself in work," "to pass an exam."

Another example of a question might be, "For what purpose do you behave or would you behave lazily?" Examples of answers might be, "to avoid doing something I'm not interested in doing," "to distract me from my commitments."

Even if you rarely or even never behave in a certain way, we still ask you to imagine the goals you might try to achieve by behaving that way.

An easy way to be sure that the answer you gave is a goal/purpose is to **test whether it makes sense when preceded by "for the purpose of."**

- Example of a valid goal/purpose: "for the purpose of... passing the exam."

- Example of invalid goal/purpose: "for the purpose of... when I have to study."

In case you do not know the meaning of any of the adjectives, you can look them up in the dictionary if you see fit.

Supplementary Material S2. CFA model of Honesty-Humility facet adjectives

Amenta and colleagues (2025) developed an assessment of Honesty-Humility facets through adjectives. The scale includes 22 adjectives: Three of them (*greedy*, *hypocritical*, and *sincere*) overlapped with those included for assessing Honesty-Humility in the HAS (Romano et al., 2023). Amenta and colleagues tested a confirmatory bifactor model including a general factor to control for response tendencies, as well as four Honesty-Humility facets: Sincerity, Fairness, Modesty, and Greed-Avoidance (see also Romano et al., 2023 for a similar approach). We examined whether that model would fit the adjective data in Studies 2 and 3.

In Study 2, the results confirmed the good fit of the model on self-report data, χ^2 (181) = 340.81, $p < .001$, CFI = .939, RMSEA = .054, SRMR = .044. Figure S1 reports model parameters in Study 2: loadings were all in the expected direction, and ranged between $\lambda = .42$ and $\lambda = .79$ in absolute value. Similar results were obtained for self-report data in Study 3, χ^2 (181) = 361.78, $p < .001$, CFI = .950, RMSEA = .050, SRMR = .047, with loadings ranging between $\lambda = .35$ and $\lambda = .83$.

In Study 3, we also tested the fit of the same model on data obtained from informants. The results confirmed the good fit of the model, χ^2 (181) = 361.78, $p < .001$, CFI = .950, RMSEA = .050, SRMR = .047, with loadings ranging between $\lambda = .35$ and $\lambda = .83$. We also examined measurement invariant between self-reports and informant-reports, by performing a series of model comparisons (Hirschfeld & von Brachel, 2014; Putnick & Bornstein, 2016). To evaluate measurement invariance, we considered ΔCFI < .01, ΔRMSEA < .015, and ΔSRMR < .03 as indicating sufficient similarity between models for testing weak invariance, and ΔCFI < .01, ΔRMSEA < .015, and ΔSRMR < .01 for supporting strong invariance (Chen, 2007; Di Pierro et al., 2022; Sass, 2011). The results are reported in Table S7 and support strong measurement invariance. Standardized parameter estimates are reported in Figures S2 (self-report) and S3 (informant-report). We used the regression method for estimating factor scores

for the four facets of Honesty-Humility, Sincerity (SI), Fairness (FA), Modesty (MO), and Greed-Avoidance (GR), to use as Honesty-Humility facet scores in Studies 2 and 3.

Figure S1. *Study 2. Standardized CFA parameters of Honesty-Humility facets, self-reported.*

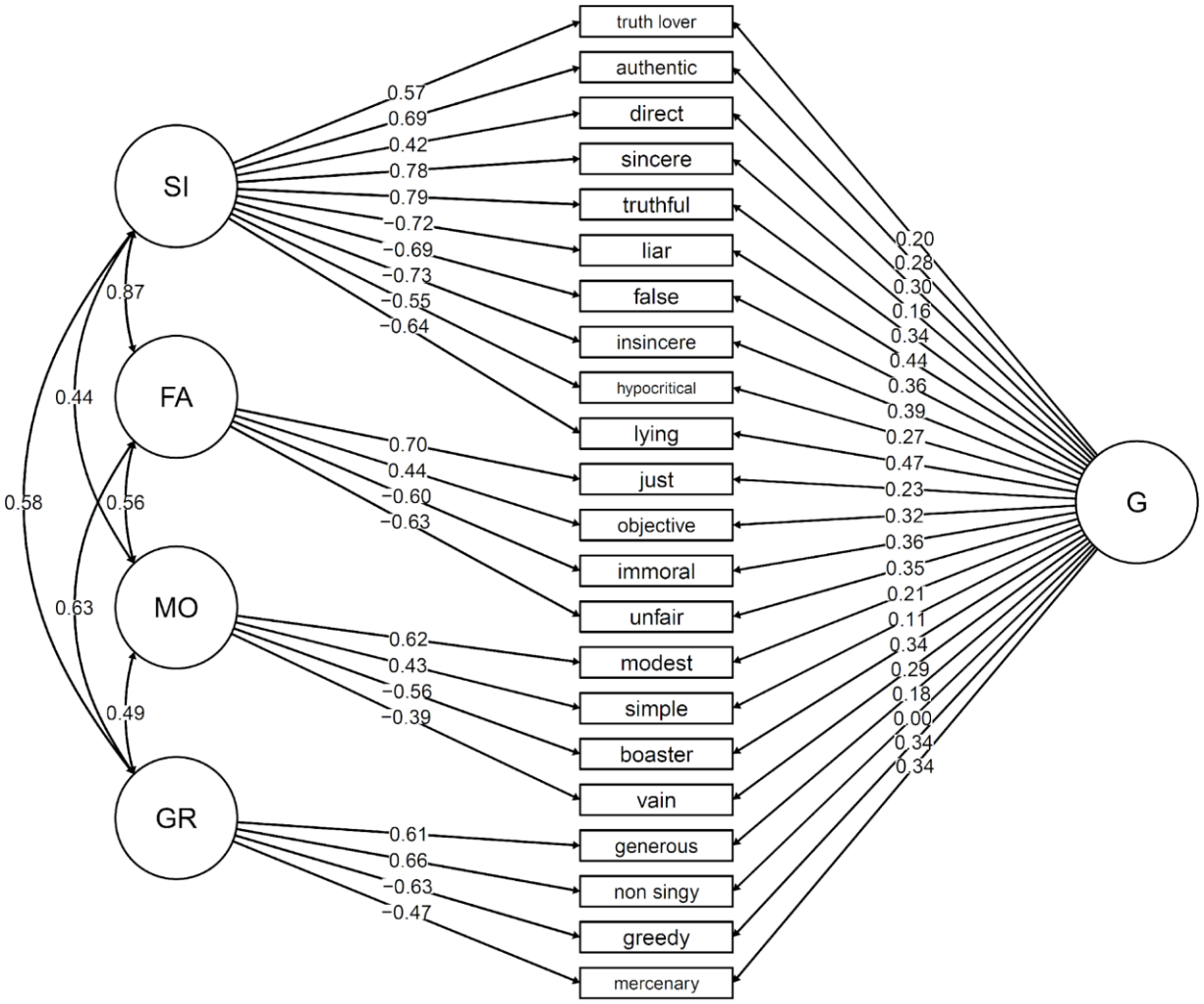
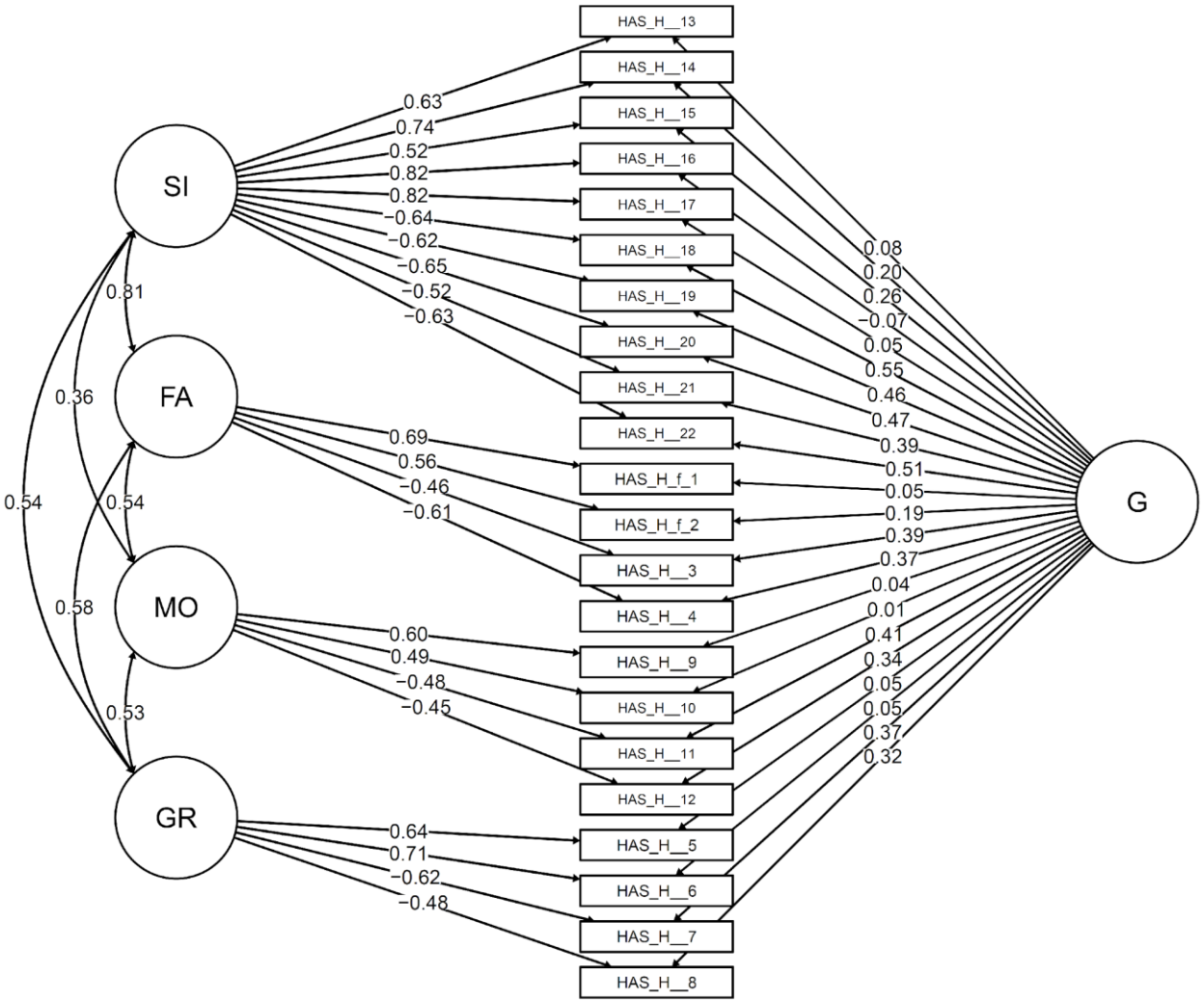


Figure S2. *Study 3. Standardized CFA parameters of Honesty-Humility facets for self-report data, the strong-invariance model (i.e., unstandardized loadings and intercepts constrained to be equal between self- and informant-report data)*.



Supplementary Material S3. Deviations from preregistration

Study 1

No deviations to report

Study 2

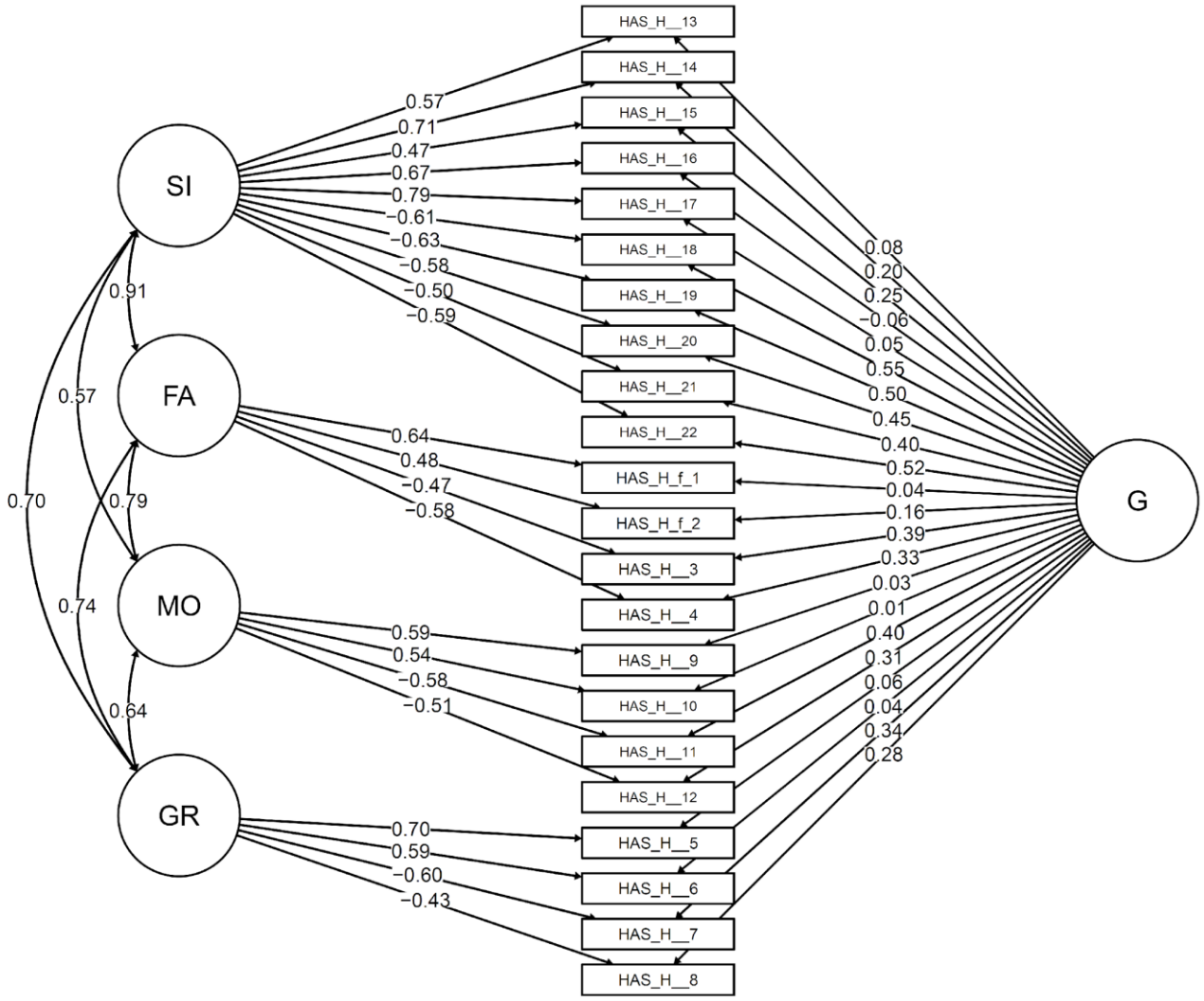
The preregistration indicated that 38 additional items would be added to the 10 items of the HAS Honesty-Humility scale to assess this trait more precisely. However, one item, "prepotente" (*bully*), was included in the HAS agreeableness scale and was not used for the assessment of Honesty-Humility. Because of an error in programming the study, instead of the item "integro" (*incorruptible*), we administered another unrelated item, "maleducato" (*impolite*). Hence, Honesty-Humility was eventually assessed by means of 46 items instead of 48 adjective items.

Study 3

The exclusion criteria for informants were not preregistered. We decided to implement those exclusion criteria to prevent including careless responses or responses from informants who were not actually acquainted with the targets.

The preregistration document mentioned 27 aspecific goals related to honesty, some of which also overlapped with goals of conscientiousness. These items were originally intended as the basis for developing a questionnaire on aspecific goals. However, we later abandoned this idea. Of the 27 items, 16 overlapped with goals of conscientiousness and were analyzed as such (Costantini et al., 2020). The remaining 11 items were included in the goal questionnaire but were not considered in the present manuscript.

Figure S3. *Study 3. Standardized CFA parameters of Honesty-Humility facets for informant-report data, the strong-invariance model (i.e., unstandardized loadings and intercepts constrained to be equal between self- and informant-report data).*



Supplementary Tables

Table S1. *Study 1 – standardized residuals of the chi-square test for 95 goal classes.*

	F	F-	M	M-	G	G-	S	S-	H	H-	E	E-	X	X-	A	A-	C	C-	O	O-
be rich*	-4.45	3.17	-3.80	-3.28	-3.23	28.59	-5.77	1.62	-5.62	8.71	-2.91	-3.03	-3.13	-2.71	-2.11	-2.10	-2.28	-2.38	-1.03	-2.70
justice and fairness*	24.86	-0.62	-2.54	-2.52	-1.36	-2.90	-2.81	-3.93	5.37	-2.89	-1.79	-1.86	-1.93	-1.04	-1.76	1.81	-1.84	-1.70	-1.78	-1.88
hide and lie*	-2.23	-1.26	-2.26	-2.24	-2.22	-2.24	-2.54	19.80	-3.25	-0.16	-1.38	-1.44	-1.49	-0.49	-1.36	-1.36	-0.70	1.03	-1.38	-1.45
sobriety*	-1.67	-2.16	18.89	-2.17	5.31	-1.68	-3.14	-3.16	1.54	-3.18	-1.34	-1.40	-1.45	1.22	-0.54	-1.32	-1.38	-1.28	-1.33	0.06
express one's point of view*	-0.25	-0.65	0.49	-2.68	-2.25	-2.69	18.19	-3.32	-3.01	-3.94	1.47	-1.73	-1.20	-1.55	-0.36	6.03	-1.71	-1.58	-1.65	-1.74
be admired*	-1.76	-2.45	-1.81	17.80	-0.35	-2.13	-2.05	-3.59	-2.32	0.35	-1.42	1.07	2.89	-0.10	-1.38	-1.91	-2.01	-1.86	7.77	-1.54
avoid false hope*	0.14	-1.39	0.86	-1.40	16.26	-1.40	-1.47	-2.03	-0.92	-2.05	0.32	-0.90	-0.93	0.47	-0.85	-0.85	-0.89	-0.82	-0.86	-0.91
think, reflect*	16.06	-2.35	-0.45	-2.76	-2.34	-2.38	-1.42	-2.88	-3.15	-3.77	-0.49	1.74	-1.27	-0.29	10.08	-1.05	2.38	1.57	1.37	1.12
loneliness, not being noticed by others*	-1.91	-1.91	15.75	-0.81	-1.90	-1.37	-2.37	-1.17	-1.57	-1.61	-1.19	-1.24	-1.28	10.98	-1.17	-1.17	-1.23	-0.22	-1.18	0.41
authenticity*	-1.04	-1.93	-0.19	-1.95	-2.37	-2.40	15.25	-3.49	1.44	-3.51	3.41	-1.54	-1.60	1.62	-1.46	-0.74	-0.85	-1.41	-0.77	-1.56
make others happy*	-0.79	-0.78	-1.30	-2.24	14.67	-1.29	-2.54	-1.16	3.05	-2.59	-1.38	0.00	1.99	-1.29	-0.60	-1.36	-1.43	-1.32	-1.38	0.69
help others*	-0.81	1.92	-3.30	-3.53	13.93	-2.74	-1.15	0.18	1.59	-3.04	-1.26	9.55	-1.93	-1.44	-2.48	1.82	-0.96	-2.40	-1.66	-2.65
gain personal advantage*	-2.87	3.76	-3.21	-1.52	-2.29	4.23	-5.58	6.63	-3.79	13.87	-2.47	-2.57	-2.66	-1.84	-2.00	-1.99	-2.55	-1.45	-1.16	-2.18
tell the truth*	-1.57	-1.57	-1.59	-1.58	-0.88	-1.58	12.06	-0.82	-0.32	-1.34	0.08	-1.02	-1.05	-0.91	-0.96	-0.96	-1.01	-0.93	-0.97	0.99
aggress*	-2.95	11.93	-3.32	3.29	-3.27	-2.32	-3.81	2.17	-4.55	9.96	-2.04	-2.12	-2.20	-1.35	-2.01	4.28	-2.10	-1.94	-2.03	-1.65
avoid hurting someone*	0.10	-2.69	0.04	-2.70	-1.47	-2.71	1.34	11.18	2.48	-2.53	-1.05	-1.74	-1.80	-0.23	0.25	-1.01	-1.12	-1.59	-1.66	-1.76
feeling superior to others*	-2.11	0.44	-2.13	11.05	-2.09	-0.10	-3.06	-2.34	-1.96	8.64	-1.31	-1.36	-1.41	-1.22	-0.48	0.33	-1.35	-1.24	-1.30	-1.37
be better than others*	-1.96	0.77	-1.44	0.75	-1.95	-0.34	-2.05	1.10	-2.46	10.90	-1.22	-1.27	-1.31	-1.13	-1.20	-1.19	-0.43	-1.16	0.50	-1.28
physical satisfaction*	-1.24	-0.37	-0.40	0.47	-0.36	10.73	-1.80	-1.81	-1.80	0.67	-0.77	-0.80	0.42	-0.71	-0.76	-0.75	-0.79	2.08	-0.76	-0.81
avoid judgment of others*	-1.93	-1.93	10.66	-1.38	0.86	-1.39	-0.37	1.61	-0.79	-2.04	-0.34	-1.25	-0.49	0.73	-1.18	-1.17	-1.24	0.67	-1.19	1.21
hurt someone*	-2.53	6.39	-2.56	2.53	-2.09	-1.70	-3.37	1.55	-3.38	10.37	-1.57	-1.64	-1.69	-1.46	-1.55	1.15	-1.62	0.58	-1.56	-1.65

	F	F-	M	M-	G	G-	S	S-	H	H-	E	E-	X	X-	A	A-	C	C-	O	O-
prove one's worth*	-1.41	0.63	1.37	10.36	-1.80	-2.25	1.23	-3.85	-1.46	-0.94	-1.00	4.40	-0.58	-0.85	-1.61	0.98	-1.69	-1.56	2.20	-0.51
be appreciated by others*	-1.09	-3.05	10.25	5.61	3.18	-3.92	-0.12	-1.61	1.08	-3.10	-1.11	0.01	0.24	-0.38	-1.50	-2.38	-2.07	-1.85	0.23	-0.46
put oneself first*	-2.00	-0.93	-0.44	-0.95	-1.99	10.18	-2.52	-0.60	-2.92	4.75	-1.24	0.30	-1.34	-0.27	-0.38	0.47	-1.28	1.43	-1.24	0.28
be true to one's ideals and principles*	4.50	-2.03	-2.08	-2.43	-2.02	-2.82	8.00	-3.82	9.92	-3.30	-1.14	-1.24	-1.32	-1.62	-1.71	-0.49	-1.80	-1.66	-1.73	0.46
be satisfied with oneself*	4.05	-0.01	-3.14	5.78	4.37	-3.11	-0.92	-5.24	9.80	-1.62	-1.91	0.24	-1.80	-1.24	-1.85	-2.24	-1.25	-1.74	-2.28	-2.45
stability of relationships*	-2.28	-2.61	-0.36	-3.29	0.40	-3.30	9.03	-2.14	9.47	-4.35	-0.49	2.84	1.65	-0.24	-2.00	-2.00	-2.10	-1.94	-1.50	-2.14
be trusted*	-0.08	-1.74	-0.96	-1.34	-2.15	-2.18	9.34	-2.86	6.27	-3.80	3.58	-0.42	-1.73	-0.80	-0.26	-0.91	-0.39	-1.52	-1.59	-1.68
win*	-1.14	0.69	-1.77	-1.76	-1.74	-1.76	-1.65	2.32	-0.77	9.30	-1.09	-0.22	-1.17	-1.01	-1.07	-0.10	-1.12	-1.03	-0.13	-1.14
avoid pain*	-1.33	-0.49	-0.96	-2.17	-0.05	-0.52	-1.31	8.35	-0.42	-2.29	0.99	0.21	-0.52	1.28	-0.26	-0.91	0.87	-0.84	-0.94	-0.44
manipulate others*	-2.09	0.47	-2.12	0.45	-1.05	-1.60	-1.92	8.10	-2.67	3.20	2.68	-0.59	-1.40	-0.36	-1.28	-0.46	-1.34	0.43	-1.29	-1.36
respect others*	3.10	-1.41	-0.69	-1.42	-0.65	-1.42	-0.41	-1.52	7.81	-2.08	-0.88	-0.91	0.15	1.69	-0.86	-0.86	-0.91	-0.84	-0.87	0.20
rebellion*	-1.08	7.75	-1.10	2.82	-1.08	-1.09	-1.57	-0.16	-0.86	0.53	-0.67	-0.70	-0.72	-0.63	-0.66	-0.66	-0.69	0.96	-0.67	-0.71
be a good example*	0.79	-0.69	-0.72	-0.70	1.55	-1.44	-1.00	-2.10	7.64	-1.58	-0.89	0.18	-0.96	-0.83	0.29	-0.87	0.20	-0.85	0.28	-0.94
happiness*	-1.09	1.69	1.88	-2.34	2.03	6.20	-2.01	-1.83	-2.04	-2.34	0.66	0.93	2.08	-0.52	6.08	-2.16	-0.42	-1.09	-1.23	-0.03
romantic love*	-1.63	-1.63	-1.65	6.19	-0.30	-0.99	0.50	-1.90	1.44	-1.45	4.08	0.91	-0.14	1.24	-0.99	-0.99	-1.04	-0.96	2.07	-1.06
get what you want at any cost*	-1.05	5.66	-2.10	-2.08	-0.51	-1.07	-3.02	5.95	-2.65	5.87	-1.29	-1.34	-1.39	-1.20	-1.27	-0.45	-1.33	-1.23	-0.48	-1.35
autonomy*	-1.11	1.76	-0.18	-0.16	-0.14	5.54	-0.22	0.46	-0.92	-1.64	-0.69	-0.72	-0.74	-0.64	-0.68	-0.68	0.73	-0.66	-0.69	0.70
make others feel well*	-0.84	-1.53	5.32	-1.54	4.75	-1.54	-2.23	-2.24	2.82	-1.76	0.13	3.16	-1.03	-0.89	3.45	-0.93	-0.98	-0.91	-0.95	-1.00
make others feel loved*	-1.16	-1.16	-0.27	-1.17	5.27	-1.17	-0.35	-1.70	1.64	-1.71	3.54	-0.75	1.87	-0.67	-0.71	-0.71	2.02	-0.69	-0.72	0.60
be safe*	-2.26	1.01	-0.15	-0.47	-2.60	4.56	-2.48	4.07	-2.50	0.58	1.53	0.24	-1.47	2.50	-0.68	-0.10	1.92	-1.17	-1.27	-0.87
be accepted*	-1.41	-1.41	4.53	0.83	-1.41	-1.42	0.69	2.31	-0.96	-2.08	0.29	-0.91	1.24	-0.82	-0.86	-0.86	1.37	-0.84	-0.87	2.43
test someone*	3.69	-1.28	-0.48	-1.29	-1.28	-0.47	-0.05	2.95	-0.66	-1.29	1.78	-0.83	-0.86	2.02	-0.78	-0.78	-0.82	0.59	0.51	-0.84
religious faith*	-0.73	-0.72	-0.73	-0.73	3.68	-0.73	1.08	-1.06	2.13	-1.07	-0.45	-0.47	-0.49	-0.42	-0.44	-0.44	-0.47	-0.43	-0.45	1.69
avoid being deceived*	-0.94	-0.94	1.29	-0.94	-0.93	-0.94	1.12	3.57	-1.37	0.25	-0.58	-0.61	-0.63	3.23	-0.57	-0.57	-0.60	-0.55	-0.58	-0.61
take the easy way*	-0.53	1.88	1.03	0.26	-1.32	1.05	-1.93	0.97	-1.93	3.25	-0.82	-0.86	-0.89	-0.77	-0.81	-0.81	0.36	0.52	0.43	-0.87
persuasion**	-1.08	2.84	-1.10	-0.11	-1.08	0.86	-0.14	1.27	-0.15	-0.18	-0.67	0.77	-0.72	-0.63	-0.66	0.89	-0.69	-0.64	-0.67	0.75
be respected	1.94	1.25	-0.18	2.61	-0.13	-1.54	-1.22	-1.74	2.32	0.75	-0.95	0.05	-1.03	-0.89	-0.94	1.26	-0.98	-0.91	-0.95	-1.00
improve oneself	1.39	-0.47	-1.12	-0.48	1.41	-1.11	-0.68	-1.60	2.03	0.16	-1.07	0.75	1.55	0.04	-0.07	-1.05	1.71	-1.01	-0.09	-0.20

	F	F-	M	M-	G	G-	S	S-	H	H-	E	E-	X	X-	A	A-	C	C-	O	O-
showing yourself as you really are	-2.06	-1.02	0.48	-1.55	-2.05	-1.56	9.50	-3.01	0.77	-2.29	10.81	-1.33	-0.63	-0.33	-1.26	-1.26	-1.32	-1.22	-1.27	-0.57
entertain others	-1.33	-1.33	-1.34	6.65	0.29	-1.34	-1.93	-1.94	-1.35	-0.22	-0.82	-0.86	15.36	-0.77	-0.81	-0.81	-0.85	-0.78	-0.82	-0.87
assert one's reasons	-0.74	6.07	-2.23	0.70	-2.19	-2.22	3.19	0.32	-1.79	-2.54	-0.61	-1.42	-1.47	-1.27	-0.58	12.52	-1.41	-0.51	-1.36	-0.71
hide my true self	-1.96	-1.96	0.72	3.47	-1.95	-1.97	-2.45	5.47	-2.46	-0.52	0.48	0.37	-1.31	11.58	-0.34	-1.19	-1.25	-1.16	-0.36	-0.47
comply with rules	2.68	-0.52	-0.55	-1.33	-1.32	-0.54	-0.17	-1.36	5.07	-1.38	-0.82	-0.86	-0.89	-0.77	-0.81	-0.81	-0.85	-0.78	-0.82	5.08
have new experiences	-1.35	-0.56	-1.37	-1.35	0.24	4.91	-1.39	-1.97	-1.97	-1.42	-0.84	3.85	1.38	1.84	-0.82	-0.82	-0.86	1.77	7.80	0.29
avoid people and situations	-1.18	-0.78	-0.07	0.75	-2.34	-1.60	-2.28	4.81	-1.44	1.89	-1.71	-1.78	-1.27	5.59	0.18	0.81	-1.17	3.49	-1.70	-0.05
devote oneself to a cause	1.82	0.74	-0.93	-1.98	2.93	-1.45	0.69	-1.70	4.22	-2.52	-1.23	-0.47	-1.32	-1.14	-1.21	-1.20	-1.26	-1.17	7.25	-1.29
personal realization	1.86	0.85	-1.58	-0.87	-2.85	2.81	-2.44	2.43	-2.22	3.81	-1.99	0.45	-1.66	-1.86	0.17	-1.96	4.57	-0.79	0.66	-1.59
prevent conflict	3.19	-1.70	-1.25	-2.20	-1.69	-2.20	3.61	3.20	0.02	-3.23	-1.36	-1.41	-1.46	-0.45	7.93	0.22	-1.40	-1.29	-1.35	2.20
avoid feeling uncomfortable	-1.78	-1.78	3.52	0.00	-1.77	-1.20	-0.85	1.31	0.01	-2.20	-1.11	-1.15	-0.33	8.92	1.74	-0.14	-1.14	1.87	-1.10	1.50
reassure someone	1.27	-1.26	-0.44	-0.42	0.44	-1.27	1.87	-1.84	3.09	-1.25	-0.78	1.71	-0.84	-0.73	3.23	-0.77	-0.81	-0.74	-0.78	-0.82
do not think	1.28	-1.11	1.99	-1.92	0.51	0.44	-1.31	-1.93	3.02	-3.41	-1.06	-1.15	-0.66	-0.90	0.23	-1.65	-0.53	11.39	-1.05	1.78
be loved	-1.21	-1.80	1.11	-0.04	1.78	-0.63	3.01	-1.34	1.69	-2.65	4.41	0.61	0.51	-1.04	-0.16	-1.10	-1.15	-1.06	-0.18	-1.17
show others that you care about them	2.20	-0.72	-1.47	-1.46	2.96	-1.46	0.02	-1.06	0.54	-1.62	7.06	1.25	1.15	-0.84	-0.89	0.27	-0.93	-0.86	-0.90	-0.95
show one's disappointment	-1.39	-0.63	-1.41	-0.64	-0.62	-1.40	2.44	-2.03	-1.47	1.26	-0.86	-0.90	-0.93	0.47	-0.85	13.67	-0.89	0.42	-0.86	0.23
not being submissive to others	-0.41	0.43	-1.27	2.10	-1.25	-0.43	-0.60	1.23	-1.83	2.41	-0.78	-0.81	-0.84	-0.73	-0.77	5.92	-0.81	-0.74	-0.78	-0.82
accomplish something, observe a commitment	2.36	0.74	-0.39	-1.44	-1.42	0.17	-2.08	-2.10	-1.30	0.61	-0.38	0.35	-1.32	-1.14	0.50	-1.20	6.09	1.48	1.32	4.34
overcome difficulties	-1.15	2.35	-1.19	-1.66	-2.14	1.31	-2.76	1.57	-2.77	-0.64	-1.33	7.56	-0.71	-0.41	6.56	1.06	0.88	0.36	0.23	-0.66
friendship	-1.14	-3.18	0.15	-1.50	1.29	-3.20	2.34	-2.91	-1.16	-3.46	1.20	0.49	26.51	-0.71	-1.41	-1.40	-1.52	-1.88	0.17	-0.56
make others proud	0.73	0.74	-0.73	-0.73	2.21	-0.73	-1.06	-1.06	0.01	-1.07	-0.45	1.72	-0.49	-0.42	-0.44	-0.44	-0.47	-0.43	4.12	1.69
loialty to family values	-0.53	-1.33	0.24	-1.33	-0.52	-1.34	-0.76	-1.94	2.15	-1.38	-0.82	1.54	-0.89	-0.77	0.45	0.46	-0.85	-0.78	-0.82	13.40
show one's feelings	-1.21	-0.33	-1.23	-1.22	-0.32	-1.22	2.08	-1.77	-1.77	-1.78	12.88	-0.78	3.00	-0.70	-0.74	3.43	-0.78	0.71	-0.75	-0.79
dealing with problems	2.07	-0.07	-0.81	-0.79	-0.77	-0.09	1.47	-1.67	-2.18	-1.17	2.39	1.17	-1.00	-0.86	6.96	0.22	0.12	1.44	-0.92	-0.97
avoid being excluded or ignored**	-0.94	-0.94	0.17	0.19	0.20	-0.94	0.29	-0.55	-0.54	1.89	-0.58	1.09	2.65	-0.54	-0.57	-0.57	-0.60	-0.55	-0.58	1.07
having fun	-1.96	0.23	1.80	1.84	0.24	1.82	-2.85	-1.28	-2.85	-1.70	-0.37	-1.27	11.34	0.68	-1.20	-1.19	-1.25	5.07	-0.36	0.34
personal growth	-0.59	-0.59	-0.60	-0.59	-0.59	1.18	1.75	0.43	0.44	-0.87	-0.37	2.29	-0.40	-0.34	-0.36	-0.36	-0.38	-0.35	-0.37	-0.39

	F	F-	M	M-	G	G-	S	S-	H	H-	E	E-	X	X-	A	A-	C	C-	O	O-
avoid or manage things you do not care about	-1.91	-0.79	-0.83	-1.36	-1.34	-0.81	-1.56	1.67	-2.38	0.01	-1.19	-1.24	-1.28	1.68	-0.29	-1.17	-1.23	17.08	-0.31	4.56
job	0.99	-0.59	1.53	1.27	-1.52	-0.63	-2.00	0.03	0.28	-1.19	-1.65	-0.80	-0.93	-1.99	-0.61	-1.09	4.48	-1.52	7.27	1.04
overcome one's limitations	1.50	-1.19	-1.20	0.59	-1.18	0.59	-1.72	-1.08	-1.73	0.19	-0.74	3.25	1.80	-0.69	-0.73	-0.72	1.94	0.76	4.87	0.56
show affection	-0.33	-0.33	-1.23	-1.22	0.56	-0.35	0.16	-1.77	1.43	-1.15	11.52	0.53	-0.81	-0.70	-0.74	0.65	-0.78	-0.72	-0.75	-0.79
starting a family	-0.54	1.42	-0.55	1.41	-0.54	-0.54	-0.79	-0.79	0.64	-0.80	-0.34	2.58	-0.36	-0.31	-0.33	-0.33	-0.35	-0.32	-0.33	2.55
have control	-0.07	-0.07	1.31	-0.08	-1.49	-1.50	-1.65	-1.15	-2.18	-1.69	0.18	1.17	-1.00	-0.86	5.83	2.47	6.57	2.60	-0.92	1.14
cultural growth	-0.87	-1.41	-1.44	-1.97	-1.40	-0.34	1.14	-2.07	-1.26	0.66	-1.22	0.37	-0.52	-1.13	1.39	-1.19	10.27	2.40	2.20	1.96
do something well, avoid mistakes	1.13	-1.75	-0.35	-2.24	-1.25	-1.77	-0.08	-2.21	0.95	-3.29	-0.64	-0.72	-1.49	-0.49	0.91	-1.36	24.68	-0.53	-1.38	-0.03
progress	1.02	-0.64	-0.65	-0.64	-0.64	-0.64	-0.93	-0.94	-0.93	-0.94	-0.40	-0.41	-0.43	-0.37	-0.39	-0.39	-0.41	2.32	12.55	-0.42
avoid being hurt	-1.37	0.96	-1.39	-1.38	0.19	-1.38	0.27	-0.31	-0.87	0.22	1.56	-0.89	-0.92	10.83	-0.84	0.39	0.29	-0.81	-0.85	0.26
managing time well	-2.03	-1.50	-0.50	-2.04	-0.97	-1.00	0.88	-2.59	-2.96	-1.86	-1.26	-0.53	-1.36	-0.30	1.25	-0.41	10.60	14.23	-1.25	5.70
getting along	-0.68	-0.68	-0.69	0.86	-0.68	-0.69	-0.99	0.13	0.13	0.11	6.80	-0.44	-0.46	-0.40	-0.42	2.04	-0.44	-0.40	-0.42	-0.45
tranquillità	-1.15	-1.15	0.79	-2.16	-0.13	-0.67	-0.94	-2.78	-1.31	-3.16	-1.33	0.85	1.45	4.57	13.64	1.06	0.13	6.05	-1.33	0.08
show you are wounded	-0.54	-0.54	-0.55	-0.54	-0.54	-0.54	0.64	-0.79	-0.79	-0.80	-0.34	-0.35	-0.36	-0.31	-0.33	12.08	-0.35	-0.32	-0.33	-0.35
vent, express anger	-1.47	-0.03	-0.78	-0.76	-1.47	-0.05	-2.14	-1.63	-2.15	0.43	6.94	0.13	0.06	-0.85	-0.90	16.25	-0.94	0.31	-0.91	-0.96
get help	-1.21	-0.33	-0.36	-1.22	-0.32	-1.22	0.16	-1.13	-0.49	0.12	14.25	-0.78	1.73	-0.70	-0.74	-0.74	-0.78	-0.72	-0.75	-0.79
be unique	-1.41	-0.66	0.06	-0.67	-1.41	-0.67	0.14	-2.07	-2.06	-1.54	-0.88	0.21	2.33	-0.82	-0.86	-0.86	-0.91	-0.84	21.50	0.20
create art	-0.77	-0.76	-0.77	-0.77	-0.76	-0.77	-1.11	-1.12	-1.12	-0.13	-0.48	-0.49	-0.51	-0.44	1.72	-0.47	-0.49	1.81	14.70	-0.50
tradition and stability	-1.47	-1.47	-0.78	-1.48	-1.47	-1.48	-2.14	-2.16	-0.57	-2.17	-0.92	-0.95	-0.99	-0.85	-0.90	-0.90	0.14	-0.87	-0.91	33.32

Note. F = fairness, *M* = modesty, G = greed avoidance, S = sincerity, H = general honesty, E = emotionality, X = extraversion, A = agreeableness vs. anger, C = conscientiousness, O = openness to experience. "" = positive pole, "-" = negative pole.

* The goal class was among one of the 46 classes identified as pertinent to Honesty-Humility or one of its facets in Study 1.

** The goal class was one of the two identified as potentially pertinent to Honesty-Humility in a previous study (Costantini et al., 2025) and thus further retained for consideration.

Table S2. *Study 2 – list of 191 preliminary goal items.*

goal	item
aggress	Essere temuto dagli altri
aggress	Fare un dispetto a qualcuno
aggress	Pareggiare i conti con qualcuno che mi ha fatto un torto
aggress	Vendicarmi
authenticity	Essere autentico
authenticity	Essere trasparente con gli altri
authenticity	Essere visto dagli altri esattamente per come sono
authenticity	Non mentire a me stesso[(a)]
authenticity	Non nascondere niente di me agli altri
autonomy	Dimostrare che posso farcela da solo
autonomy	Essere autonomo
autonomy	Non dipendere dagli altri
avoid being deceived	Evitare di essere ingannato
avoid being deceived	Non farmi prendere in giro
avoid being excluded or ignored	Evitare di essere ignorato dagli altri
avoid being excluded or ignored	Non essere escluso dagli altri
avoid false hope	Evitare di avere aspettative troppo elevate, che possono essere in seguito disilluse
avoid false hope	Evitare di crearmi false speranze
avoid hurting someone	Evitare di far soffrire una persona amata
avoid hurting someone	Evitare di ferire qualcuno
avoid hurting someone	Non arrecare fastidi agli altri[(a)]
avoid judgment of others	Evitare il giudizio degli altri
avoid judgment of others	Non avere paura di essere giudicato
avoid judgment of others	Non sentirmi giudicato
avoid pain	Evitare di subire le conseguenze spiacevoli delle mie azioni[(a)]
avoid pain	Evitare le sofferenze
avoid pain	Evitare problemi
avoid pain	Non provare dolore
be a good example	Essere un buon esempio per gli altri
be a good example	Ispirare gli altri
be accepted	Essere accettato dagli altri
be accepted	Ottenere il consenso degli altri
be admired	Attirare l'attenzione di qualcuno
be admired	Essere al centro dell'attenzione
be admired	Essere bello e attraente
be admired	Far colpo sugli altri
be admired	Farmi notare
be admired	Ricevere ammirazione
be admired	Ricevere attenzioni
be appreciated by others	Compiacere gli altri
be appreciated by others	Essere apprezzato

be appreciated by others	Essere ben visto dagli altri
be appreciated by others	Fare una bella impressione
be appreciated by others	Non essere banale agli occhi degli altri
be appreciated by others	Piacere agli altri
be better than others	Avere autorità sugli altri[(a)]
be better than others	Emergere rispetto agli altri
be better than others	Essere il/la migliore
be better than others	Essere un leader
be rich	Aumentare il mio reddito
be rich	Ottenere tutto ciò che voglio
be rich	Poter comprare ciò che desidero
be rich	Risparmiare i soldi[(a)]
be rich	Soddisfare i miei capricci
be safe	Non correre rischi
be safe	Non farmi del male
be safe	Non mettermi in pericolo
be safe	Non mettermi nei guai
be safe	Proteggermi
be safe	Salvaguardarmi
be safe	Stare al sicuro
be safe	Stare in salute
be satisfied with oneself	Avere fiducia in me stesso
be satisfied with oneself	Avere stima di me stesso
be satisfied with oneself	Mantenere alta o aumentare la mia autostima
be satisfied with oneself	Sentirmi appagato
be satisfied with oneself	Sentirmi soddisfatto di me stesso
be satisfied with oneself	Stare bene con me stesso
be satisfied with oneself	Stare in pace con me stesso[(a)]
be true to one's ideals and principles	Avere la coscienza pulita[(a)]
be true to one's ideals and principles	Avere saggezza, una comprensione matura della vita
be true to one's ideals and principles	Avere saldi principi
be true to one's ideals and principles	Perseguire i miei ideali
be trusted	Fare in modo che gli altri si fidino di me
be trusted	Guadagnarmi la fiducia degli altri
be trusted	Ispirare fiducia negli altri
express one's point of view	Dire le cose come stanno[(a)]
express one's point of view	Esprimere il mio punto di vista
express one's point of view	Evitare fraintendimenti[(a)]
express one's point of view	Far capire agli altri ciò che penso
feeling superior to others	Essere invidiato dagli altri[(a)]
feeling superior to others	Mostrare di essere colto
feeling superior to others	Mostrare la mia superiorità
feeling superior to others	Sentirmi superiore agli altri

gain personal advantage	Ottenere un vantaggio personale
gain personal advantage	Trarre un guadagno
get what you want at any cost	Avere ciò che voglio a qualsiasi costo[(a)]
get what you want at any cost	Ottenere qualcosa anche se non la merito[(a)]
get what you want at any cost	Ottenere qualcosa illecitamente[(a)]
get what you want at any cost	Ottenere un vantaggio a discapito di altri[(a)]
happiness	Essere felice
happiness	Migliorare le mie condizioni di vita
happiness	Vivere bene
happiness	Vivere sereno
help others	Aiutare qualcuno
help others	Comprendere le altre persone
help others	Non far sentire solo qualcuno
help others	Offrire sostegno a qualcuno
help others	Proteggere qualcuno[(a)]
hide and lie	Raggirare qualcuno[(a)]
hide and lie – (removed)	Mentire per proteggere me stesso[(a)(c)]
hide and lie – (removed)	Nascondere i propri sbagli[(a)(b)]
hide and lie - hide	Nascondere una verità che farebbe male agli altri[(a)(b)]
hide and lie - hide	Tenere un segreto[(a)(b)]
hide and lie - lie	Ingannare gli altri[(a)(b)]
hide and lie - lie	Nascondere le mie vere intenzioni[(a)(b)]
hurt someone	Far soffrire qualcuno
hurt someone	Fare del male a qualcuno
hurt someone	Ferire qualcuno
justice and fairness - fairness	Dare un giudizio imparziale[(a)(b)]
justice and fairness - fairness	Premiare solo chi se lo merita[(a)(b)]
justice and fairness - justice	Non creare disuguaglianze[(a)(b)]
justice and fairness - justice	Ottenere giustizia[(a)(b)]
loneliness, not being noticed by others	Evitare che gli altri si accorgano della mia presenza
loneliness, not being noticed by others	Evitare di espormi
loneliness, not being noticed by others	Non sentirmi al centro dell'attenzione
loneliness, not being noticed by others	Passare inosservato
loneliness, not being noticed by others	Stare per i fatti miei
loneliness, not being noticed by others	Stare solo
make others feel loved	Chiarire i malintesi con gli altri
make others feel loved	Far sapere a qualcuno che può contare su di me
make others feel loved	Far sentire agli altri la mia vicinanza
make others feel loved	Far sentire qualcuno amato
make others feel well	Far sentire gli altri a proprio agio
make others feel well	Infondere sicurezza negli altri
make others happy	Far felice qualcuno

make others happy	Far stare bene gli altri
make others happy	Fare del bene
manipulate others	Fare in modo che altri facciano ciò che conviene a me
manipulate others	Fare in modo che gli altri facciano ciò che voglio io
manipulate others	Ingraziarmi qualcuno
manipulate others	Manipolare il comportamento di altre persone
persuasion	Convincere gli altri
persuasion	Far prevalere la mia idea o il mio punto di vista su quello degli altri
persuasion	Far vedere agli altri che ho ragione
persuasion	Persuadere gli altri della mia idea
physical satisfaction	Concedermi dei piaceri[(a)]
physical satisfaction	Provare piacere carnale[(a)]
physical satisfaction	Soddisfare istinti e pulsioni[(a)]
prove one's worth	Dimostrare le mie capacità
prove one's worth	Far vedere quanto valgo
prove one's worth	Farmi valere
prove one's worth	Mostrare di essere sicuro di me
prove one's worth	Mostrare il mio valore
prove one's worth	Non dovermi mai accontentare
prove one's worth	Ottenere ciò che merito
put oneself first	Mettere i miei interessi e bisogni prima di quelli degli altri[(a)]
put oneself first	Mettere me stesso al primo posto[(a)]
put oneself first	Tenere per me ciò che è mio, non condividere[(a)]
rebellion	Disobbedire alle regole
rebellion	Ribellarmi
rebellion	Trasgredire le regole
rebellion	Violare le convenzioni sociali
religious faith	Avere fede religiosa
religious faith	Praticare le tradizioni religiose
religious faith	Rispettare il mio credo
religious faith	Trovare un rifugio spirituale
respect others	Essere sempre corretto con gli altri
respect others	Mostrare rispetto
respect others	Rispettare gli altri
romantic love	Essere emotivamente connesso a un partner romantico
romantic love	Essere innamorato
romantic love	Trovare un partner
sobriety	Accontentarsi di quello che si ha[(a)]
sobriety	Non darmi delle arie con gli altri[(a)]
sobriety	Non essere troppo appariscente agli occhi degli altri[(a)]
sobriety	Non mettere in ombra altre persone[(a)]
sobriety	Non volere troppo[(a)]
sobriety	Vivere una vita semplice[(a)]
stability of relationships	Avere relazioni solide con gli altri
stability of relationships	Mantenere buoni rapporti con qualcuno

take the easy way	Evitare di fare fatica[(a)]
take the easy way	Evitare di faticare per raggiungere i miei obiettivi[(a)]
take the easy way	Prendere delle scorciatoie per ottenere ciò che voglio[(a)]
tell the truth	Dire la verità senza omissioni[(a)]
tell the truth	Non ingannare nessuno[(a)]
tell the truth	Scoprire la verità[(a)]
tell the truth	Smascherare le persone false[(a)]
test someone	Accertarsi dell'onestà di qualcuno[(a)]
test someone	Capire se mi posso fidare di qualcuno[(a)]
test someone	Provare la lealtà di qualcuno[(a)]
think, reflect	Analizzare bene le situazioni
think, reflect	Non lasciare nulla al caso
think, reflect	Prendere buone decisioni
think, reflect	Prendere decisioni giuste
think, reflect	Prevedere le conseguenze delle mie azioni
think, reflect	Riordinare le idee
think, reflect	Trovare la soluzione migliore quando ho un problema
think, reflect	Valutare tutte le opzioni prima di prendere una decisione
win	Essere vincente
win	Vincere

Note. (a) the item was generated in Study1; (b) In Study 2, items of goals “justice and fairness” and “hide and lie” were eventually assigned to more specific goals (i.e., “justice”, “fairness, “hide”, “lie”) or removed. We indicated the specific goals for those items, or that the item was removed.

Table S3. *Study 2. Oblimin-rotated component loadings (pattern matrix) of items belonging to goals Justice and Fairness*

Item	*Justice*	*Fairness*
11	-.09	**.85**
77	.14	**.70**
118	**.71**	.25
182	**.87**	-.13

Note. Eigenvalues before rotation were 1.50, 1.09, .80, and .60. The correlation between the components was r = .09. Loadings > .40 are bolded. Item text is reported in Table S2.

Table S4. *Study 2. Oblimin-rotated component loadings (pattern matrix) of items belonging to goals Hide and Lie*

Item	*Lie*	*Hide*
22	-.21	**.70**
86	**.83**	-.07
107	.01	**.73**
113	**.70**	.13
130*	**.40**	**.49**
139*	**.50**	**.46**
155	**.79**	-.13

Note. Eigenvalues before rotation were 2.64, 1.22, 0.86, and 0.73, 0.64, 0.50, 0.40. The correlation between the components was $r = .20$. Loadings > .40 are bolded. Item text is reported in Table S2.

* The item was removed from the final set because of sizeable cross-loadings.

Table S5. *Study 2. Reliability of 50 goal goals of the preliminary GHGD questionnaire.*

Goal	Nitem	Cronbach's alpha
aggress	4	.73
authenticity	5	.75
autonomy	3	.81
avoid being deceived	2	.50
avoid being excluded or ignored	2	.74
avoid false hope	2	.63
avoid hurting someone	3	.68
avoid judgment of others	3	.50
avoid pain	4	.65
be a good example	2	.63
be accepted	2	.70
be admired	7	.88
be appreciated by others	6	.84
be better than others	4	.82
be rich	5	.71
be safe	8	.83
be satisfied with oneself	7	.83
be true to one's ideals and principles	4	.61
be trusted	3	.77
express one's point of view	4	.69
fairness	2	.41
feeling superior to others	4	.74
gain personal advantage	2	.69
get what you want at any cost	4	.73
happiness	4	.76
help others	5	.86
hide	2	.36
hurt someone	3	.73
justice	2	.46
lie	3	.71
loneliness, not being noticed by others	6	.80
make others feel loved	4	.79
make others feel well	2	.69
make others happy	3	.81
manipulate others	4	.82
persuasion	4	.83
physical satisfaction	3	.63
prove one's worth	7	.84
put oneself first	3	.60
rebellion	4	.82
religious faith	4	.74
respect others	3	.71
romantic love	3	.85
sobriety	6	.72
stability of relationships	2	.59
take the easy way	3	.71

tell the truth	4	.59
test someone	3	.64
think, reflect	8	.82
win	2	.82

Table S6. *Study 2. Correlation between goals and HEXACO traits.*

	H	E	X	A	C	O	SI	FA	MO	GR
Aggress	-.43***	-.06	-.05	-.42***	-.13*	-.05	-.31***	-.36***	-.30***	-.34***
Authenticity[a]	.57***	.02	.35***	.27***	.23***	.30***	.64***	.59***	.37***	.52***
Autonomy[a]	.43***	.00	.18**	.22***	.29***	.24***	.44***	.45***	.26***	.33***
Avoid being deceived[a]	.31***	.05	.04	.10	.23***	.18**	.34***	.35***	.19***	.15**
Avoid being excluded or ignored	.02	.22***	.23***	-.06	.03	.04	.08	.11	-.02	.01
Avoid false hope	.17**	.16**	-.14*	.16**	.03	-.12*	.10	.15**	.27***	.08
Avoid hurting someone[a]	.50***	.20***	.14*	.32***	.18**	.14*	.41***	.47***	.44***	.47***
Avoid judgment of others	.08	.32***	-.06	.00	-.03	.07	.10	.13*	.14*	.07
Avoid pain	.12*	.14*	-.09	.13*	.11	-.15**	.10	.14*	.17**	.03
Be a good example	.36***	-.05	.39***	.16**	.29***	.26***	.41***	.41***	.22***	.38***
Be accepted	-.02	.20***	.17**	-.10	.03	.00	.04	.07	-.02	.00
Be admired	-.14*	.04	.38***	-.17**	-.01	.11*	.01	-.01	-.20***	-.08
Be appreciated by others	.05	.19***	.32***	-.03	.08	.13*	.14*	.16**	.02	.08
Be better than others	-.08	-.25***	.43***	-.18**	.18**	.23***	.11	.08	-.23***	-.05
Be rich	.00	-.04	.20***	-.02	.08	.13*	.11	.13*	-.05	-.06
Be safe	.45***	.08	.03	.33***	.37***	-.07	.38***	.44***	.40***	.29***
Be satisfied with oneself[a]	.48***	-.02	.30***	.32***	.30***	.34***	.51***	.51***	.30***	.34***
Be true to one's ideals and principles[a]	.62***	.05	.21***	.28***	.36***	.32***	.62***	.65***	.39***	.46***
Be trusted	.37***	.07	.34***	.19***	.22***	.22***	.40***	.42***	.23***	.39***
Express one's point of view[a]	.43***	.01	.24***	.13*	.22***	.31***	.53***	.51***	.25***	.34***
Fairness	.20***	-.16**	.03	.06	.16**	.16**	.21***	.23***	.08	.13*
Feeling superior to others	-.32***	-.13*	.20***	-.29***	.01	.10	-.13*	-.18**	-.34***	-.26***
Gain personal advantage	-.20***	-.11	.06	-.09	.02	-.03	-.12*	-.10	-.15*	-.27***
Get what you want at any cost[b]	-.44***	-.09	.03	-.24***	-.20***	-.05	-.30***	-.34***	-.33***	-.37***
Happiness[a]	.45***	.05	.17**	.35***	.26***	.21***	.45***	.46***	.30***	.28***
Help others[a]	.53***	.14*	.34***	.26***	.25***	.24***	.48***	.51***	.44***	.59***
Hide	.15*	.03	-.06	.11	.05	.06	.10	.15**	.21***	.18**
Hurt someone[b]	-.34***	-.10	-.05	-.22***	-.15**	-.07	-.27***	-.31***	-.24***	-.23***
Justice[a]	.43***	.27***	.15*	.16**	.14*	.22***	.40***	.43***	.28***	.40***
Lie[b]	-.56***	-.09	-.07	-.31***	-.17**	-.12*	-.48***	-.50***	-.33***	-.39***
Loneliness, not being noticed by others	-.06	.06	-.58***	.10	-.08	-.20***	-.15**	-.14*	.12*	-.07

Make others feel loved[a]	.52***	.24***	.32***	.27***	.18**	.28***	.47***	.51***	.41***	.54***
Make others feel well	.38***	.12*	.37***	.18**	.17**	.22***	.37***	.40***	.27***	.43***
Make others happy[a]	.53***	.13*	.31***	.30***	.25***	.24***	.46***	.51***	.44***	.55***
Manipulate others[b]	-.47***	-.09	.07	-.29***	-.14*	-.02	-.33***	-.34***	-.34***	-.37***
Persuasion	-.23***	-.12*	.12*	-.24***	.02	.04	-.07	-.10	-.25***	-.23***
Physical satisfaction	.05	-.03	.20***	.07	.04	.18**	.12*	.10	-.02	.04
Prove one's worth	.26***	.01	.39***	.07	.23***	.34***	.36***	.37***	.10	.24***
Put oneself first	-.17**	-.12*	-.03	-.04	.02	.02	-.07	-.08	-.18**	-.22***
Rebellion	-.35***	-.03	.02	-.26***	-.31***	.13*	-.23***	-.29***	-.27***	-.15**
Religious faith	.18**	-.10	.10	.08	.17**	.02	.22***	.20***	.13*	.14*
Respect others[a]	.65***	.04	.18**	.42***	.38***	.14*	.58***	.62***	.52***	.56***
Romantic love	.24***	.13*	.25***	.11*	.14*	.10	.25***	.25***	.15**	.20***
Sobriety[a]	.40***	.04	-.12*	.32***	.19***	-.15*	.24***	.31***	.47***	.34***
Stability of relationships[a]	.44***	.13*	.28***	.25***	.21***	.22***	.45***	.47***	.32***	.33***
Take the easy way	-.25***	.00	-.11	-.08	-.14*	-.04	-.18**	-.15*	-.14*	-.26***
Tell the truth[a]	.48***	.05	.12*	.16**	.25***	.24***	.56***	.51***	.29***	.43***
Test someone[a]	.30***	.07	.09	.12*	.22***	.20***	.34***	.36***	.23***	.26***
Think, reflect	.41***	.01	.10	.19***	.36***	.20***	.44***	.46***	.28***	.29***
Win	-.06	-.14*	.36***	-.12*	.11	.17**	.14*	.09	-.23***	-.07

Note. H = Honesty-Humility, E = Emotionality, X = eXtraversion, A = Agreeableness vs. Anger, C = Conscientiousness, O = Openness to experience, SI = Sincerity, FA = Fairness, MO = Modesty, GR = Greed-avoidance.

[a] The goal was selected as part of the goals of honesty (GH) scale.

[b] The goal was selected as part of the goals of dishonesty (GD) scale.

*p < .05, **p < .01, ***p < .001

Table S7. *Study 3. Measurement invariance of self- and informant-report forms of the Adjective Checklist of Honesty*.

Model	χ^2	*df*	*p*-value	RMSEA	SRMR	CFI	AIC	BIC
1. Configural Inv.	671.55	362	<.001	.053	.045	.942	38958.65	39788.07
2. Weak Inv.	760.64	401	<.001	.054	.067	.932	38969.74	39627.10
Δ2-1	89.09	39	<.001	.001	.022	-.009	11.09	-160.98
3. Strong Inv.	811.7	418	<.001	.056	.068	.926	38986.80	39569.16
Δ3-2	51.06	17	<.001	.001	.001	-.006	17.06	-57.94

Note. RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; CFI = comparative fit index; AIC = Akaike information criterion; BIC = Bayesian information criterion.

Table S8. *Study 3. Component loadings of goals.*

	Self-report		Informant report	
	C1	C2	C1	C2
	GH			
authenticity	**.77**	-.06	**.81**	.07
autonomy	**.54**	.13	**.59**	-.15
avoid being deceived	**.69**	.27	**.71**	.16
avoid hurting someone	**.65**	-.28	**.67**	-.28
be satisfied with oneself	**.71**	.04	**.70**	-.11
be true to one's ideals and principles	**.79**	-.05	**.79**	-.18
express one's point of view	**.77**	.14	**.81**	.16
happiness	**.66**	.01	**.56**	-.28
help others	**.78**	-.10	**.84**	-.10
justice	**.72**	.00	**.71**	.02
make others feel loved	**.77**	-.10	**.88**	-.02
make others happy	**.74**	-.14	**.81**	-.14
respect others	**.73**	-.22	**.75**	-.19
sobriety	**.49**	-.11	**.60**	-.07
stability of relationships	**.76**	-.04	**.75**	-.01
tell the truth	**.79**	.07	**.83**	.18
test someone	**.67**	.37	**.84**	.32
get what you want at any cost	-.07	**.82**	.01	**.86**
lie	-.11	**.81**	-.08	**.86**
hurt someone	-.15	**.63**	-.08	**.71**
manipulate others	.07	**.87**	.02	**.88**

Note. Component loadings emerged in self- and informant-reports after oblimin rotation. The highest loading for each goal is reported in bold. The correlations among the two components were r = -.27 for self-reports and r = -.24 for informant-reports.

Table S9. *Study 3. Pearson's correlations of goals with personality traits and facets, in self-report data*

	GH	GD	GC	GU
H (HAS)	.61***	-.45***	.48***	-.15**
E (HAS)	.06	-.06	.01	.22***
X (HAS)	.25***	-.03	.10*	-.16**
A (HAS)	.30***	-.39***	.29***	-.19***
C (HAS)	.30***	-.14**	.40***	-.29***
O (HAS)	.24***	-.08	.13*	-.08
H (HEXACO-PI-R)	.35***	-.68***	.18***	-.25***
E (HEXACO-60)	.21***	-.16**	.15**	.14**
X (HEXACO-60)	.21***	-.05	.11*	-.27***
A (HEXACO-60)	.06	-.26***	.00	-.17***
C (HEXACO-60))	.32***	-.27***	.46***	-.37***
O (HEXACO-60)	.25***	-.14**	.14**	-.05
SI (ACH)	.55***	-.32***	.41***	-.12*
FA (ACH)	.59***	-.34***	.49***	-.12*
MO (ACH)	.45***	-.29***	.41***	-.02
GR (ACH)	.48***	-.32***	.32***	-.10*
SI (HEXACO-PI-R)	.32***	-.57***	.16***	-.20***
FA (HEXACO-PI-R)	.39***	-.58***	.32***	-.29***
MO (HEXACO-PI-R)	.27***	-.51***	.11*	-.15**
GR (HEXACO-PI-R)	.06	-.36***	-.07	-.10

Note. Correlations were estimated on the entire sample of N = 400 individuals. GH = Goals of Honesty, GD = Goals of Dishonesty, GC = Goals of Conscientiousness, GU = Goals of Unconscientiousness.
* $p < .05$, ** $p < .01$, *** $p < .001$

Table S10. *Study 3. Pearson's correlations of goals with personality traits and facets in informant-report data*

	GH	GD	GC	GU
H (HAS)	.50***	-.63***	.33***	-.39***
E (HAS)	.01	-.03	-.11	.18
X (HAS)	.33***	-.06	.26**	-.07
A (HAS)	.29**	-.46***	.22*	-.25**
C (HAS)	.38***	-.27**	.55***	-.28**
O (HAS)	.26**	-.04	.07	-.04
H (HEXACO-PI-R)	.31***	-.70***	.13	-.49***
E (HEXACO-60)	.25**	-.07	.19*	.15
X (HEXACO-60)	.42***	-.04	.37***	-.12
A (HEXACO-60)	.14	-.42***	.06	-.29**
C (HEXACO-60))	.41***	-.31***	.58***	-.46***
O (HEXACO-60)	.22*	-.08	.08	-.12
SI (ACH)	.59***	-.48***	.44***	-.29**
FA (ACH)	.59***	-.54***	.43***	-.31***
MO (ACH)	.42***	-.47***	.28**	-.27**
GR (ACH)	.53***	-.49***	.27**	-.36***
SI (HEXACO-PI-R)	.22*	-.72***	.10	-.51***
FA (HEXACO-PI-R)	.41***	-.56***	.27**	-.45***
MO (HEXACO-PI-R)	.24*	-.57***	.11	-.39***
GR (HEXACO-PI-R)	.12	-.41***	-.07	-.24**

Note. Correlations were estimated on N = 116 individuals who were rated by at least one informant. GH = Goals of Honesty, GD = Goals of Dishonesty, GC = Goals of Conscientiousness, GU = Goals of Unconscientiousness. * $p < .05$, ** $p < .01$, *** $p < .001$

Table S11. *Study 3. Multiple Poisson regression models predicting observed cheating behavior in the Spot-the-difference task on the full sample of N = 400 individuals.*

	B coefficient Step 1	*B coefficient Step 2*
	Model 1	
Intercept	2.242***[2.209, 2.274]	2.216***[2.183, 2.249]
H (HAS)	-0.040 [-0.082, 0.002]	0.104***[0.055, 0.152]
E (HAS)	-0.156***[-0.082, 0.002]	-0.097***[-0.134, -0.060]
X (HAS)	0.156***[0.122, 0.190]	0.168***[0.133, 0.202]
A (HAS)	-0.106***[-0.145, -0.067]	-0.046*[-0.087, -0.005]
C (HAS)	0.062**[0.024, 0.099]	0.048*[0.010, 0.086]
O (HAS)	0.014 [-0.020, 0.048]	0.015 [-0.019, 0.050]
GH		-0.082***[-0.122, -0.043]
GD		0.235***[0.202, 0.267]
R_N^2	.499	.708
	Model 2	
Intercept	2.218***[2.184, 2.251]	2.214***[2.181, 2.247]
H (HEXACO-PI-R)	-0.245***[-0.279, -0.211]	-0.183***[-0.228, -0.138]
E (HEXACO-60)	-0.062***[-0.096, -0.028]	-0.054**[-0.090, -0.019]
X (HEXACO-60)	0.174***[0.140, 0.209]	0.180***[0.144, 0.216]
A (HEXACO-60)	-0.009 [-0.044, 0.026]	-0.004 [-0.039, 0.031]
C (HEXACO-60))	-0.015 [-0.048, 0.018]	-0.006 [-0.041, 0.028]
O (HEXACO-60)	-0.060***[-0.092, -0.028]	-0.061***[-0.093, -0.028]
GH		0.037 [0.000, 0.073]
GD		0.111***[0.070, 0.151]
R_N^2	.688	.712
	Model 3	
Intercept	2.241***[2.208, 2.273]	2.216***[2.183, 2.249]
SI (ACH)	0.038 [-0.041, 0.117]	0.110**[0.030, 0.190]
FA (ACH)	-0.068 [-0.162, 0.026]	-0.048 [-0.144, 0.047]
MO (ACH)	0.047 [-0.008, 0.101]	0.051 [-0.003, 0.106]
GR (ACH)	-0.061*[-0.112, -0.010]	-0.036 [-0.087, 0.015]
E (HAS)	-0.155***[-0.191, -0.120]	-0.087***[-0.125, -0.049]
X (HAS)	0.167***[0.131, 0.202]	0.177***[0.141, 0.213]
A (HAS)	-0.110***[-0.149, -0.070]	-0.026 [-0.067, 0.015]
C (HAS)	0.061**[0.023, 0.099]	0.059**[0.021, 0.097]
O (HAS)	0.028 [-0.008, 0.065]	0.021 [-0.016, 0.059]
GH		-0.074***[-0.114, -0.034]
GD		0.226***[0.194, 0.258]
R_N^2	.507	.706
	Model 4	
Intercept	2.204***[2.171, 2.238]	2.200***[2.166, 2.234]
SI (HEXACO-PI-R)	-0.143***[-0.183, -0.104]	-0.105***[-0.147, -0.062]
FA (HEXACO-PI-R)	0.081***[0.040, 0.123]	0.108***[0.065, 0.151]
MO (HEXACO-PI-R)	-0.063**[-0.102, -0.025]	-0.034 [-0.074, 0.007]
GR (HEXACO-PI-R)	-0.200***[-0.238, -0.162]	-0.189***[-0.227, -0.150]
E (HEXACO-60)	-0.112***[-0.148, -0.076]	-0.097***[-0.134, -0.059]
X (HEXACO-60)	0.124***[0.088, 0.160]	0.135***[0.098, 0.172]
A (HEXACO-60)	-0.017 [-0.053, 0.019]	-0.014 [-0.049, 0.022]
C (HEXACO-60))	-0.052**[-0.087, -0.018]	-0.042*[-0.077, -0.006]
O (HEXACO-60)	-0.063***[-0.095, -0.031]	-0.061***[-0.094, -0.027]
GH		0.020 [-0.094, -0.027]
GD		0.128***[0.087, 0.169]
R_N^2	.758	.779
	Model 5	
Intercept	2.268***[2.236, 2.299]	2.249***[2.217, 2.281]
GC	-0.048**[-0.079, -0.017]	-0.010 [-0.061, 0.040]
GU	0.130***[0.099, 0.161]	0.028 [-0.007, 0.064]
GH		0.016 [-0.036, 0.069]

GD			0.214^{***}[0.181, 0.248]
	R^2_N	.162	.445

Note. This table is analogous to table 7, but the models were fitted on the full sample of N = 400 individuals. In all models, cheating frequency in the Spot-the-difference task is the dependent variable. In Step 1, a different set of predictors is entered in each model. Model 1 = HEXACO-traits assessed with the HAS (Romano et al., 2023); Model 2 = HEXACO-traits assessed with the HEXACO-PI (Lee & Ashton, 2004); Model 3 = HEXACO traits other than Honesty-Humility assessed with the HAS, plus facets of Honesty-Humility assessed with ACH (Amenta et al., 2025); Model 4 = HEXACO traits other than Honesty-Humility, plus Honesty-Humility facets assessed with the HEXACO-PI; Model 5 = Goals of Conscientiousness (GC) and Unconscientiousnes (GU; Costantini et al., 2020a). In Step 2, Goals of Honesty (GH) and Dishonesty (GD) are included as additional predictors in all models. For each model, we also report R^2_N = Nagelkerke's (1991) R-squared index.
GH = Goals of Honesty, GD = Goals of Dishonesty, GC = Goals of Conscientiousness, GU = Goals of Unconscientiousness, H = Honesty-Humility, E = Emotionality, X = eXtraversion, A = Agreeableness vs. Anger, C = Conscientiousness, O = Openness to experience, SI = Sincerity, FA = Fairness, MO = Modesty, GR = Greed-Avoidance.
95% Confidence intervals for each parameter are reported in square brackets.
* $p < .05$, ** $p < .01$, *** $p < .001$
Statistical significance of individual regression coefficients was assessed using z-tests.

4

Motivational Dynamics of Honesty-Humility: An Ecological Momentary Assessment Study of Goals, States, and Trait Development

Galkina, A., Viechtbauer, W., Di Masi, A., Amenta, S., Perugini, M., Costantini, G. (2025), Motivational Dynamics of Honesty-Humility: An Ecological Momentary Assessment Study of Goals, States, and Trait Development. (Submitted).

Abstract

Research has demonstrated that goals explain personality variation, but the motivational processes' role in Honesty-Humility expression remains underexplored. Understanding how goals relate to Honesty-Humility would have important implications for personality theory and interventions. A sample of 198 participants completed a three-phase study: Baseline assessment (T1), 15-day ecological momentary assessment (EMA) with five daily prompts, and follow-up assessment (T2). Participants rated the momentary importance of 21 goals (17 honesty-related, four dishonesty-related) and their Honesty-Humility states, along with lying behavior. Multilevel modeling examined between- and within-person associations of goals with Honesty-Humility states and lying behavior at contemporaneous and lagged levels. A latent-growth model estimated change trajectories in goals and states during EMA and their associations with trait-level change. Results confirmed the main pre-registered hypothesis that goals predicted Honesty-Humility at the contemporaneous level. Similar results were found at the lagged level. Lying behavior was predicted by goals at the contemporaneous level but not across time. The latent-growth model demonstrated that state-level changes in Honesty-Humility mediated goal effects in predicting trait Honesty-Humility changes from T1 to T2. These findings support a dynamic understanding of Honesty-Humility, demonstrating that momentary goal prioritization robustly predicts honest behavior and can facilitate bottom-up personality change.

Introduction

One of the fundamental issues of personality psychology has traditionally been to provide descriptions of systematic differences between individuals (traits) that could explain differences in their behavior (Allport, 1961; Heller et al., 2007; McCrae & Costa, 1987; McCrae & Sutin, 2018). Various models describing personality dimensions were proposed, among which the most prominent is the five-factor model (Costa & McCrae, 1992; Goldberg, 2002) and the more recent six-factor HEXACO model (Ashton et al., 2014; Lee & Ashton, 2008). However, some have argued that such structural models need to be more clearly connected to psychological processes (Baumert et al., 2017), for two reasons: first, to link higher-level personality organization to the lower-level causal mechanisms of behavior, and second, to account for the fact that there is no perfect correspondence between observed personality structures and underlying processes, which may interact in much more complex patterns than what would be predicted solely by structural models (Cramer et al., 2012; Mõttus et al., 2020).

In fact, recent personality research has revealed a complex interplay between stable structural characteristics defined by traits and more dynamic components reflected in states, which resulted in an integrative approach to personality (Fleeson, 2001; John & Chaplin, 1988; Mischel, 2004). In this view, personality states have been defined as short-term, continuous, concrete ways of acting, feeling, and/or thinking (Fleeson, 2001). An individual's state/behavior varies from moment to moment, and these variable states form a certain distribution. The central tendency of such a distribution would represent this individual's trait. Thus, being directly connected, traits and states are considered to share a certain degree of isomorphism and to be described by the same terms and scales (Fleeson, 2001; Fleeson et al., 2002; Fleeson & Jayawickreme, 2015).

Studies have shown that traits and states are related to each other (Fleeson & Gallagher,

2009; Heller et al., 2007); however, the importance of disentangling the role of these two components to study their specific role is often emphasized in the literature (Allemand & Mehl, 2017; Augustine & Larsen, 2012; Bleidorn, 2009; R. E. Wilson et al., 2017). While, in general, those who scale high on a certain trait appear to engage more often in trait-relevant behaviors (Wu & Clark, 2003), one of the key questions of personality psychology is to explain why an individual with a certain level of trait expression behaves differently in different situations (Baumert et al., 2017).

Ecological Momentary Assessment (EMA) studies represent an important research instrument that enables the exploration of the interrelationships between personality traits, states, and their behavioral manifestations, because they allow researchers to study both stable and variable aspects of personality (Fleeson & Noftle, 2012). This method allows for the assessment of personality manifestations on repeated occasions, and thus enables the estimation of the dynamics of personality also within an individual at the idiographic level, in addition to a between-person level (Conner et al., 2009). The possibility of disentangling within- (WP) from between-person (BP) effects is particularly relevant for distinguishing between an average manifestation of a trait that differs from person to person from its situational manifestation that can be different for the same person in different circumstances. Fleeson (2001) demonstrated the empirical validity of this approach for personality research using EMA data. This work and its later empirical developments became the basis of the Whole Trait Theory (Conner et al., 2009; Fleeson, 2004; Fleeson & Jayawickreme, 2015).

WTT (Fleeson & Jayawickreme, 2015) provides a framework for understanding the WP variability by conceptualizing personality traits as density distributions of states, where states represent momentary manifestations of the trait that vary systematically in response to social-cognitive mechanisms. As demonstrated by McCabe and Fleeson (2012, 2016), goals substantially account for personality state variance, indicating that motivational processes serve

as key mechanisms driving momentary trait expression. This state-level understanding connects to broader developmental processes through theoretical proposals by Wrzus and Roberts (2017), which posit that long-term personality development occurs through repeated short-term, situational processes, emphasizing dynamic, process-oriented mechanisms rather than static individual differences as the primary drivers of personality change.

Recent views of personality emphasize the importance of goals for both state and trait variation (Bleidorn et al., 2010; Buchinger et al., 2024; Costantini et al., 2020; Costantini & Perugini, 2018; Di Sarno et al., 2023; Fleeson & Jayawickreme, 2015; Quirin et al., 2020). In this view, for example, changes in the significance of a goal (e.g., the goal to have fun) might change the probability of being engaged in certain behaviors (e.g., being talkative), and these changes, over time, may add up in bottom-up personality trait change (McCabe & Fleeson, 2012, 2016; Wrzus & Roberts, 2017).

Previous studies connecting personality and goals focused primarily on the Big Five traits (e.g., Bleidorn et al., 2010; Buchinger et al., 2024; Costantini et al., 2020; Costantini & Perugini, 2018; Lüdtke et al., 2009; Reisz et al., 2013), which do not encompass Honesty-Humility, the dimension that was later added in the HEXACO model (Ashton & Lee, 2008, 2020). However, Honesty-Humility represents a fundamental cornerstone of societal functioning, demonstrating strong associations with prosocial and moral behaviors and attitudes (Ashton & Lee, 2008). The examination of honesty as both a moral virtue and a personality characteristic holds significant relevance for understanding and predicting socially meaningful behaviors (Miller et al., 2021). Limited empirical investigation has examined the personal goals that underlie honest and dishonest behavioral expression. This research gap is particularly notable given the growing theoretical consensus that honesty cannot be adequately understood through behavioral manifestations alone (e.g., truth-telling or cheating avoidance), but requires examination of the underlying motivational structures and goal orientations that

guide such behaviors (Miller, 2020; Miller et al., 2021; A. T. Wilson, 2018). Certain individuals may engage in truthful behavior not merely for being compliant with social norms, but through commitment to the superordinate goal of maintaining personal trustworthiness. On the other hand, others may engage in deceptive behavior for prosocial motives, such as emotional protection of others. This ambiguity complicates the direct correspondence between honesty, behavioral outcomes, and motivational antecedents. While individuals high in Honesty-Humility demonstrate consistent truthful behavior even when deception might serve prosocial functions (Thielmann et al., 2024), suggesting internalized moral imperatives, parallel research indicates that high Honesty-Humility individuals may engage in dishonest behavior when such actions are perceived to serve moral or benevolent objectives (Choshen-Hillel et al., 2020; McArthur et al., 2024; Ścigała et al., 2020). These empirical patterns suggest that Honesty-Humility may represent not merely behavioral predispositions, but rather a motivational framework, thereby necessitating investigating the specific goals that characterize this personality dimension.

Although, as stated earlier, the understanding of Honesty-Humility cannot be limited only to demonstrating or abstaining from cheating behavior, this trait remains crucial in explaining lying, cheating behavior, and exploiting others (Heck et al., 2018; McArthur et al., 2024). While deceptive behavior encompasses a spectrum ranging from intentional hiding of harmful behavior to inconsequential lies told for prosocial purposes (e.g., avoiding hurting others' feelings), empirical research reveals that lying constitutes a ubiquitous aspect of daily social interaction (Vrij, 2008). Previous works indicated substantial individual variation in lying frequency, with approximately 60% of individuals reporting zero daily lies, while a minority engage in frequent deception (Serota et al., 2022; Serota & Levine, 2015). Research has identified multiple motivational categories underlying deceptive behavior, such as instrumental gain maximization, loss avoidance, relational bond maintenance, protection of others, social

politeness, privacy protection, entertainment, hurting others, and other motivations (Caspi & Gorsky, 2006; DePaulo et al., 2004; DePaulo, Kashy, et al., 1996; Levine et al., 2016). These examples showed that lying and deceptive behavior are characterized by a significant motivational complexity.

Not surprisingly, most scholars have focused predominantly on the motives of dishonesty and lying as the opposite of honesty. For example, motivated by cases of unethical conduct within organizational settings, research in behavioral ethics and behavioral business ethics has examined the antecedents of dishonest behavior (De Cremer & Moore, 2020). Several studies explored the problem of goals of dishonesty in the academic context, revealing a particular role of achievement goals in fostering cheating behavior (Daumiller & Janke, 2020; Janke et al., 2019) and the negative association between prioritizing learning goals and dishonesty in students and academicians (Jordan, 2001; Marsden et al., 2005). These methodological approaches, however, embody an implicit premise that individuals face a binary choice when confronting ethical dilemmas — namely, the option to deceive or to be truthful. Moreover, previous research has identified the most important goals associated with Honesty-Humility and showed that goals of honesty and goals of dishonesty represented two distinguishable and mildly negatively associated motivations, which played different roles in predicting actual cheating behavior (Galkina et al., 2025). Thus, focusing on the goals related to honesty, in addition to those of dishonesty and deception, is fundamental for our understanding of both the trait Honesty-Humility and the respective behavior of individuals in different situations. This approach would be in line with the theoretical framework stating that honesty represents a multifaceted phenomenon that transcends singular behavioral acts but rather constitutes a communicative process that is inherently context-dependent and cannot be divorced from situational factors (Cooper et al., 2023).

Thus, in the current work, we focus on the momentary states of Honesty-Humility and their

interplay with the subjective importance of related goals. For this study, we used a questionnaire validated in previous work (Galkina et al., 2025), the Goals of Honesty and Dishonesty (GH-GD) questionnaire, which we adapted to investigate the momentary goals in an intensive longitudinal framework (for a similar approach, see Di Sarno et al., 2023). This study was pre-registered (https://aspredicted.org/cdss-24q4.pdf) with the main hypothesis that the momentary experience of specific goals related to honesty and dishonesty would predict the state of Honesty-Humility at the contemporaneous level. In order to account for both intra- and interindividual differences, independent variables were split into a between-person (grand-mean centered mean state score) and a within-person part (person-mean centered individual state scores). Furthermore, additional pre-registered analyses were performed on the temporal associations between goals and Honesty-Humility states and temporal and contemporaneous associations between goals and self-reported lying behavior. Moreover, we investigated the psychometric properties of the EMA version of the GH/GD-questionnaire, assessing its factor structure at the WP and BP levels. Finally, we explored whether the change occurring during the EMA phase related to mid-term change, as assessed before and after the study.

Materials and methods

Participants and procedure

Participants were recruited through the university's Sona System, word of mouth, printed announcements on campus, and social media. All participants provided informed consent online before completing the first survey. The study was judged as involving minimal risk for the participants by the Committee for Research in Psychology (CRIP) of the University of Milan-Bicocca (protocol number RM-2022-590 as of November 18th, 2022).

The study involved three main phases: a baseline questionnaire at the start (Time 1, T1), a 15-day Ecological Momentary Assessment (EMA) period, and a final questionnaire at the

end (Time 2, T2). Both the T1 and T2 questionnaires were completed via the Qualtrics survey platform, while the EMA phase utilized a mobile app that we called *eSampler*, which we developed based on the Experience Sampler project (Thai & Page-Gould, 2018) and made available for Android smartphones, due to strict requirements for app development in the Apple/iOS environment. Participants were informed about the requirement of possessing an Android smartphone in the study announcements.

At the beginning of the questionnaire, participants were asked to install *eSampler* and were provided with instructions on how to download and set up the app correctly. Both written and video instructions were tailored to match frequently used smartphone brands (Samsung, Huawei, Xiaomi). Participants owning other Android-based smartphones were provided with access to all of these instructions as well as general guidelines for setting up the app. Before continuing the survey, participants were asked to confirm the successful app installation or request help. This was done to avoid cases where participants completed the T1 questionnaire but could not proceed with the EMA phase for technical reasons. Twenty participants requested help with app installation, 11 of whom successfully installed the app with the assistance and completed the T1 survey afterward.

The T1 questionnaire contained three attention-check questions, designed to resemble the main survey items, but instructing participants to select a specific response (e.g., “This is a check question: Please select four and proceed with the questionnaire”). Following pre-registered criteria, participants who missed more than one attention check were disqualified from the study (N = 5).

During the EMA phase, participants were prompted to respond to short surveys five times daily at fixed intervals (10:00, 13:00, 16:00, 19:00, 22:00) over 15 days (for a total of 75 prompts). The eSampler app sent three additional reminders every 15 minutes after the prompt if the questionnaire was not completed. After one hour, if not completed, the questionnaire was

considered as missed. The day after finishing the EMA phase, participants received the T2 questionnaire, which included the same measures administered at T1.

Two hundred and seventy-four participants successfully completed the T1 questionnaire. Following pre-registered criteria, we excluded from analysis those participants who responded to less than 20% of EMA surveys, i.e., less than 15 surveys out of 75, or who did not provide any EMA data (N = 76). If not otherwise specified, analyses refer to the remaining sample of N = 198 participants. Most of these participants also completed the T2 survey (N = 191). The socio-demographic characteristics of both the total sample and the subsample, with the completed T2 assessments, are presented in Table 1.

Table 1. *Socio-demographic characteristics of the final sample(s)*

	Total sample (T1 and at least 20% EMA)	**Subsample (T2 completed)**
Sample size (*N* females, *N* males, *N* non-binary)	198 (132, 65, 1)	191 (125, 65, 1)
Age: mean (SD, range)	28.17 (10.59, 18 - 66)	28.35 (10.72, 18 - 66)
EMA observations per participant: mean (sd, range)	47.10 (17.28, 15 - 71)	47.90 (16.90, 15 - 71)
Total EMA observations	9 326	9 149
N (%) by level of Education		
Bachelor's degree or higher	113 (57,07%)	109 (57,07%)
High school diploma	81 (40,91%)	78 (40,84%)
Lower than high school	4 (2,02%)	4 (2,09%)
Android smartphone brand, *N* (%)		
Samsung	76 (38,38%)	71 (37,17%)
Xiaomi	55 (27,78%)	53 (27,75%)
Huawei	24 (12,12%)	24 (12,57%)
Other	43 (21,72%)	43 (22,51%)

Note: SD, Standard Deviation

Participants who completed the T1 and T2 questionnaires and at least 75% of EMA prompts (i.e., a minimum of 52 out of 70) received a €10 token. The requirements were clearly outlined in the Informed Consent Form at the beginning of the study. Participants were also advised to notify the researchers of mobile application issues to ensure that affected prompts

were not counted as missed and did not affect the reward.

Measures

T1 and T2 assessments included the questionnaires in the following order.

Personality traits

The *HEXACO Adjective Scale* (HAS; Romano et al., 2023) is an instrument containing 60 adjectives corresponding to the six main dimensions of personality according to the HEXACO model. The Honesty-Humility trait was further investigated with additional items from the Adjective Checklist of Honesty (Amenta et al., 2025). These items were administered together as they share the same format (as in Amenta et al., 2025). Participants were asked to indicate the extent to which each of the adjectives (e.g., "loyal") described them on a scale from 1 (*it does not describe me at all*) to 7 (*it describes me completely*). Cronbach's α reliabilities for the trait subscales at T1 (N = 200) and T2 (N = 191) ranged between 0.78 and 0.90. The ACH Scale can be found in Supplementary Materials S18.

The 60-item version of *HEXACO-PI* (Ashton & Lee, 2009) assesses six major personality traits with 10 items each. The trait Honesty-Humility was measured in more detail with the full 32-item HEXACO-PI scale, which includes four main facets of the trait: Truthfulness/Sincerity, Modesty, Fairness, and Greed-Avoidance (Ashton et al., 2006). Participants indicated their level of agreement with each item (e.g., "I would be quite bored by a visit to an art gallery"), on a scale from 1 (*strongly disagree*) to 5 (*strongly agree*). Cronbach's α reliabilities at T1 (N = 198) ranged between 0.72 and 0.91 for the traits and between 0.79 and 0.85 for the Honesty-Humility facets. At T2 (N = 191), alphas ranged between 0.70 and 0.93 for the traits and between 0.83 and 0.86 for the Honesty-Humility facets.

Subjective importance of honesty- and dishonesty-related goals

The Goals of Honesty/Dishonesty (GH/GD, Galkina et al., 2025) questionnaire assesses

the subjective importance assigned by participants to a list of 78 goals related to honesty (64 items) and dishonesty (14 items). Participants rated the extent to which each goal was important to them, on a scale from 1 (*not at all*) to 5 (*extremely*). In the previous study (Galkina et al., 2025), the goal items were grouped into broader classes: 17 honesty-related and four dishonesty-related classes (see Table S3 for the list of classes). The item scores were first averaged for each goal class, and then the class scores were averaged to compute two separate scores, one for the honesty-related goals (T1 $\alpha = 0.90$, N = 198; T2 $\alpha = 0.93$, N = 191), and another for dishonesty-related goals (T1 $\alpha = 0.73$, N = 198; T2 $\alpha = 0.83$, N = 191).

During the EMA phase, participants completed the following questionnaires.

Momentary Honesty-Humility states

Participants rated the extent to which the ten adjectives assessing Honesty-Humility (haughty, greedy, dishonest, faithful, hypocritical, loyal, honest, sincere, snob, and humble; Romano et al., 2023) characterized them over the preceding hour, using a 7-point Likert scale ranging from 1 (*does not describe me at all during the last hour*) to 7 (*describes me completely during the last hour*). Reliability was estimated for both the WP and BP levels of the measure within a multilevel factor analysis framework (Geldhof et al., 2014) using the *lavaan* R package (version 0.6-17, Rosseel, 2012) and the *semTools* R package (version 0.5-6, Jorgensen et al., 2012). They were $\alpha_{BP} = 0.90$ and $\alpha_{WP} = 0.71$. The intraclass correlation coefficient (ICC) was 0.60 (see also section *Preliminary analysis*).

Momentary importance of honesty- and dishonesty-related goals

The EMA version for the subjective importance of goals of honesty and dishonesty (GH/GD-EMA) questionnaire assessed participants' self-perceived importance of goals related to honesty and dishonesty in the hour before the beep. To develop this measure, we adapted the original GH/GD questionnaire (Galkina et al., 2025) by selecting 21 goal items. Items were

selected based on their suitability for capturing momentary variations of GH/GD (e.g., "help someone", "pursue my ideals"), each representing a goal class associated with Honesty, as identified in the previous study, to ensure the breadth of theoretical content. Participants rated the extent to which each goal had been important to them during the hour before the prompt, on a scale from 1 (*not at all*) to 5 (*extremely*). See Table S5 for the full list of items and related classes. . Seventeen goal items represented honesty-related goals, and the remaining four referred to the goals of dishonesty. Reliability estimates were $\alpha_{BP} = 0.98$ and $\alpha_{WP} = 0.86$ for honesty-related goals, and $\alpha_{BP} = 0.93$ and $\alpha_{WP} = 0.47$ for dishonesty-related goals. ICCs were 0.65 and 0.54 for honesty- and dishonesty-related goals, respectively.

Lying behavior (frequency)

Participants reported on their lying behavior, measured by a single item: "*How many times did you tell lies during the last hour?*" (see Serota et al., 2022) with six response options ranging from *"0"* to *"5 or more"*.

Statistical analyses

Analyses were performed in R (R Core Team, 2022). R code and the data can be found in the Open Science Framework

(https://osf.io/9z67s/?view_only=0826c142ae1845cea599b218756e8187).

The statistical analyses described in this section are summarized in Table 2.

Table 2. *Summary of statistical analysis.*

Step of analyses	Type of analyses	Models estimated
Preliminary analyses	Descriptive statistics	
	Pearson correlations (EMA scales)	
	ICCs (EMA scales)	
	Goals factor structure	Unidimensional, bidimensional structure with

		correlated factors, bidimensional structure with orthogonal factors
	Item-specific ICCs	
	Item-specific distributional properties	
Pre-registered main hypothesis	Stepwise linear mixed-effects regression models predicting state Honesty-Humility	Step 0 predictors: days (temporal trend) Step 1 predictors: + GH and GD (WP component) Step 2 predictors: + GH and GD (BP component) Step 3 predictors: + GH and GD (at T1) Step 4 predictors: + HEXACO Honesty-Humility (at T1) Step 5 predictors: + other HEXACO traits (at T1)
Additional analyses	Stepwise *temporal* linear mixed-effects regression models predicting state Honesty-Humility (lag 1)	Step 0 predictors: days (temporal trend) Step 1 predictors: + Honesty-Humility state at lag 1 (autoregressive effect) Step 2 predictors: + GH and GD at lag 1
	Stepwise Poisson mixed-effects regression models predicting lying behavior	Step 0 predictors: days (temporal trend) Step 1 predictors: + GH and GD (WP component) Step 2 predictors: + GH and GD (BP component) Step 3 predictors: + GH and GD (at T1) Step 4 predictors: + HEXACO Honesty-Humility (at T1) Step 5 predictors: + other HEXACO traits (at T1)
	Stepwise *temporal* Poisson mixed-effects regression	Step 0 predictors: days (temporal trend)

	models predicting lying behavior	Step 1 predictors: + lying behavior at lag 1 (autoregressive effect) Step 2 predictors: + Honesty-Humility state at lag 1 Step 3 predictors: + GH and GD at lag 1
	Latent Growth Model	Mediation model: trait Honesty-Humility change (from T1 to T2) was predicted by both the intercept and slope factors of daily Honesty-Humility, which in turn, were predicted by the corresponding intercept and slope factors of GH and GD.

Preliminary Analyses

Descriptive statistics were calculated for variables measured at T1, EMA, and T2, accompanied by Pearson correlations and ICCs for the EMA variables. ICCs quantify the proportion of variance attributable to BP differences, reflecting variability in random intercepts in a mixed-model framework (Lorah, 2018). The ICCs provide insight into the partitioning of the variance components within the measurement framework. Specifically, these coefficients quantify the proportion of total item variance attributable to systematic differences between individuals, as distinguished from within-individual variation and error components (Eisele et al., 2022). For instance, an ICC value of .30 indicates that 30% of the observed variance in the corresponding item stems from between-individual differences, while the remaining 70% reflects within-individual variability and measurement error.

Preliminary analyses examined the relationships between each momentary goal related to honesty or dishonesty and state Honesty-Humility by computing correlations between person-level aggregated means for each goal and state honesty. Additionally, mixed models

were employed to investigate these relationships further, with state Honesty-Humility as the dependent variable and goals as independent variables, both individually in separate models and collectively in a single model. In subsequent analyses, individual goals were combined into composite scores representing goals related to honesty and dishonesty.

Preliminary Analyses - Factor Structure of Goals

The present investigation systematically examined the factor structure of the EMA version of the GH/GD questionnaire through a comprehensive multilevel CFA approach. This examination was conducted at both the WP and BP levels, providing a thorough understanding of the instrument's psychometric properties across distinct analytical dimensions. The multilevel approach recognizes that psychological constructs may manifest differently at the level of intra-individual variability (WP) versus inter-individual differences (BP), necessitating separate evaluation at each level.

Three competing theoretical models were subjected to empirical scrutiny at each analytical level. The first model implied a unidimensional factor structure, suggesting that all observed variables load onto a single latent construct. The second model hypothesized a bidimensional structure characterized by correlated GH and GD factors, allowing for theoretical overlap and shared variance between these constructs. The third model specified orthogonal factors within a two-factor framework, assuming statistical independence between the GH and GD dimensions.

The evaluation of these competing models employed partially saturated modeling procedures, following the methodological recommendations of Ryu and West (2009). Saturated models impose no structural constraints and consequently achieve a perfect fit with the observed data. When assessing the fit of the WP model, a saturated BP model was specified, and conversely, when evaluating the BP model, a saturated WP model was employed. This

methodological choice addresses a critical concern in mixed modeling—that model misspecification at higher analytical levels may remain undetected when both levels are tested simultaneously.

Model adequacy was quantitatively assessed using a comprehensive suite of level-specific fit indices, including the level-specific Chi-square statistic, Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Squared Residual (SRMR). Acceptable model fit was defined by CFI values of 0.90 or higher, RMSEA values of 0.08 or lower, and SRMR values of 0.08 or lower, following the established standards (Bentler, 1990; Hu & Bentler, 1999).

Preliminary Analyses - Item Characteristics

The comprehensive descriptive statistics and ICCs for all measurement items are presented in Table S1 of the Supplementary Materials, and covariance matrices are documented in Tables S6-S7. The observed ICCs ranged from 0.296 to 0.565, reflecting the different extent to which inter-individual differences contribute to total item variance relative to within-individual fluctuations and measurement error.

Examination of the distributional properties revealed that several items exhibited marked deviations from normality, as evidenced by the skewness and kurtosis values reported in Table S1. These statistical indices provide quantitative indicators of the extent to which response distributions diverge from the multivariate normal distribution that serves as a fundamental assumption underlying the CFA model. Such violations of normality assumptions are frequently encountered and indeed anticipated in empirical research of this nature.

The observed distributional violations carry important implications for model estimation and statistical inference. Previous research has demonstrated that deviations from multivariate normality can result in inflated $\chi 2$ test statistics, leading to the erroneous rejection

of correctly specified models (Curran et al., 1996). While the maximum likelihood estimator employed in the current CFA analyses demonstrates robustness against minor violations of normality assumptions, more substantial deviations can significantly influence analytical outcomes and parameter estimates (Brown, 2015).

It is important to acknowledge that alternative estimators designed to accommodate violations of normality assumptions are available in the statistical literature; however, these robust estimation procedures have not yet been implemented within the mixed-effects modeling capabilities of the lavaan package (Rosseel, 2012). This methodological constraint necessitates cautious interpretation of the present findings, particularly regarding results obtained at the BP level, where model fit indices may exhibit systematic bias due to the distributional violations.

Pre-registered analyses

To evaluate the pre-registered hypothesis concerning the contemporaneous association between state goals and state honesty, we employed a stepwise mixed modeling approach. This analysis was conducted using the *lme4* and *lmerTest* packages (Bates et al., 2015; Kuznetsova et al., 2017). In these models, state Honesty-Humility served as the dependent variable. The predictors consisted of composite scores for the subjective importance of honesty- and dishonesty-related goals (in the EMA phase). Each predictor was partitioned into a BP component (the grand-mean centered mean state score) and a WP component (person-mean centered individual state scores), following the procedures outlined by Di Sarno et al. (2020, 2023) and Wang and Maxwell (2015).

To account for temporal trends, all models included a linear time covariate (scaled in days from 0 to 14)[1]. Random slopes were specified for both time and the WP components of

[1] An alternative possibility would have been to center the variable day as well, i.e., representing the first day as -7, the last day as +7, thus representing the 8th day as zero. In this way, the intercept would have represented the predicted value of honesty-humility in the central moment of the data-collection. These two coding strategies do

the goal predictors. Further, additional predictors assessed at baseline (T1), mean-centered prior to inclusion, were sequentially added to the models. These comprised baseline measures of honest and dishonest goal importance, baseline honesty, operationalized as the scale score of the Honesty-Humility factor of HAS (Romano et al., 2023) and ACH (Amenta et al., 2025), and the remaining HEXACO personality dimensions (scale scores of HAS factors). For each step in the modeling process, we reported marginal R^2 values, reflecting the proportion of variance explained by the fixed effects (Johnson, 2014; Nakagawa & Schielzeth, 2013). These statistics were computed using the *performance* package (Lüdecke et al., 2021) and *MuMIn* package in R (Bartoń, 2025).

Additional pre-registered analyses

Temporal effects on state honesty

In addition to contemporaneous relationships, we investigated the temporal dynamics underlying trait-goal associations. Specifically, we examined whether state Honesty-Humility at time t_i was predicted by state Honesty-Humility (autoregressive effects) and goals (lagged effects) measured at the preceding time point t_{i-1}, where *i* denotes the observation number ranging from 1 to 75. Throughout all analyses, we controlled for linear effects of time, to account for temporal trends in the data, though we did not include random slopes for temporal trends in these temporal models due to model convergence issues.

Lag-1 variables for state Honesty-Humility and goals were derived from their WP components. Following established methodological protocols, we excluded lags when participants missed scheduled surveys and removed overnight lags to maintain temporal continuity (Bringmann et al., 2016). Consequently, these temporal analyses were conducted on

not affect the estimates or interpretation of coefficients other than the intercept. Albeit these two options are equally valid, we reasoned that interpreting the intercept as the estimated levels of honesty-humility at the beginning of the data collection would yield more intuitively interpretable results and it would be coherent with past similar work (Di Sarno et al., 2023)

a reduced dataset comprising 6,109 EMA observations from our total sample (N = 198).

Prediction of lying behavior

We fitted a series of Poisson mixed-effects regression models predicting momentary lying behavior with a number of predictors added in each step, mirroring the procedure followed for predicting Honesty-Humility states. For the contemporaneous model, predictors included composite scores for the subjective importance of honesty- and dishonesty-related goals (in the EMA phase) at the WP and BP levels, baseline measures of goal importance and personality traits, while controlling for a linear temporal trend. In the temporal model, we estimated the autoregressive effects, and the temporal effect of goals and state Honesty-Humility (lagged effects) measured at the preceding time point t_{i-1}, where *i* denotes the observation number ranging from 1 to 75. This modeling strategy allowed testing whether goals would predict lying behavior above and beyond state honesty.

Latent Growth Modeling of Trait Stability and Change

To examine the dynamic mechanisms underlying both stability and change in trait honesty, we modeled a longitudinal process using participants who completed assessments at both time points (N = 191). Our primary objective was to determine whether fluctuations in momentary goals and state Honesty-Humility could account for patterns of trait-level change from baseline (T1) to follow-up (T2). Though traits are usually considered to be quite stable, and relative changes are expected to happen in association with major events across the life span (Roberts et al., 2006), some studies have demonstrated that trait changes can occur even within much shorter time frames, such as weeks (e.g., Hudson & Fraley, 2016; Stieger et al., 2022).

Mirroring Di Sarno and colleagues (2023), we employed latent growth curve modeling (McArdle & Nesselroade, 2003) within a structural equation modeling (SEM) framework, implemented through the *lavaan* R package (Rosseel, 2012). This analytical technique enables

the modeling of individual developmental trajectories by extracting continuous latent factors that capture both initial status and systematic patterns of change across repeated measurements. Latent growth modeling is thus particularly well-suited for examining how changes in momentary experiences may contribute to longer-term personality development.

EMA responses were aggregated to create daily composite scores. Specifically, we computed daily mean values for state Honesty-Humility and momentary goals by averaging each participant's EMA responses within each assessment day. These daily aggregates served as manifest indicators for subsequent latent factor construction, providing a more stable and reliable basis for modeling individual trajectories while reducing the influence of momentary fluctuations. Using these daily person-level averages, we specified a two-factor latent growth model for each of the three focal EMA constructs: honest goals, dishonest goals, and state honesty. The intercept factor was defined by constraining all factor loadings to 1 across the 15-day assessment period, thereby representing each participant's baseline level or initial status for each construct during the EMA procedure. In contrast, the slope factor incorporated linearly increasing factor loadings ranging from 0 to 14, corresponding to each successive day of the assessment period. This specification allowed the slope factor to capture individual differences in the rate and direction of linear change throughout the monitoring window.

The change in trait Honesty-Humility was quantified as a latent factor (ΔH) extracted from the difference scores across all Honesty-Humility facets measured by the ACH questionnaire (Amenta et al., 2025). These change scores were calculated as the difference between follow-up and baseline assessments for each facet ($\Delta_{f1...f4}$ = $f1...f4_{T2}$ − $f1...f4_{T1}$, where f1−f4 represent the respective Honesty-Humility facets). This approach allowed us to capture comprehensive trait-level change while accounting for measurement error across multiple indicators.

We tested a comprehensive mediation model to examine both direct and indirect

pathways through which momentary experiences might influence trait development. In this model, trait Honesty-Humility change (ΔH) was predicted by both the intercept and slope factors of daily Honesty-Humility. These state Honesty-Humility factors were, in turn, predicted by the corresponding intercept and slope factors of honest and dishonest goals. This analytical strategy enabled us to investigate whether goal-directed states serve as mechanisms through which momentary experiences translate into longer-term personality change, while simultaneously examining the direct effects of baseline levels and patterns of change in state Honesty-Humility on trait-level outcomes.

Results

Factor structure of momentary honesty- and dishonesty-related goals

We verified preliminarily the factor structure both at the WP and BP levels. At the WP level, a two-factor model with correlated factors provided the best fit, fitting the data significantly better than a single-factor model and a model with orthogonal GH and GD factors (see Table S2 in Supplementary Materials for model fit and comparisons). The fit measures of the model with two correlated factors indicated a reasonable but not ideal fit, respecting two out of three cut-offs and being close to the third one (RMSEA = 0.058; CFI = 0.867; SRMR = 0.051). The correlation between factors GH and GD was low, but positive, 0.266, $p < .001$. Standardized item loadings on the two correlated factors at the WP level can be found in Table S8.

At the BP level, the model with two correlated factors fitted significantly better than the single-factor one, but only slightly better than the model with independent GH and GD factors. Though the χ^2 test of model differences is significant, there is not enough evidence to prefer the correlated model over the orthogonal one based on established criteria (in particular, RMSEA and CFI). We interpret this as evidence of some correlation present between the factors, which is, though, not high enough to show a considerable difference from the

orthogonal model, which assumes a correlation of 0. For consistency, the correlated factor structure was retained (the structure of the WP and BP factor models with the standardized loadings is illustrated in Figure S3). As in the case of the WP structure, the overall fit of the model was reasonable though not ideal, achieving two out of the three specified criteria (RMSEA = 0.032; CFI = 0.961; SRMR = 0.117). The correlation between the factors GH and GD was low, but positive, similar to the WP structure: 0.299, $p < 0.001$. Table S9 provides estimations of standardized item loadings on the two correlated factors at the BP level.

Test of the pre-registered hypothesis and incremental validity of goals

The main purpose of this work was to explore whether the momentary experience of specific GH and GD goals, as assessed through the GH/GD-EMA questionnaire, could predict state Honesty-Humility at the contemporaneous level. The results of a mixed-effects regression model confirmed this hypothesis, taking into account a possible temporal trend (see Table 3). Goals of honesty were positively associated with state Honesty-Humility at the WP level (Step 1), suggesting that participants reported higher state Honesty-Humility when they prioritized honesty goals more highly. In contrast, goals related to dishonesty negatively predicted self-reported Honesty-Humility states. This model, however, explained only 7% of variance (R^2 = 0.069). The associations with the goals were similar when the BP parts of goals were added to the model (Step 2). Additionally, these BP components showed significant relationships with state Honesty-Humility as well, in the expected direction. In total, this model explained 33% of the variance (R^2 = 0.329). In general, participants who typically gave more importance to honesty-related goals through the EMA phase (and less to those of dishonesty) were more likely to describe their states as honest/humble.

Additionally, further analyses demonstrated that the relationships with goals remained significant even after controlling for baseline T1 goal measures (Step 3, R^2 = 0.336) and for the Honesty-Humility trait assessed at T1 (Step 4, R^2 = 0.367). Interestingly, the importance of

the honest and dishonest goals assessed at T1 did not themselves predict state Honesty-Humility significantly, after taking into account the same goals assessed during the EMA phase and trait Honesty-Humility assessed at T1. Overall, the Step 4 model demonstrated the best fit, while adding the remaining HEXACO-60 dimensions at baseline in Step 5 neither considerably added explained variance to the model ($R^2 = 0.371$) nor enhanced the model fit based on AIC and BIC criteria.

Table 3. *Hierarchical mixed-effects multiple regression models predicting state honesty*

	Step 0			Step 1			Step 2			Step 3			Step 4			Step 5		
IVs	***B***	**SE**	***t***	***B***	**SE**	***t***	***B***	**SE**	***t***	***B***	**SE**	***t***	***B***	**SE**	***t***	**B**	**SE**	**t**
Intercept	5.85	0.048	121.18***	5.78	0.049	117.82***	5.78	0.035	165.99***	5.78	0.034	168.92***	5.78	0.031	183.71***	5.78	0.031	186.28***
day	-0.02	0.005	-3.74***	-0.00	0.003	-1.55	-0.01	0.003	-1.61	-0.01	0.003	-1.59	-0.01	0.003	-1.60	-0.01	0.003	-1.62
GH (wpc)				0.42	0.022	18.96***	0.43	0.022	19.30***	0.43	0.022	19.26***	0.42	0.022	19.34***	0.43	0.022	19.37***
GD (wpc)				-0.21	0.018	-11.60***	-0.21	0.018	-11.57***	-0.21	0.018	-11.61***	-0.21	0.018	-11.62***	-0.21	0.018	-11.65***
GH (bpc)							*0.55*	*0.042*	*13.00****	*0.50*	*0.046*	*10.93****	*0.46*	*0.042*	*10.79****	*0.45*	*0.042*	*10.81****
GD (bpc)							-0.64	0.055	-11.59***	-0.58	0.066	-8.86***	-0.53	0.061	-8.78***	-0.52	0.060	-8.59***
GH (T1)										0.24	0.087	2.73**	0.03	0.088	0.29	0.06	0.091	0.70
GD (T1)										-0.05	0.081	-0.57	0.09	0.077	1.13	0.09	0.077	1.23
Honesty-Humility (T1)													0.37	0.059	6.23***	0.40	0.062	6.42***
Conscientiousness (T1)																-0.05	0.031	-1.76
Emotionality (T1)																0.02	0.033	0.56
Extraversion (T1)																-0.02	0.026	-0.83
Agreeableness (T1)																0.02	0.033	0.74
Openness (T1)																-0.04	0.033	-1.09
	R^2 =0.005			R^2 =0.069			R^2 =0.329			R^2 =0.336			R^2 = 0.367			R2 = 0.371		
	AIC = 16633			AIC = 13795			AIC = 13670			AIC = 13665			AIC = 13635			AIC = 13639		
	BIC = 16676			BIC = 13902			BIC = 13791			BIC = 13801			BIC = 13778			BIC = 13817		

Note: N = 198 participants (Subsample 1), 9,367 observations (due to missing values).

B, Unstandardized regression estimates; bpc, between-person centered; GD, dishonesty-related goals; GH, honesty-related goals; IVs, Independent variables; SE, Standard Error; wpc, within-person centered.

All T1 predictors are mean-centered.

*p<0.05; **p<0.01; ***p<0.001.

Additional pre-registered analyses

Temporal effects on state honesty

As reported in Table 4 (Step 1), the Honesty-Humility states showed a significant autoregressive effect (b = .19, p <.001), often referred to as inertia (Hamaker & Grasman, 2014). This autoregressive effect provides evidence for a certain stability of the Honesty-Humility state across the measurement occasions within the same day. This model, however, accounted for only 2% of the variance (R^2 = 0.016). As can be seen from Step 2 in Table 4 (R^2 = 0.016), there was a significant lagged effect of both honesty- and dishonesty-related goals on state Honesty-Humility at the subsequent measurement occasion. The direction of these associations was positive for goals of honesty and negative for goals of dishonesty, which was in line with our expectations. To assess the incremental validity of lagged effects, we compared the autoregressive and lagged models against a baseline model that only incorporates the linear effect of time (Step 0). Both models added a limited amount of variance (2%). According to the AIC, model fit improved when both autoregressive effects and lagged effects of goals were included. However, the BIC identified the autoregressive model as having the best fit.

Table 4. *Autoregressive and cross-lagged effects of goals on state honesty*

	Step 0			Step 1			Step 2		
IVs	***B***	**SE**	***t***	***B***	**SE**	**t**	***B***	**SE**	***t***
Intercept	5.84	0.053	110.23***	5.83	0.048	120.55***	5.81	0.053	109.32***
day	-0.01	0.002	-4.91***	-0.01	0.002	-3.15**	-0.00	0.002	-2.50*
State Honesty-Humility (wpc), lag 1				0.19	0.021	9.15***	0.16	0.020	7.69***
GH (wpc), lag 1							0.05	0.016	3.10**
GD (wpc), lag 1							-0.05	0.016	-2.89**
	R^2 = 0.002			R^2 = 0.016			R^2 = 0.016		
	AIC = 11008			AIC= 10498			AIC = 10469		
	BIC = 11035			BIC = 10545			BIC = 10577		

Note: N = 198 (Subsample 1), 6,109 observations (due to missing data).

B, Unstandardized regression estimates; bpc, between-person centered; GD, dishonesty-related goals; GH, honesty-related goals; IVs, Independent variables; SE, Standard Error; wpc, within-person centered.

*p<0.05; **p<0.01; ***p<0.001.

Prediction of lying behavior

A histogram of lying behavior can be found in Supplementary Materials (Figure S1). We first performed an analysis using individual goals as predictors (mirroring the analysis reported in Table S4 in the Supplementary Materials, predicting Honesty-Humility states). As expected, all four goals of dishonesty (Hurt Someone, Manipulate, Lie, Get At Any Cost) were positively associated with lying behavior. Three goals related to honesty, Satisfaction with Self, Happiness, and Tell the Truth, predicted lying behavior negatively. However, four goals of honesty turned out to predict lying positively: Avoid Hurting (B = .09, p = .002), Avoid Deception (B = .13, p < .001), Justice (B = .13, p < .001), and Test Someone (B = .09, p = .005). These relations remained mainly unchanged when estimated in a single mixed model, with the exception of goals Avoid Hurting and Justice, which ceased to be significant predictors of Lying. Full details of this analysis can be found in the Supplementary Materials (Table S13).

Additionally, we examined contemporaneous and temporal effects of the goals of the self-reported lying behavior, while controlling for a potential linear trend of time. The results showed that, at the contemporaneous level (Table 5), only goals related to dishonesty, but not those of honesty, predicted lying at a significant level (Step 1). Goals related to dishonesty were positively associated with the self-reported lies at the WP level, suggesting that participants reported having told more lies when they gave higher importance to goals of dishonesty. The model, however, explained only 2% of the variance ($R^2 = 0.021$). Surprisingly, when the BP parts of goals were added to the model (Step 2), the importance of goals of honesty, and not only those of dishonesty, at the WP level became a significant predictor of lying behavior in contrast to the previous model. The variance explained also increased to 19% ($R^2 = 0.187$). This effect might potentially be explained by a BP component of goals acting as a suppressor variable, which means that its inclusion suppresses the variance of the WP part of honest goals that is irrelevant for predicting lying behavior (Tzelgov & Henik, 1991). As for the BP components themselves, only the dishonesty goals showed significant relationships with lying behavior. This suggests that participants who typically gave more importance to dishonesty-related goals through the EMA phase reported a higher number of lies told at the momentary level as well. Further analyses revealed that the outlined relationships with goals remained stable after controlling for baseline T1 goal measures (Step 3, $R^2 = 0.187$) and the Honesty-Humility trait assessed at T1 (Step 4, $R^2 = 0.206$). Moreover, the importance of the dishonest goals and the trait level assessed at T1 predicted the lying behavior significantly. The baseline importance of dishonest goals was identified as a negative predictor of lies in the EMA phase, however, we attribute this effect to multicollinearity between predictors in the model (Alin, 2010), specifically between baseline dishonesty goals and the BP component of EMA goals of dishonesty. In fact, when the BP part of dishonest goals is taken out from the model, the T1 dishonesty-related goals become a positive predictor of lying behavior as expected.

Table 5. *Hierarchical mixed-effects multiple regression models predicting the lying behavior*

	Step 0			Step 1			Step 2			Step 3			Step 4			Step 5		
IVs	***B***	**SE**	***z***	***B***	**SE**	***z***	***B***	**SE**	***z***	***B***	**SE**	***z***	***B***	**SE**	***z***	***B***	**SE**	***z***
Intercept	-1.60	0.107	-14.94***	-1.64	0.111	-14.77***	-1.72	0.105	-16.36***	-1.72	0.104	-16.51***	-1.71	0.101	-16.93***	-1.73	0.102	-17.02***
day	-0.02	0.011	-1.54	-0.02	0.010	-2.12*	-0.00	0.009	-0.52	-0.00	0.009	-0.53	-0.01	0.009	-0.72	-0.00	0.009	-0.38
GH (wpc)				-0.09	0.064	-1.47	-0.17	0.059	-2.90**	-0.17	0.059	-2.90**	-0.15	0.059	-2.58**	-0.16	0.059	-2.74**
GD (wpc)				0.53	0.065	8.15***	0.48	0.055	8.64***	0.47	0.055	8.68***	0.48	0.054	8.91***	0.48	0.054	8.94***
GH (bpc)							-0.26	0.130	-1.98*	-0.24	0.139	-1.73	-0.18	0.137	-1.33	-0.20	0.136	-1.50
GD (bpc)							1.46	0.171	8.56***	1.64	0.200	8.18***	1.52	0.200	7.61***	1.59	0.197	8.10***
GH (T1)										-0.32	0.253	-1.27	0.05	0.277	0.19	0.03	0.284	0.12
GD (T1)										-0.48	0.237	-2.02*	-0.68	0.242	-2.80**	-0.64	0.238	-2.67**
Honesty-Humility (T1)													-0.58	0.191	-3.01**	-0.49	0.204	-2.39*
Conscientiousness (T1)																-0.24	0.099	-2.46*
Emotionality (T1)																0.04	0.105	0.35
Extraversion (T1)																0.07	0.083	0.85
Agreeableness (T1)																0.18	0.105	1.74
	Step 0			**Step 1**			**Step 2**			**Step 3**			**Step 4**			**Step 5**		
IVs	***B***	**SE**	***z***	***B***	**SE**	***z***	***B***	**SE**	***z***	***B***	**SE**	***z***	***B***	**SE**	***z***	***B***	**SE**	***z***
Openness (T1)																0.05	0.107	0.44
	R^2 =0.001			R^2 =0.021			R^2 =0.187			R^2 =0.197			R^2 = 0.206			R^2 = 0.233		
	AIC = 12841			AIC = 12522			AIC = 12461			AIC = 12461			AIC = 12454			AIC = 12454		
	BIC = 12877			BIC = 12622			BIC = 12576			BIC = 12589			BIC = 12590			BIC = 12626		

Note: N = 198 (Subsample 1), 9,419 observations (due to missing data).

B, Unstandardized regression estimates; bpc, between-person centered; GD, dishonesty-related goals; GH, honesty-related goals; IVs, Independent variables; SE, Standard Error; wpc, within-person centered.

All T1 predictors are mean-centered.

*p<0.05; **p<0.01; ***p<0.001.

Adding the remaining HEXACO-60 dimensions at baseline in Step 5 did not improve the model fit based on AIC and BIC criteria and added only a small amount of variance explained (R^2 = 0.233). Overall, according to the AIC, Step 4 with the Honesty-Humility trait at T1 included and Step 5 with all T1 traits showed the best fit, whereas the BIC indicated a more parsimonious model of Step 2 (with only WP and BP goal components) as having the best fit. In addition, more complex models did not add considerably more variance explained compared to Step 2. Table 6 illustrates the temporal effects of the Honesty-Humility state and goals on lying behavior. The outcomes of the Step 1 model indicated that lies had no significant autoregressive effect, which means that the lying behavior at one measurement occasion does not predict lying behavior at the following time interval. Moreover, adding lagged variables of Honesty-Humility state (Step 2) and goals (Step 3) neither revealed any significant effect (apart from a weak effect of dishonest goals) nor improved the model fit. In general, all temporal models explained less than 1% of the variance.

Table 6. *Autoregressive and cross-lagged effects predicting lying behavior*

	Step 0			Step 1			Step 2			Step 3		
IVs	***B***	**SE**	***z***	***B***	**SE**	***z***	***B***	**SE**	***z***	***B***	**SE**	***z***
Intercept	-1.56	0.111	-14.02***	-1.56	0.112	-13.96***	-1.56	0.112	-13.91***	-1.56	0.112	-13.87***
day	-0.03	0.013	-2.75**	-0.03	0.012	-2.80**	-0.03	0.012	-2.84**	-0.03	0.012	-2.92**
Number of lies, lag 1				-0.00	0.071	-0.08	-0.02	0.069	-0.32	-0.03	0.074	-0.42
Honesty state (wpc), lag 1							-0.07	0.066	-1.08	-0.10	0.081	-1.26
GH (wpc), lag 1										0.09	0.077	1.19
GD (wpc), lag 1										0.09	0.080	1.10
	R^2 = 0.005			R^2 = 0.005			R^2 = 0.006			R^2 = 0.007		
	AIC = 8167			AIC= 8 156			AIC = 8 163			AIC = 8 175		
	BIC = 8 201			BIC = 8 216			BIC = 8 257			BIC = 8 356		

Note: N = 198 (Subsample 1), 6,100 observations (due to missing data).

B, Unstandardized regression estimates; bpc, between-person centered; GD, dishonesty-related goals; GH, honesty-related goals; IVs, Independent variables; SE, Standard Error; wpc, within-person centered.

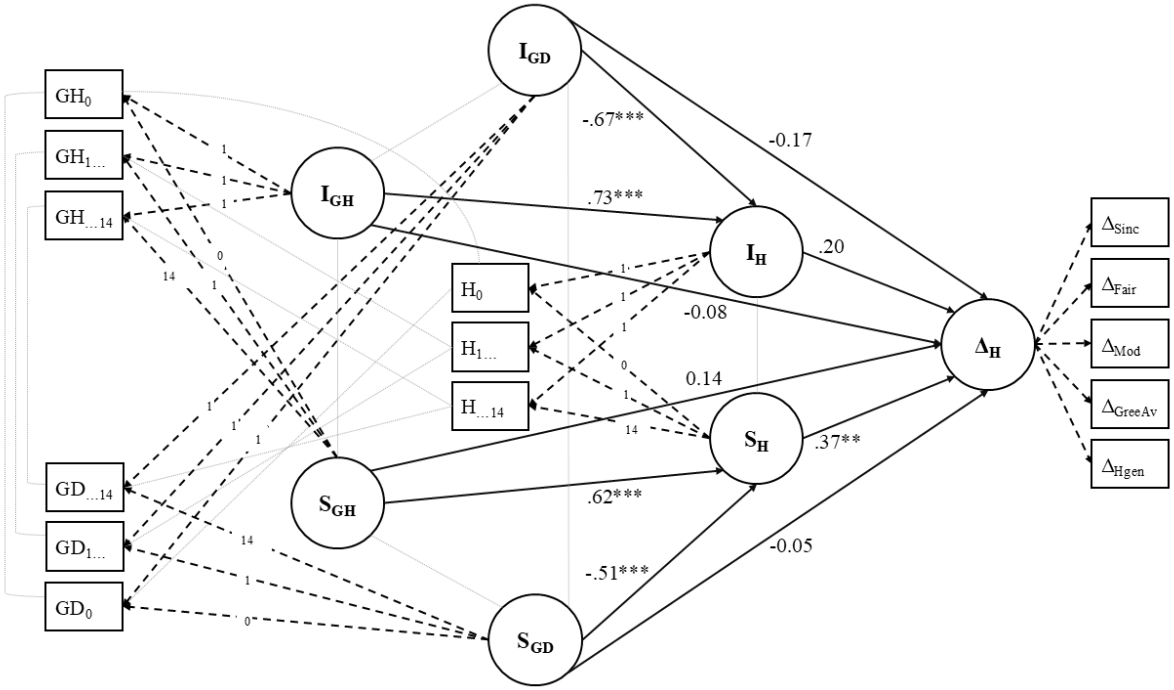
*p<0.05; **p<0.01; ***p<0.001

Latent Growth Process

The examination of mean differences in the HAS Honesty-Humility trait between T1 and T2 revealed no statistically significant changes across the measurement period. This was the case both for the total trait score and for all Honesty-Humility facets individually. Comprehensive distributions of delta scores and mean difference analyses are presented in Figure S2 of the Supplementary Materials. This indicates that aggregate Honesty-Humility levels remained relatively stable across all participants throughout the 15-day investigation period, suggesting the absence of systematic mean-level change at the group level.

Despite the lack of significant mean-level changes, participants demonstrated some degree of intraindividual variability across the various Honesty-Humility dimensions. The standard deviations of T2-T1 difference scores for Honesty-Humility facets were the following: Sincerity (0.50), Fairness (0.60), Modesty (0.64), Greed-Avoidance (0.72), and general Honesty-Humility items (0.54), as detailed in Figure S2. Although there was no significant change in the mean levels of Honesty-Humility between T1 and T2, it was still possible that the levels of the trait increased for some participants, while others' levels decreased (Roberts et al., 2006).To comprehensively model these patterns of individual changes, following Di Sarno et al. (2023), we fit a latent growth model, and it demonstrated acceptable fit: $\chi^2(1235)=2264.823$, $p<0.001$; CFI=0.908; RMSEA=0.066; SRMR=0.094. The comprehensive structural representation of this model is illustrated in Figure 1.

Figure 1. *Latent Growth Model*



Note: N = 191 participants (Subsample with completed T2). Dashed black arrows: factor loadings; Solid black arrows: regression coefficients; Dotted gray lines: covariances; Round nodes: latent factors; Square nodes: observed indicators; I - latent intercept; S - latent slope (growth factor); GH - goals of honesty; GD - goals of dishonesty; H - honesty; ΔSinc/ΔGreeAv - Delta Scores of Honesty-Humility facets (T2 - T1); ΔH - Delta score of the seven general HAS Honesty-Humility items. Residual covariances between H and GH, H and GD, GH and GD were estimated only within days and were constrained to be equal (i.e., constant) across days. Due to some missing data, the model was fit with "full information" maximum likelihood estimation. The figure reports standardized regression coefficients. Loadings and intercepts for the ΔH measurement model and variances and covariances of latent variables and indicators are omitted to improve readability but can be found in Supplementary Materials in Tables S15 – S17.
*p<0.05; **p<0.01; ***p<0.001

The latent growth model elucidated a number of critical dynamic relationships that characterize the complex interplay between goal-oriented behaviors and Honesty-Humility dynamics.

Direct versus indirect pathways.

First, in contrast to conscientiousness (Di Sarno et al., 2023), no direct association between honesty (GH) or dishonesty-related (GD) goals and corresponding changes in trait Honesty-Humility (ΔH) emerged. However, indirect pathways through state-level changes were identified. Specifically, growth in goals of honesty over time during the EMA phase (S_{GH}) was linked to an increase in Honesty-Humility states during EMA (S_H), while growth in goals of dishonesty (S_{GD}) was associated with a decline in EMA Honesty-Humility (S_H). In turn, growth in state Honesty-Humility observed during EMA (S_H) subsequently predicted increases in trait Honesty-Humility (ΔH) from T1 to T2. This pattern suggests that the changes in subjective goal importance (S_{GH}, S_{GD}) were related to the changes in trait level (ΔH) only through changes in the Honesty-Humility state (SH), establishing a sequential mediation pathway.

Testing the indirect effects.

In order to confirm the mediating role of state-level changes, we additionally tested the indirect effects of the latent slopes of goals (S_{GH}, S_{GD}) through changes in the latent slope of Honesty-Humility states (S_H) on T2-T1 trait Honesty-Humility change. As expected, these indirect effects were significantly different from zero for both honesty-related goals ($\beta = 0.23$, $p = 0.008$) and dishonesty-related goals ($\beta = -0.19$, $p = 0.011$).

Initial levels versus growth trajectories.

A different pattern was found for the initial levels (latent intercepts) of EMA goals compared to their growth trajectories. Although initial levels of goal importance (I_{GH}, I_{GD}) were

associated with initial levels of EMA-measured Honesty-Humility (I_H), the latter did not predict the change in trait Honesty-Humility (ΔH). Similarly, there was no indirect effect of goal intercepts (I_{GH}, I_{GD}) on trait change (ΔH) through the initial level of Honesty-Humility states (I_H).

Discussion

This study explored the role of goals related to honesty and dishonesty in the dynamics of Honesty-Humility states through an intensive longitudinal measurement design. Our pre-registered hypothesis was confirmed, demonstrating a clear contemporaneous relationship between honesty-related goals and state Honesty-Humility in daily life. Furthermore, we identified those goals that were associated with lying behavior in everyday situations. Additionally, we investigated the psychometric properties of the GH/GD-EMA questionnaire, which is an important analytical step that is, however, not often done in existing studies (Hall et al., 2021). Finally, exploratory analyses revealed insights into the dynamic processes connecting motives, momentary personality states, and enduring personality traits.

The results of the present study provide significant empirical evidence for a dynamic and processual understanding of honesty, consistent with the principles of WTT (Fleeson & Jayawickreme, 2015) and previous studies that showed that goals substantially accounted for personality state variance (McCabe & Fleeson, 2012, 2016). Our results extend this evidence to the domain of Honesty-Humility, showing that goals serve as predictors of behavioral states for this trait, as well as for lying behavior. In addition, our results demonstrated that goals also predicted future Honesty-Humility states. Overall, our findings support the idea that Honesty-Humility is not a static characteristic, but manifests through dynamic states that fluctuate systematically as a function of the momentary goals pursued by the individual. The results of our study also align with the broader theoretical framework proposed by Costantini and colleagues (Costantini et al., 2020; Costantini & Perugini, 2018), who argued for the

importance of uncovering the motivational core of personality traits to fully understand their processes both at the state and trait levels. The significant associations between honesty-related goals and state Honesty-Humility provide empirical support for the notion that goals serve as crucial regulatory mechanisms in personality expression and change (Buchinger et al., 2024; Hudson & Roberts, 2014; Robinson et al., 2015). The robustness of the associations found, which remain significant even after controlling for baseline levels of trait Honesty-Humility and goal importance at T1, suggests that the observed effects are not simply a reflection of stable differences between individuals but represent active, dynamic processes in daily life. This result is particularly important in light of the debate on the distinction between trait and state assessments (Augustine & Larsen, 2012), highlighting how momentary assessment captures aspects of personality functioning not completely reflected by traditional dispositional measures. The lagged effects observed in our study further extend previous findings by Bleidorn (2009), who demonstrated that personality states are linked to current social roles and major life goals. Our results suggest that this linkage operates not only contemporaneously but also across time, with goals at one moment predicting subsequent personality state expression. This temporal dynamic is crucial for understanding how personality unfolds in the stream of daily experience and how momentary motivational processes can accumulate into longer-term personality patterns.

Particularly interesting is the evidence that dishonesty goals negatively predict both contemporaneous and future Honesty-Humility states and positively predict contemporaneous lying behavior. This pattern suggests that lying is not simply due to the absence of goals directed to honesty, but represents an active process guided by specific dishonest goals. The CFA results that give preference to a two-factor structure and predictive patterns for lying behavior converge to support the broader theoretical framework that goals for dishonesty constitute a distinct motivational system rather than simply representing the absence of honesty

goals — a finding that contradicts what a unidimensional model would suggest (see also Costantini et al., 2020; Di Sarno et al., 2023). While dishonesty goals demonstrated a clear positive relationship with lying behavior, honesty goals exhibited a more nuanced and complex association with deceptive conduct. Specifically, certain honesty-related goals showed positive correlations with lying behavior (Table S13), while simultaneously maintaining positive relationships (though not always significant) with Honesty-Humility (Table S4). These goals include Avoid Hurting Someone, Avoid Deception, Justice, and Test Someone.

The goal Avoid Hurting Someone likely reflects the phenomenon of other-oriented prosocial lies — deceptive communications motivated by benevolent intentions to protect others from emotional distress or negative consequences. This explanation aligns with research demonstrating that individuals may engage in deceptive behavior when motivated by altruistic concerns, highlighting the moral complexity inherent in honest-dishonest behavioral dichotomies (Levine et al., 2016; E. E. Levine & Lupoli, 2022).

The remaining paradoxical motivations — Avoid Deception, Justice, and Test Someone — might represent a previously underexplored category of defensive or protective deception. These goals potentially reflect individuals' attempts to safeguard their own truth-related or justice-oriented rights through the strategic dishonesty of others. For instance, individuals might engage in deceptive behavior to protect themselves from perceived injustices, to maintain their moral integrity in corrupt environments, or to expose others' dishonesty through deceptive means. While this phenomenon has received limited attention in the existing literature, it suggests a sophisticated moral reasoning process whereby individuals may justify deceptive behavior as a means of preserving higher-order ethical principles.

The identification of these paradoxical goal-behavior relationships also has practical implications for interventions designed to promote ethical behavior and prevent antisocial behavior (see e.g., Brooks et al., 2013; Epley & Tannenbaum, 2017; Osafo et al., 2021).

Understanding that some forms of deception may be motivated by fundamentally honest intentions suggests that straightforward approaches to reducing dishonest behavior may be insufficient or even counterproductive. Instead, interventions might benefit from addressing the underlying motivational conflicts that give rise to these complex behavioral patterns.

Another result of the present study concerns evidence for bottom-up processes of personality change. Although the absence of significant group-level changes in the Honesty-Humility trait over a two-week period is consistent with literature on trait stability in the short term (Roberts & DelVecchio, 2000), the presence of inter-individual variability in change suggests that some individuals experience some degree of modifications. The latent growth model revealed that changes in honesty goals during the EMA period predict changes in Honesty-Humility states, which in turn mediate the association with change in the Honesty-Humility trait from T1 to T2. The observed mediation pattern aligns with the theoretical framework by Wrzus and Roberts (2017). Empirical findings revealed differential effects between initial goal and trait levels (intercepts) versus goal and trait changes over time (slopes), providing empirical validation for the proposed dynamic developmental mechanisms over static individual difference explanations. In the case of Honesty-Humility, the effects of the trajectories of goals measured during the EMA phase on trait change Honesty-Humility were mediated by state trajectories during the EMA phase. This pattern partly diverges from Di Sarno et al. (2023) findings on conscientiousness, where both direct and indirect pathways were observed from goal changes to trait changes. Our findings on Honesty-Humility reveal instead a full mediation model, where the effects of goal changes on trait Honesty-Humility operate exclusively through changes in Honesty-Humility states. The absence of direct effects in our study suggests that, for Honesty-Humility, the pathway from motivational processes to trait change is more strongly channeled through experiential states than for conscientiousness.

The results of the present study also contribute to the growing literature on the importance

of ecological momentary assessment, which enables the distinction between WP and BP effects, for understanding personality functioning in daily life (Allemand & Mehl, 2017; Di Sarno et al., 2023). The use of mixed-effects models on disaggregated WP and BP effects (Wang & Maxwell, 2015) revealed interesting patterns in Honesty-Humility dynamics. Significant associations at both levels of analysis suggest that the importance of honesty-related goals operates through both intra-individual mechanisms (momentary variations in goal activation) and inter-individual mechanisms (stable differences in the priority attributed to honesty-related goals).

Evidence for bottom-up personality change processes has important practical implications. If momentary goals are related both to immediate Honesty-Humility states and longer-term trait development, interventions targeting goal activation in daily life could represent a promising approach for promoting ethical behavior (Allemand & Flückiger, 2017). This could be particularly relevant in organizational contexts, where fostering cultures emphasizing honesty-related goals might contribute to more ethical workplace behaviors (Pletzer et al., 2019b). Research on workplace interventions has identified several pathways for enhancing Honesty-Humility in organizational settings. Individual-level nudging approaches, such as visual cues with social norms or prosocial message framing (Ayal et al., 2021; Hillebrandt & Barclay, 2022), were found to activate honesty-related goals in daily work situations, aligning with the goal-state-trait pathway identified in personality development research. Structural modifications of task design that promote deliberative processing lowered the employees' rule-breaking behavior (Derfler-Rozin et al., 2016). Additionally, ethical leadership emerged as critical for facilitating Honesty-Humility development through modeling appropriate conduct and reinforcing ethical decision-making (X. Liu et al., 2020; Mayer et al., 2009). However, our finding that momentary dishonesty goals actively predict both reduced Honesty-Humility states and increased lying behavior suggests that interventions

should not only promote positive ethical goals but also address factors that activate antisocial motivations. This dual approach could be more effective than interventions focusing solely on promoting honesty.

Several limitations of the present study merit consideration. First, using self-report measures for all constructs of interest raises questions regarding social desirability and response bias, particularly relevant in the honesty domains connected to moral behavior, such as Honesty-Humility (see also DeVries et al., 2018), although there is evidence that self-reported honesty measures also reflect substantial valid variance (e.g., Lee et al., 2008; Zettler et al., 2020). Future studies could benefit from integrating objective behavioral measures and external observations. Second, our study gives a first hint into the interplay of goals and momentary state Honesty-Humility in the change process of the trait. However, demonstration of understanding of the causal processes involving goals would require an experimental study, in which the goals could be manipulated (McCabe & Fleeson, 2016; Stieger et al., 2021). Additionally, investigating the situational factors that activate different types of goals could provide deeper insights into the person-situation interactions that shape honest behavior (e.g., Rauthmann, 2016).

Conclusions

Taken together, our results contribute significantly to the theoretical understanding of Honesty-Humility as a dynamic and situationally responsive trait, with both important implications for personality theorizing and potential applications in the domain of behavioral ethics. Integrating WP and BP perspectives provides a more nuanced understanding of how Honesty-Humility operates in daily life, bridging together the trait and state approaches to personality. Through ecological momentary assessment, the research reveals that motivational goals serve as the crucial bridge between situations and honest or dishonest behaviors. The study uncovers moral complexity where some honesty-related goals, such as Avoiding Hurting,

can paradoxically increase deceptive behavior, highlighting the nuanced relationship between moral intentions and actions. The findings provide evidence for bottom-up personality change, where repeated patterns of daily goal-state behaviors contribute to longer-term trait development. Overall, the study not only advances our theoretical understanding of honesty but also offers potential practical pathways for developing interventions that could enhance ethical behavior in real-world contexts.

References

Alin, A. (2010). Multicollinearity. *WIREs Computational Statistics*, *2*(3), 370–374. https://doi.org/10.1002/wics.84

Allemand, M., & Flückiger, C. (2017). Changing personality traits: Some considerations from psychotherapy process-outcome research for intervention efforts on intentional personality change. *Journal of Psychotherapy Integration*, *27*(4), 476–494. https://doi.org/10.1037/int0000094

Allemand, M., & Mehl, M. R. (2017). Personality assessment in daily life: A roadmap for future personality development research. In *Personality Development Across the Lifespan* (pp. 437–454). Academic Press. https://doi.org/10.1016/B978-0-12-804674-6.00001-6

Allport, G. W. (1961). *Pattern and growth in personality.* Holt, Reinhart & Winston.

Amenta, S., Galkina, A., Romano, D., Perugini, M., & Costantini, G. (2025). *A lexical examination of the facets of Honesty-Humility: The Adjective Checklist of Honesty*. OSF. https://doi.org/10.31234/osf.io/yv7un_v1

Ashton, M. C., & Lee, K. (2008). The HEXACO Model of Personality Structure and the Importance of the H Factor. *Social and Personality Psychology Compass*, *2*(5), 1952–1962. https://doi.org/10.1111/j.1751-9004.2008.00134.x

Ashton, M. C., & Lee, K. (2009). The HEXACO–60: A Short Measure of the Major Dimensions of Personality. *Journal of Personality Assessment*. https://www.tandfonline.com/doi/full/10.1080/00223890902935878

Ashton, M. C., & Lee, K. (2020). Objections to the HEXACO Model of Personality Structure—And why those Objections Fail. *European Journal of Personality*, *34*(4), 492–510. https://doi.org/10.1002/per.2242

Ashton, M. C., Lee, K., & De Vries, R. E. (2014). The HEXACO Honesty-Humility, Agreeableness, and Emotionality Factors: A Review of Research and Theory. *Personality and Social Psychology Review*, *18*(2), 139–152. https://doi.org/10.1177/1088868314523838

Ashton, M. C., Lee, K., de Vries, R. E., Perugini, M., Gnisci, A., & Sergi, I. (2006). The HEXACO model of personality structure and indigenous lexical personality dimensions in Italian, Dutch, and English. *Journal of Research in Personality*, *40*(6), 851–875. https://doi.org/10.1016/j.jrp.2005.06.003

Augustine, A. A., & Larsen, R. J. (2012). Is a Trait Really the Mean of States?: Similarities

and Differences Between Traditional and Aggregate Assessments of Personality. *Journal of Individual Differences*, *33*(3), 131–137. https://doi.org/10.1027/1614-0001/a000083

Ayal, S., Celse, J., & Hochman, G. (2021). Crafting messages to fight dishonesty: A field investigation of the effects of social norms and watching eye cues on fare evasion. *Organizational Behavior and Human Decision Processes*, *166*, 9–19. https://doi.org/10.1016/j.obhdp.2019.10.003

Bartoń, K. (2025). *MuMIn: Multi-Model Inference. R package version 1.48.11* [Computer software]. https://CRAN.R-project.org/package=MuMIn.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Baumert, A., Schmitt, M., Perugini, M., Johnson, W., Blum, G., Borkenau, P., Costantini, G., Denissen, J. J. A., Fleeson, W., Grafton, B., Jayawickreme, E., Kurzius, E., MacLeod, C., Miller, L. C., Read, S. J., Roberts, B., Robinson, M. D., Wood, D., & Wrzus, C. (2017). Integrating Personality Structure, Personality Process, and Personality Development. *European Journal of Personality*, *31*(5), 503–528. https://doi.org/10.1002/per.2115

Bleidorn, W. (2009). Linking personality states, current social roles and major life goals. *European Journal of Personality*, *23*(6), 509–530. https://doi.org/10.1002/per.731

Bleidorn, W., Kandler, C., Hülsheger, U. R., Riemann, R., Angleitner, A., & Spinath, F. M. (2010). Nature and nurture of the interplay between personality traits and major life goals. *Journal of Personality and Social Psychology*, *99*(2), 366–379. https://doi.org/10.1037/a0019982

Bringmann, L. F., Pe, M. L., Vissers, N., Ceulemans, E., Borsboom, D., Vanpaemel, W., Tuerlinckx, F., & Kuppens, P. (2016). Assessing Temporal Emotion Dynamics Using Networks. *Assessment*, *23*(4), 425–435. https://doi.org/10.1177/1073191116645909

Brooks, J., Narvaez, D., & Bock, T. (2013). MORAL MOTIVATION, MORAL JUDGMENT, AND ANTISOCIAL BEHAVIOR. *Moral Motivation*, *9*(2).

Brown, T. A. (2015). Confirmatory factor analysis for applied research. Guilford Press.

Buchinger, L., Entringer, T. M., Richter, D., Wagner, G. G., Gerstorf, D., & Bleidorn, W. (2024). Codevelopment of life goals and the Big Five personality traits across adulthood and old age. *Journal of Personality and Social Psychology*, *126*(2), 346–368. https://doi.org/10.1037/pspp0000477

Caspi, A., & Gorsky, P. (2006). Online Deception: Prevalence, Motivation, and Emotion. *CyberPsychology & Behavior*, *9*(1), 54–59. https://doi.org/10.1089/cpb.2006.9.54

Choshen-Hillel, S., Shaw, A., & Caruso, E. M. (2020). Lying to appear honest. *Journal of Experimental Psychology: General*, *149*(9), 1719–1735. https://doi.org/10.1037/xge0000737

Conner, T. S., Tennen, H., Fleeson, W., & Barrett, L. F. (2009). Experience Sampling Methods: A Modern Idiographic Approach to Personality Research. *Social and Personality Psychology Compass*, *3*(3), 292–313. https://doi.org/10.1111/j.1751-9004.2009.00170.x

Cooper, B., Cohen, T. R., Huppert, E., Levine, E. E., & Fleeson, W. (2023). Honest Behavior: Truth-Seeking, Belief-Speaking, and Fostering Understanding of the Truth in Others. *Academy of Management Annals*, *17*(2), 655–683. https://doi.org/10.5465/annals.2021.0209

Costa, P. T., & McCrae, R. R. (1992). The Five-Factor Model of Personality and Its Relevance to Personality Disorders. *Journal of Personality Disorders*, *6*(4), 343–359. https://doi.org/10.1521/pedi.1992.6.4.343

Costantini, G., & Perugini, M. (2018). A framework for testing causality in personality research. *European Journal of Personality*, *32*(3), 254–268. https://doi.org/10.1002/per.2150

Costantini, G., Saraulli, D., & Perugini, M. (2020). Uncovering the Motivational Core of Traits: The Case of Conscientiousness. *European Journal of Personality*, *34*(6), 1073–1094. https://doi.org/10.1002/per.2237

Cramer, A. O. J., Van Der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S. H., Kendler, K. S., & Borsboom, D. (2012). Dimensions of Normal Personality as Networks in Search of Equilibrium: You Can't like Parties if you Don't like People. *European Journal of Personality*, *26*(4), 414–431. https://doi.org/10.1002/per.1866

Curran, P. J., West, S. G., & Finch, J. F. (1996). The Robustness of Test Statistics to Nonnormality and Specification Error in Confirmatory Factor Analysis.

Daumiller, M., & Janke, S. (2020). Effects of performance goals and social norms on academic dishonesty in a test. *British Journal of Educational Psychology*, *90*(2), 537–559. https://doi.org/10.1111/bjep.12310

De Cremer, D., & Moore, C. (2020). Toward a Better Understanding of Behavioral Ethics in the Workplace. *Annual Review of Organizational Psychology and Organizational Behavior*, *7*(1), 369–393. https://doi.org/10.1146/annurev-orgpsych-012218-015151

DePaulo, B. M., Ansfield, M. E., Kirkendol, S. E., & Boden, J. M. (2004). Serious Lies. *BASIC AND APPLIED SOCIAL PSYCHOLOGY*, *26*(2 & 3), 147–167.

DePaulo, B. M., Kashy, D. A., Kirkendol, S. E., & Wyer, M. M. (1996). Lying in Everyday Life. *Journal of Personality and Social Psychology*, *70*(5), 979–995.

Derfler-Rozin, R., Moore, C., & Staats, B. R. (2016). Reducing Organizational Rule Breaking Through Task Variety: How Task Design Supports Deliberative Thinking. *Organization Science*, *27*(6), 1361–1379. https://doi.org/10.1287/orsc.2016.1094

DeVries, R. E., Hilbig, B. E., Zettler, I., Dunlop, P. D., Holtrop, D., Lee, K., & Ashton, M. C. (2018). *Still misinterpreting lie scales: Reply to Feldman's rejoinder*. https://research.vu.nl/en/publications/still-misinterpreting-lie-scales-reply-to-feldmans-rejoinder

Di Sarno, M., Costantini, G., Richetin, J., Preti, E., & Perugini, M. (2023). Why are you (un)conscientious? The dynamic interplay of goals, states, and traits in everyday life. *Journal of Personality*, *91*(4), 977–991. https://doi.org/10.1111/jopy.12701

Di Sarno, M., Zimmermann, J., Madeddu, F., Casini, E., & Di Pierro, R. (2020). Shame behind the corner? A daily diary investigation of pathological narcissism. *Journal of Research in Personality*, *85*, 103924. https://doi.org/10.1016/j.jrp.2020.103924

Eisele, G., Kasanova, Z., & Houben, M. (2022). Questionnaire Design and Evaluation. In *I. Myin-Germeys & P. Kuppens (Eds.), The open handbook of Experience Sampling Methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies* (2nd ed., pp. 71–90). Center for Research on Experience Sampling and Ambulatory Methods Leuven.

Epley, N., & Tannenbaum, D. (2017). Treating Ethics as a Design Problem. *Behavioral Science*, *3*(2).

Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*(6), 1011–1027. https://doi.org/10.1037/0022-3514.80.6.1011

Fleeson, W. (2004). Moving Personality Beyond the Person-Situation Debate: The Challenge and the Opportunity of Within-Person Variability. *Current Directions in Psychological Science*, *13*(2), 83–87. https://doi.org/10.1111/j.0963-7214.2004.00280.x

Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology*, *97*(6), 1097–1114. https://doi.org/10.1037/a0016786

Fleeson, W., & Jayawickreme, E. (2025). Getting from states to traits: Whole Trait Theory's explanatory and developmental engine. *European Journal of Personality*, 08902070251366709. https://doi.org/10.1177/08902070251366709

Fleeson, W., & Jayawickreme, E. (2015). Whole Trait Theory. *Journal of Research in Personality*, *56*, 82–92. https://doi.org/10.1016/j.jrp.2014.10.009

Fleeson, W., Malanos, A. B., & Achille, N. M. (2002). An intraindividual process approach to the relationship between extraversion and positive affect: Is acting extraverted as "good" as being extraverted? *Journal of Personality and Social Psychology*, *83*(6), 1409–1422. https://doi.org/10.1037/0022-3514.83.6.1409

Fleeson, W., & Noftle, E. E. (2012). Personality Research. In *Handbook of Research Methods for Studying Daily Life / edited by Matthias R. Mehl, Tamlin S. Conner*. The Guilford Press.

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72–91. https://doi.org/10.1037/a0032138

Goldberg, L. R. (2002). An Alternative "Description of Personality": The Big-Five Factor Structure. In *Personality and Personality Disorders*. Routledge.

Hall, M., Scherner, P. V., Kreidel, Y., & Rubel, J. A. (2021). A Systematic Review of Momentary Assessment Designs for Mood and Anxiety Symptoms. *Frontiers in Psychology*, *12*, 642044. https://doi.org/10.3389/fpsyg.2021.642044

Hamaker, E. L., & Grasman, R. P. P. P. (2014). To center or not to center? Investigating inertia with a multilevel autoregressive model. *Frontiers in Psychology*, *5*, 1492. https://doi.org/10.3389/fpsyg.2014.01492

Heck, D. W., Thielmann, I., Moshagen, M., & Hilbig, B. E. (2018). Who lies? A large-scale reanalysis linking basic personality traits to unethical decision making. *Judgment and Decision Making*, *13*(4), 356–371. https://doi.org/10.1017/S1930297500009232

Heller, D., Komar, J., & Lee, W. B. (2007). The Dynamics of Personality States, Goals, and Well-Being. *Personality and Social Psychology Bulletin*, *33*(6), 898–910. https://doi.org/10.1177/0146167207301010

Hillebrandt, A., & Barclay, L. J. (2022). How COVID-19 can promote workplace cheating behavior via employee anxiety and self-interest – And how prosocial messages may overcome this effect. *Journal of Organizational Behavior*. https://doi.org/10.1002/job.2612

Hudson, N. W., & Fraley, R. C. (2015). Volitional personality trait change: Can people choose

to change their personality traits? *Journal of Personality and Social Psychology*, *109*(3), 490–507. https://doi.org/10.1037/pspp0000021

Hudson, N. W., & Roberts, B. W. (2014). Goals to change personality traits: Concurrent links between personality traits, daily behavior, and goals to change oneself. *Journal of Research in Personality*, *53*, 68–83. https://doi.org/10.1016/j.jrp.2014.08.008

Janke, S., Daumiller, M., & Rudert, S. C. (2019). Dark Pathways to Achievement in Science: Researchers' Achievement Goals Predict Engagement in Questionable Research Practices. *Social Psychological and Personality Science*, *10*(6), 783–791. https://doi.org/10.1177/1948550618790227

John, O. P., & Chaplin, W. F. (1988). *Conceptions of States and Traits: Dimensional Attributes With Ideals as Prototypes*.

Johnson, P. C. D. (2014). Extension of Nakagawa & Schielzeth's R^2_{GLMM} to random slopes models. *Methods in Ecology and Evolution*, *5*(9), 944–946. https://doi.org/10.1111/2041-210x.12225

Jordan, A. E. (2001). College Student Cheating: The Role of Motivation, Perceived Norms, Attitudes, and Knowledge of Institutional Policy. *Ethics & Behavior*, *11*(3), 233–247. https://doi.org/10.1207/S15327019EB1103_3

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2012). *semTools: Useful Tools for Structural Equation Modeling* (p. 0.5-7) [Dataset]. https://doi.org/10.32614/CRAN.package.semTools

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lee, K., & Ashton, M. C. (2008). The HEXACO Personality Factors in the Indigenous Personality Lexicons of English and 11 Other Languages. *Journal of Personality*, *76*(5), 1001–1054. https://doi.org/10.1111/j.1467-6494.2008.00512.x

Levine, Ali, M. V., Dean, M., Abdulla, R. A., & Garcia-Ruano, K. (2016). Toward a Pan-cultural Typology of Deception Motives. *Journal of Intercultural Communication Research*, *45*(1), 1–12. https://doi.org/10.1080/17475759.2015.1137079

Levine, E. E., & Lupoli, M. J. (2022). Prosocial lies: Causes and consequences. *Current Opinion in Psychology*, *43*, 335–340. https://doi.org/10.1016/j.copsyc.2021.08.006

Liu, X., Liao, H., Derfler-Rozin, R., Zheng, X., Wee, E. X. M., & Qiu, F. (2020). In line and out of the box: How ethical leaders help offset the negative effect of morality on creativity. *Journal of Applied Psychology*, *105*(12), 1447–1465.

https://doi.org/10.1037/apl0000489

Lüdecke, D., Makowski, D., Ben-Shachar, M. S., Patil, I., Waggoner, F., Wiernik, B. M., & Thériault, R. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Lüdtke, O., Trautwein, U., & Husemann, N. (2009). Goal and Personality Trait Development in a Transitional Period: Assessing Change and Stability in Personality Development. *Personality and Social Psychology Bulletin*, *35*(4), 428–441. https://doi.org/10.1177/0146167208329215

Marsden, H., Carroll, M., & Neill, J. T. (2005). Who cheats at university? A self-report study of dishonest academic behaviours in a sample of Australian university students. *Australian Journal of Psychology*, *57*(1), 1–10.

Mayer, D. M., Kuenzi, M., Greenbaum, R., Bardes, M., & Salvador, R. (Bombie). (2009). How low does ethical leadership flow? Test of a trickle-down model. *Organizational Behavior and Human Decision Processes*, *108*(1), 1–13. https://doi.org/10.1016/j.obhdp.2008.04.002

McArdle, J. J., & Nesselroade, J. R. (2003). Growth curve analysis in contemporary psychological research. I. B. Weiner (Ed.). In *Handbook of psychology* (pp. 447--480). Hoboken. https://doi.org/10.1002/0471264385. wei0218

McArthur, J., McPhee, P., & Ternes, M. (2024). Examining the acceptability and likelihood of self-centered and other-centered lying using the HEXACO. *Personality and Individual Differences*, *220*, 112526. https://doi.org/10.1016/j.paid.2023.112526

McCabe, K. O., & Fleeson, W. (2012). What Is Extraversion For? Integrating Trait and Motivational Perspectives and Identifying the Purpose of Extraversion. *Psychological Science*, *23*(12), 1498–1505. https://doi.org/10.1177/0956797612444904

McCabe, K. O., & Fleeson, W. (2016). Are traits useful? Explaining trait manifestations as tools in the pursuit of goals. *Journal of Personality and Social Psychology*, *110*(2), 287–301. https://doi.org/10.1037/a0039490

McCrae, R. R., & Costa, P. T. (1987). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology*, *52*(1), 81–90. https://doi.org/10.1037/0022-3514.52.1.81

McCrae, R. R., & Sutin, A. R. (2018). A Five–Factor Theory Perspective on Causal Analysis. *European Journal of Personality*, *32*(3), 151–166. https://doi.org/10.1002/per.2134

Miller, C. B. (2020). Motivation and the Virtue of Honesty: Some Conceptual Requirements

and Empirical Results. *Ethical Theory and Moral Practice*, *23*(2), 355–371. https://doi.org/10.1007/s10677-019-10055-1

Miller, C. B., Fleeson, W., Cohen, T. R., Furr, R. M., & Jayawickreme, E. (2021). *Honesty: Big questions, philosophical issues, and empirical research (white paper for the Honesty Project grant)*.

Mischel, W. (2004). Toward an Integrative Science of the Person. *Annual Review of Psychology*, *55*(1), 1–22. https://doi.org/10.1146/annurev.psych.55.042902.130709

Mõttus, R., Wood, D., Condon, D. M., Back, M. D., Baumert, A., Costantini, G., Epskamp, S., Greiff, S., Johnson, W., Lukaszewski, A., Murray, A., Revelle, W., Wright, A. G. C., Yarkoni, T., Ziegler, M., & Zimmermann, J. (2020). Descriptive, Predictive and Explanatory Personality Research: Different Goals, Different Approaches, but a Shared Need to Move Beyond the Big Few Traits. *European Journal of Personality*, *34*(6), 1175–1201. https://doi.org/10.1002/per.2311

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*, 133–142. https://doi.org/doi: 10.1111/j.2041-210x.2012.00261.x

Osafo, E., Paros, A., & Yawson, R. M. (2021). Valence–Instrumentality–Expectancy Model of Motivation as an Alternative Model for Examining Ethical Leadership Behaviors. *Sage Open*, *11*(2), 21582440211021896. https://doi.org/10.1177/21582440211021896

Pletzer, J. L., Bentvelzen, M., Oostrom, J. K., & de Vries, R. E. (2019). A meta-analysis of the relations between personality and workplace deviance: Big Five versus HEXACO. *Journal of Vocational Behavior*, *112*, 369–383. https://doi.org/10.1016/j.jvb.2019.04.004

Quirin, M., Robinson, M. D., Rauthmann, J. F., Kuhl, J., Read, S. J., Tops, M., & DeYoung, C. G. (2020). The Dynamics of Personality Approach (DPA): 20 Tenets for Uncovering the Causal Mechanisms of Personality. *European Journal of Personality*, *34*(6), 947–968. https://doi.org/10.1002/per.2295

R Core Team. (2022). *R: A Language and Environment for Statistical Computing.* [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Rauthmann, J. F. (2016). Motivational Factors in the Perception of Psychological Situation Characteristics. *Social and Personality Psychology Compass*, *10*(2), 92–108. https://doi.org/10.1111/spc3.12239

Reisz, Z., Boudreaux, M. J., & Ozer, D. J. (2013). Personality traits and the prediction of personal goals. *Personality and Individual Differences*, *55*(6), 699–704.

https://doi.org/10.1016/j.paid.2013.05.023

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*(1), 3–25. https://doi.org/10.1037/0033-2909.126.1.3

Roberts, B. W., Walton, K. E., & Viechtbauer, W. (2006). Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. Psychological Bulletin, 132(1), 1–25. https://doi.org/10.1037/0033-2909.132.1.1

Robinson, O. C., Noftle, E. E., Guo, J., Asadi, S., & Zhang, X. (2015). Goals and plans for Big Five personality trait change in young adults. *Journal of Research in Personality*, *59*, 31–43. https://doi.org/10.1016/j.jrp.2015.08.002

Romano, D., Costantini, G., Richetin, J., & Perugini, M. (2023). The HEXACO Adjective Scales and Its Psychometric Properties. *Assessment*, *30*(8), 2510–2532. https://doi.org/10.1177/10731911231153833

Rosseel, Y. (2012). lavaan: An *R* Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Ścigała, K. A., Schild, C., & Zettler, I. (2020). Dishonesty as a signal of trustworthiness: Honesty-Humility and trustworthy dishonesty. *Royal Society Open Science*, *7*(10), 200685. https://doi.org/10.1098/rsos.200685

Serota, K. B., & Levine, T. R. (2015). A Few Prolific Liars: Variation in the Prevalence of Lying. *Journal of Language and Social Psychology*, *34*(2), 138–157. https://doi.org/10.1177/0261927X14528804

Serota, K. B., Levine, T. R., & Docan-Morgan, T. (2022). Unpacking variation in lie prevalence: Prolific liars, bad lie days, or both? *Communication Monographs*, *89*(3), 307–331. https://doi.org/10.1080/03637751.2021.1985153

Stieger, M., Allemand, M., Roberts, B. W., & Davis, J. P. (2022). Mindful of personality trait change: Are treatment effects on personality trait change ephemeral and attributable to changes in states? *Journal of Personality*, *90*(3), 375–392. https://doi.org/10.1111/jopy.12672

Stieger, M., Flückiger, C., Rüegger, D., Kowatsch, T., Roberts, B. W., & Allemand, M. (2021). Changing personality traits with the help of a digital personality change intervention. *Proceedings of the National Academy of Sciences*, *118*(8), e2017548118. https://doi.org/10.1073/pnas.2017548118

Tett, R. P., Toich, M. J., & Ozkum, S. B. (2021). Trait Activation Theory: A Review of the Literature and Applications to Five Lines of Personality Dynamics Research. *Annual*

Review of Organizational Psychology and Organizational Behavior, *8*, 199–233. https://doi.org/10.1146/annurev-orgpsych-012420062228

Thai, S., & Page-Gould, E. (2018). ExperienceSampler: An open-source scaffold for building smartphone apps for experience sampling. *Psychological Methods*, *23*(4), 729–739. https://doi.org/10.1037/met0000151

Thielmann, I., Hilbig, B. E., Klein, S. A., Seidl, A., & Heck, D. W. (2024). Cheating to benefit others? On the relation between Honesty-Humility and prosocial lies. *Journal of Personality*, *92*(3), 870–882. https://doi.org/10.1111/jopy.12835

Tzelgov, J., & Henik, A. (1991). Suppression Situations in Psychological Research: Definitions, Implications, and Applications. *Psychological Bulletin*, *109*(3), 524–536. https://doi.org/10.1037/0096-1523.17.3.792

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities.* John Wiley & Sons.

Wang, L. (Peggy), & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, *20*(1), 63–83. https://doi.org/10.1037/met0000030

Wilson, A. T. (2018). Honesty as a Virtue. *Metaphilosophy*, *49*(3), 262–280. https://doi.org/10.1111/meta.12303

Wilson, R. E., Thompson, R. J., & Vazire, S. (2017). Are fluctuations in personality states more than fluctuations in affect? *Journal of Research in Personality*, *69*, 110–123. https://doi.org/10.1016/j.jrp.2016.06.006

Wrzus, C., & Roberts, B. W. (2017). Processes of personality development in adulthood: The TESSERA framework. *Personality and Social Psychology Review*, *21*(3), 253–277. https://doi.org/10.1177/1088868316652279

Wu, K. D., & Clark, L. A. (2003). Relations between personality traits and self-reports of daily behavior. *Journal of Research in Personality*, *37*(4), 231–256. https://doi.org/10.1016/S0092-6566(02)00539-

Supplementary Materials

Figure S1. *Mean number of lies at each observation by person*

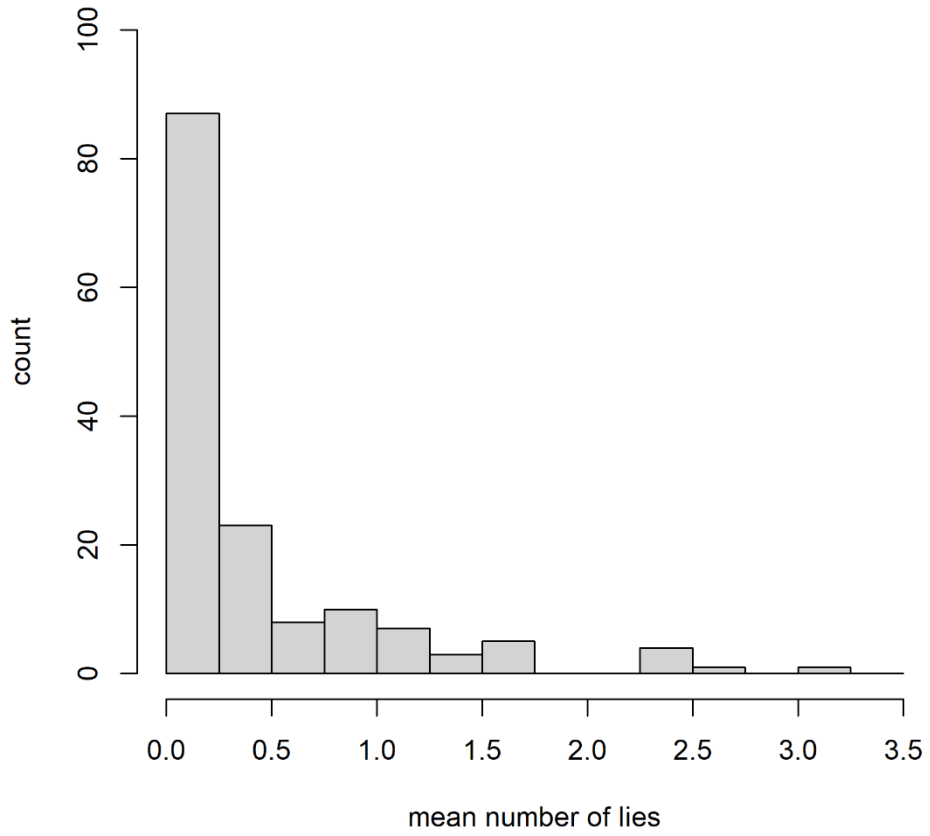
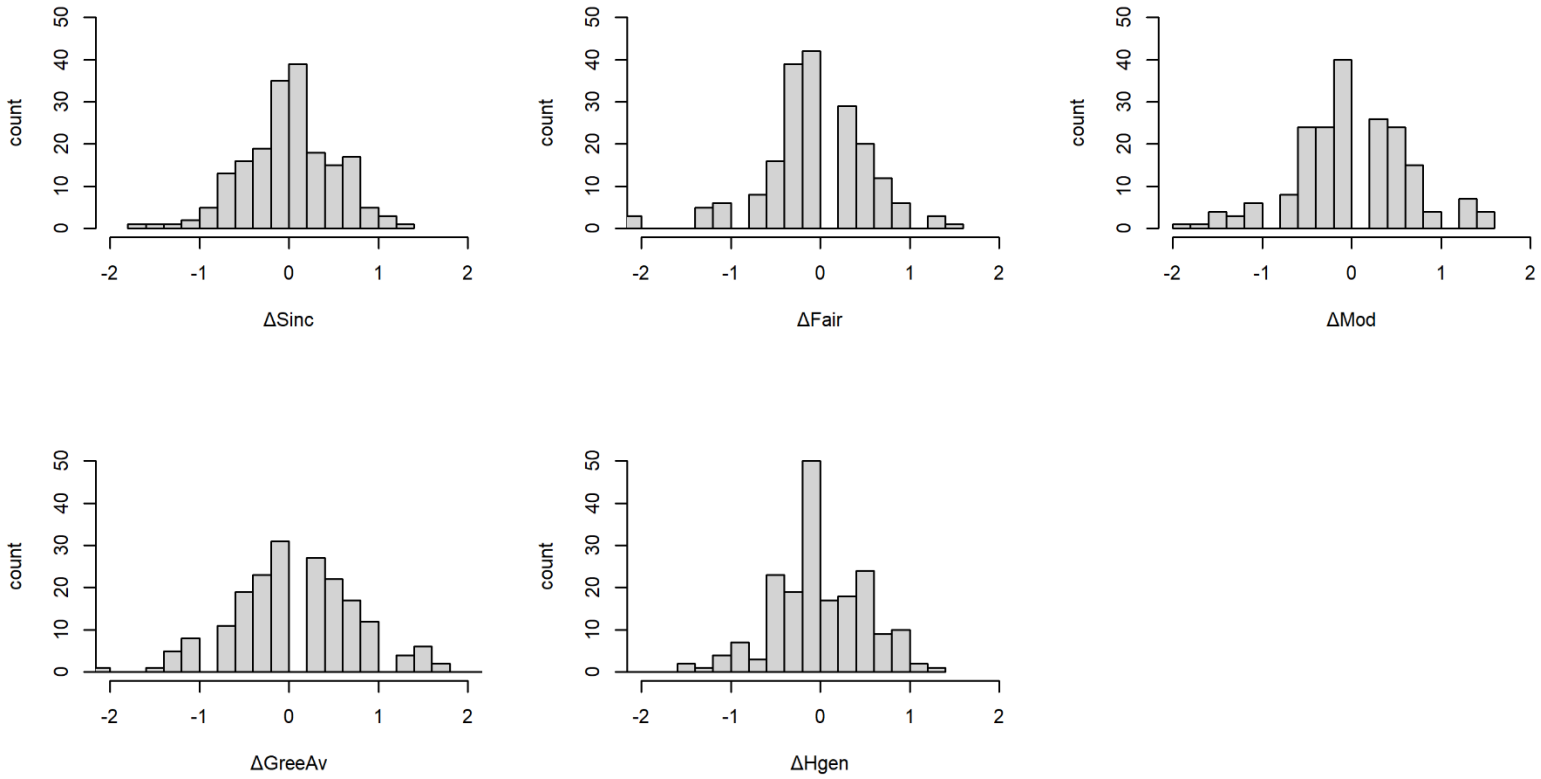


Figure S2. *Delta scores of trait HAS + ACH Honesty-Humility.*



Note: N = 191; For mean differences between T1 and T2 see Table S14.

Figure S3. *Structures at the WP and BP levels with standardized loadings and factor correlations. GD - goals of dishonesty, GH - goals of honesty. All coefficients are significant at p < 0.001.*

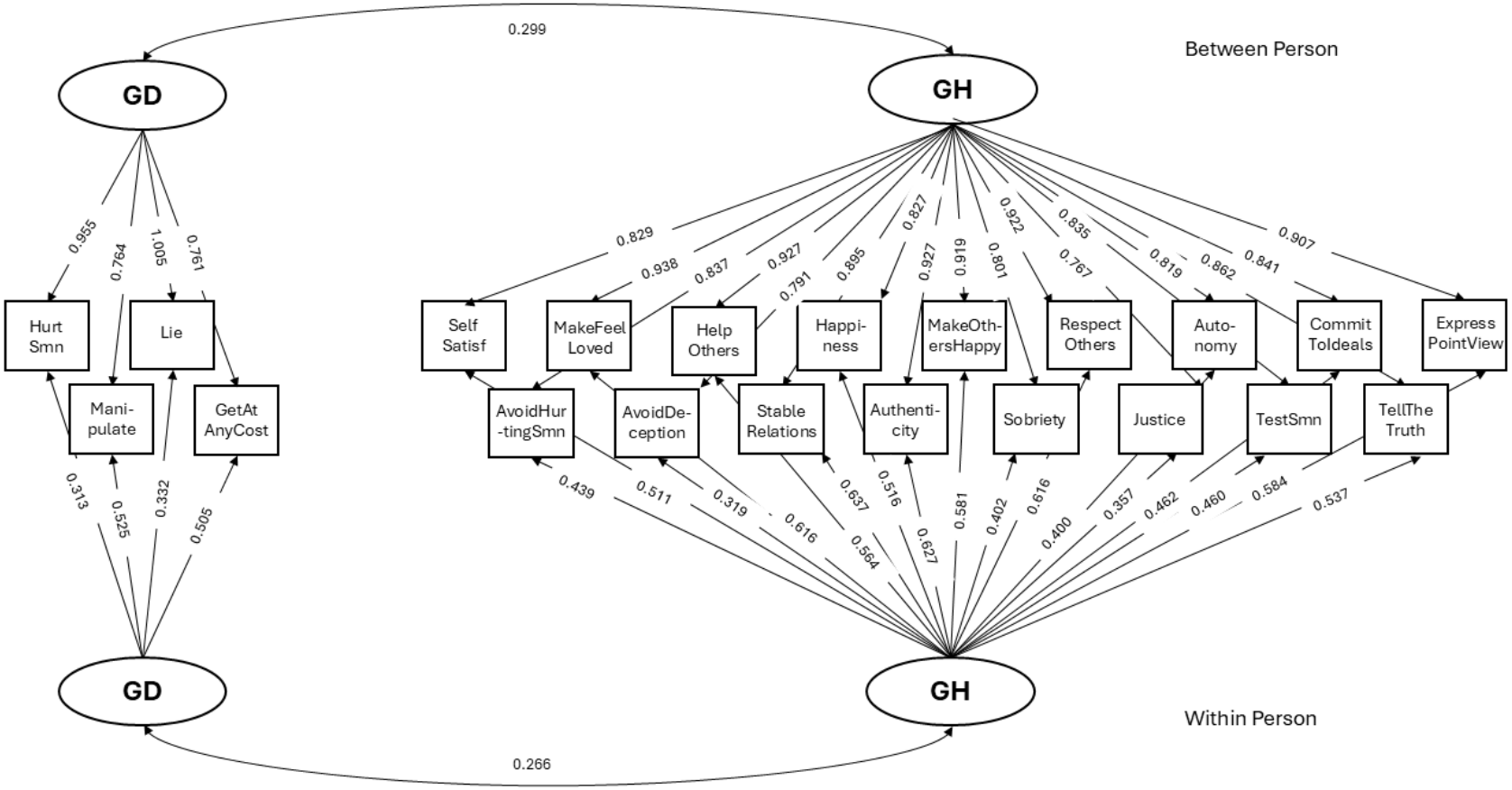


Table S1. *Descriptive statistics of goal items.*

Item	Mean	SD Person-level means	Skewness Person-level means	Kurtosis Person-level means	SD Deviations from person-level means	Skewness Deviations from person-level means	Kurtosis Deviations from person-level means	ICC
Satisfaction with Self	3.60	0.83	-0.59	3.03	0.96	-0.51	3.96	0.413
Make Others Feel Loved	3.30	0.85	-0.24	2.46	1.05	-0.14	3.13	0.380
Help Others	3.23	0.90	-0.34	2.48	1.09	-0.02	3.03	0.388
Happiness	3.67	0.80	-0.49	3.04	0.93	-0.53	4.09	0.411
Hurt Someone	1.43	0.53	1.98	7.68	0.78	1.88	8.45	0.296
Make Others Happy	3.29	0.86	-0.36	2.75	1.07	-0.12	3.16	0.382
Respect Others	3.57	0.82	-0.72	3.34	1.03	-0.46	3.60	0.374
Autonomy	3.66	0.85	-0.57	2.98	0.99	-0.47	3.89	0.410
Commit To Ideals	3.23	0.94	-0.24	2.24	1.01	-0.05	3.41	0.453
Express My Point of View	3.33	0.84	-0.32	2.63	1.07	-0.22	3.21	0.369
Avoid Hurting	2.81	0.99	0.07	2.00	1.11	0.16	3.50	0.430
Avoid Deception	2.45	1.09	0.50	2.25	0.99	0.53	4..24	0.535
Stable Relations	3.48	0.80	-0.45	2.95	1.04	-0.33	3.34	0.354
Manipulate Others	1.84	0.74	1.17	4.43	0.91	1.14	5.06	0.384

Authenticity	3.33	0.87	-0.28	2.51	1.01	-0.20	3.56	0.410
Sobriety	2.65	1.00	0.23	2.22	1.07	0.18	3.78	0.454
Justice	2.46	1.10	0.45	2.18	0.94	0.58	4.44	0.565
Lie	1.43	0.58	1.84	6.10	0.76	1.79	8.63	0.346
Test Someone	2.75	1.05	0.25	2.10	0.99	0.39	3.82	0.514
Get What You Want At Any Cost	1.96	0.78	0.90	3.59	0.98	0.86	4.34	0.371
Tell The Truth	3.13	0.97	-0.11	2.17	1.03	-0.00	3.64	0.457

Note. Participants rated the extent to which each goal was important to them, on a scale from 1 (*not at all*) to 5 (*extremely*). ICC - Intraclass correlation.

Table S2. *Model fit and comparisons*

Model	CFI	RMSEA	SRMR	df	AIC	BIC	χ^2	*p*	Diff. χ^2	Df diff.	***p*** diff.
Within Person level (N = 9 353 observations)											
Model 1: Correlated GH and GD	0.867	0.058	0.051	188	531975.167	534082.486	6187.483	<0.001			
Model 2: Single factor	0.838	0.064	0.056	189	533274.787	535374.962	7489.1	<0.001	1301.6	1	<0.001
Model 3: Orthogonal GH and GD	0.861	0.060	0.060	189	532235.925	534336.100	6450.24	<0.001	262.76	1	<0.001
Between Person level (N = 198 persons)											
Model 1: Correlated GH and GD	0.961	0.032	0.117	188	527733.275	529840.593	1945.590	<0.001			
Model 2: Single factor	0.948	0.036	0.144	189	528312.806	530412.981	2527.122	<0.001	581.53	1	<0.001
Model 3: Orthogonal GH and GD	0.961	0.032	0.195	189	527748.144	529848.319	1962.46	<0.001	16.869	1	<0.001

Note: Model comparisons were conducted with Model 1 of the respective level. CFI - Comparative Fit Index; GD - goals of dishonesty; GH - goals of honesty; RMSEA - Root Mean Square Error of Approximation. RMSEA (adjusted) was calculated based on the formula in (Ryu & West, 2009).

Preliminary analyses – the Scales

Table S3 displays descriptive statistics of the T1 and T2 variables of interest. In the EMA phase (9,443 observations), the mean level of the 198-person means of state Honesty-Humility was 5.76 (SD = 0.73, 4.09 - 6.98). The mean score of person means of honesty- and dishonesty-related goals were 3.17 (SD = 0.78, 1.07 - 5) and 1.67 (SD = 0.6, 1- 4.05), respectively.

As shown in Table S4, the aggregated person-means across the EMA phase of all but two goals, Avoid Deception (r = .10, p = .16) and Justice (r = .08, p = .28), were significantly correlated with state Honesty-Humility. One goal, Test Someone (r = .18, p = .01), showed a weak correlation with state Honesty-Humility. All goals, except for Get What You Want At Any Cost (B = -.02, p = .12), predicted state Honesty-Humility in individual multilevel models in the anticipated directions. When estimated in a single multilevel model, 15 goals out of 21 demonstrated significant unique associations with state Honesty-Humility, while six goals did not. The correlation matrix displaying aggregated goal scores and all other variables can be found in Tables S10-S12.

Table S3. *Descriptive statistics of T1 and T2 variables, including goals related to honesty and dishonesty*

Variable	T1 M(SD)	T1 Range	T2 M(SD)	T2 Range
HEXACO Honesty-Humility	3.87 (0.53)	2.34 - 4.94	3.84 (0.58)	2.16 - 4.94
HAS Honesty-Humility	5.71 (0.59)	3.69 - 6.97	5.74 (0.60)	3.38 - 6.90
HAS Emotionality	4.49 (0.84)	1.90 - 6.40	4.37 (0.90)	2.10 - 6.30
HAS Extraversion	4.34 (1.13)	1.50 - 7.00	4.34 (1.08)	1.10 - 6.90
HAS Agreeableness	5.43 (0.86)	1.70 - 7.00	5.51 (0.84)	2.80 - 7.00
HAS Conscientiousness	5.01 (0.93)	2.50 - 7.00	5.03 (0.95)	2.80 - 7.00
HAS Openness	4.88 (0.82)	2.50 - 6.80	4.93 (0.84)	2.70 - 6.90
GD	1.57 (0.45)	1.00 - 3.48	1.59 (0.53)	1.00 - 3.96
GH	4.22 (0.38)	2.95 - 4.98	4.09 (0.47)	2.11 - 4.95
Satisfaction With Self (GH)	4.54 (0.47)	2.29 - 5.00	4.39 (0.58)	1.86 - 5.00
Make Others Feel Loved (GH)	4.34 (0.54)	2.00 - 5.00	4.24 (0.64)	1.00 - 5.00
Help Others (GH)	4.28 (0.59)	2.60 - 5.00	4.17 (0.64)	1.80 - 5.00
Happiness (GH)	4.60 (0.39)	3.00 - 5.00	4.45 (0.49)	2.25 - 5.00

Make Others Happy (GH)	4.38 (0.60)	2.33 - 5.00	4.25 (0.66)	2.00 - 5.00
Respect Others (GH)	4.34 (0.59)	2.33 - 5.00	4.30 (0.62)	2.33 - 5.00
Autonomy (GH)	4.37 (0.60)	2.33 - 5.00	4.31 (0.68)	2.00 - 5.00
Commit to Ideals (GH)	4.40 (0.55)	2.75 - 5.00	4.32 (0.62)	2.00 - 5.00
Express My Point of View (GH)	4.09 (0.57)	2.00 - 5.00	4.06 (0.60)	2.25 - 5.00
Avoid Hurting Someone (GH)	4.32 (0.58)	2.00 - 5.00	4.16 (0.68)	2.00 - 5.00
Avoid Deception (GH)	4.28 (0.69)	2.00 - 5.00	3.87 (0.90)	1.00 - 5.00
Stable Relations (GH)	4.38 (0.59)	2.00 - 5.00	4.30 (0.65)	2.00 - 5.00
Authenticity (GH)	3.98 (0.63)	2.40 - 5.00	3.97 (0.66)	2.20 - 5.00
Sobriety (GH)	3.54 (0.68)	1.50 - 5.00	3.54 (0.69)	1.67 - 5.00
Justice (GH)	4.20 (0.72)	2.00 - 5.00	3.85 (0.84)	1.50 - 5.00
Test Someone (GH)	3.86 (0.79)	1.67 - 5.00	3.51 (0.87)	1.67 - 5.00
Tell The Truth (GH)	3.90 (0.63)	2.00 - 5.00	3.82 (0.70)	1.50 - 5.00
Hurt Someone (GD)	1.28 (0.55)	1.00 - 3.67	1.26 (0.54)	1.00 - 4.00
Manipulate Others (GD)	1.83 (0.68)	1.00 - 4.50	1.81 (0.75)	1.00 - 4.75
Lie (GD)	1.38 (0.50)	1.00 - 3.67	1.45 (0.64)	1.00 - 4.33
Get What You Want At Any Cost (GD)	1.81 (0.67)	1.00 - 3.75	1.83 (0.67)	1.00 - 4.00

Note: GD, Dishonesty-related goals; GH, Honesty-related goals.

Table S4. *Associations of individual goals with state Honesty-Humility*

	Pearson correlations (person means)				Individual multilevel models (random-intercept and slope)		Multilevel multiple regression (random-intercept)	
Goal class	***r***	***B***	**SE**	***t***	***R^2***	***B***	**SE**	***t***
Satisfaction With Self	0.41***	0.18	0.015	12.17***	0.06	0.04	0.007	6.13***
Make Others Feel Loved	.46***	0.17	0.015	11.34***	0.06	0.03	0.006	5.23***
Help Others	.46***	0.14	0.013	10.51***	0.05	0.01	0.006	2.28*
Happiness	.39***	0.18	0.015	12.29***	0.06	0.06	0.007	8.82***
Hurt someone	-.40***	-0.12	0.016	-7.83***	0.02	-0.10	0.007	-13.67***
Make Others happy	.39***	0.16	0.014	11.76***	0.06	0.03	0.006	5.25***
Respect Others	.55***	0.19	0.014	13.86***	0.08	0.06	0.006	9.44***
Autonomy	.45***	0.16	0.014	10.93***	0.05	0.04	0.006	6.74***
Commit to Ideals	.32***	0.14	0.014	10.21***	0.05	0.02	0.006	3.99***
Express My Point of View	.41***	0.13	0.013	9.91***	0.04	0.01	0.006	1.59
Avoid Hurting Someone	.27***	0.10	0.013	7.76***	0.03	-0.00	0.005	-0.42
Avoid deception	.10	0.05	0.014	3.51***	0.01	-0.01	0.006	-0.92
Stable relations	.45***	0.18	0.015	12.39***	0.06	0.04	0.006	5.47***
Manipulate	-.32***	-0.03	0.011	-2.50*	0.00	-0.04	0.006	-6.16***
Authenticity	.52***	0.19	0.014	13.16***	0.07	0.05	0.007	8.04***
Sobriety	.27***	0.09	0.011	8.22***	0.02	0.01	0.005	1.18
Justice	.08	0.07	0.015	4.53***	0.01	-0.00	0.006	-0.50
Lie	-.44***	-0.14	0.018	-7.89***	0.02	-0.10	0.007	-13.43***
Test someone	.18*	0.10	0.014	7.46***	0.02	0.01	0.006	1.80
Get What You Want At Any Cost	-.34***	-0.02	0.011	-1.64	0.00	-0.03	0.006	-4.47***
Tell the truth	.46***	0.16	0.014	11.06***	0.06	0.05	0.006	8.12***
						R^2 = 0.27		

Note: N = 198 participants; 9,443 observations.

B, Unstandardized regression estimates; SE, Standard Error.

*p<0.05; **p<0.01; ***p<0.001.

Table S5. *List of goal items related to honesty and dishonesty and their relative classes.*

Item (Italian)	Item (English)	Goal Class
1. Sentirmi soddisfatto di me stesso	Feeling satisfied with myself	Satisfaction With Self
2. Far sentire agli altri la mia vicinanza	Making others feel my closeness	Make Others Feel Loved
3. Aiutare qualcuno	Helping someone	Help Others
4. Essere felice	Being happy	Happiness
5. Ferire qualcuno	Hurting someone	Hurt someone
6. Far felice qualcuno	Making someone happy	Make Others happy
7. Rispettare gli altri	Respecting others	Respect Others
8. Essere autonomo	Being autonomous	Autonomy
9. Perseguire i miei ideali	Pursuing my ideals	Commit to Ideals
10. Esprimere il mio punto di vista	Expressing my point of view	Express My Point of View
11. Evitare di ferire qualcuno	Avoid hurting someone	Avoid Hurting Someone
12. Evitare di essere ingannato	Avoid being deceived	Avoid deception
13. Avere relazioni solide con gli altri	Having strong relationships with others	Stable relations
14. Fare in modo che altri facciano ciò che conviene a me	Getting others to do what is convenient for me	Manipulate
15. Essere trasparente con gli altri	Being transparent with others	Authenticity
16. Non darmi delle arie con gli altri	Not to put on airs with others	Sobriety
17. Ottenere giustizia	Getting justice	Justice
18. Ingannare gli altri	Deceiving others	Lie
19. Capire se mi posso fidare di qualcuno	Understanding whether I can trust someone	Test someone
20. Avere ciò che voglio a qualsiasi costo	Having what I want at any cost	Get What You Want At Any Cost
21. Dire la verità senza omissioni	Telling the truth without omissions	Tell the truth

Table S6. *Covariance matrix of person means of EMA goal items.*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. Satisfaction With Self	0.672	0.498	0.494	0.592	0.033	0.464	0.472	0.559	0.628	0.553	0.449	0.467	0.46	0.143	0.525	0.461	0.457	0.036	0.493	0.163	0.523
2. Make Others Feel Loved	0.498	0.696	0.635	0.494	0.062	0.637	0.599	0.495	0.54	0.571	0.609	0.53	0.579	0.18	0.605	0.567	0.559	0.066	0.581	0.169	0.604
3. Help Others	0.494	0.635	0.768	0.484	0.088	0.669	0.607	0.541	0.599	0.577	0.67	0.65	0.543	0.193	0.609	0.624	0.635	0.092	0.677	0.192	0.63
4. Happiness	0.592	0.494	0.484	0.641	0.04	0.504	0.453	0.507	0.578	0.523	0.456	0.459	0.459	0.153	0.499	0.442	0.437	0.052	0.47	0.172	0.513
5. Hurt someone	0.033	0.062	0.088	0.04	0.25	0.099	0.027	0.033	0.128	0.082	0.169	0.258	0.046	0.214	0.07	0.165	0.259	0.249	0.23	0.235	0.096
6 .Make Others happy	0.464	0.637	0.669	0.504	0.099	0.721	0.572	0.468	0.557	0.536	0.643	0.588	0.551	0.215	0.569	0.588	0.615	0.11	0.629	0.214	0.576
7. Respect Others	0.472	0.599	0.607	0.453	0.027	0.572	0.655	0.532	0.522	0.543	0.585	0.499	0.546	0.112	0.617	0.562	0.462	0.022	0.541	0.094	0.621
8. Autonomy	0.559	0.495	0.541	0.507	0.033	0.468	0.532	0.712	0.589	0.543	0.473	0.501	0.467	0.126	0.531	0.535	0.468	0.035	0.546	0.143	0.548
9. Commit to Ideals	0.628	0.54	0.599	0.578	0.128	0.557	0.522	0.589	0.896	0.618	0.589	0.702	0.525	0.241	0.588	0.642	0.718	0.134	0.717	0.272	0.606
10. Express My Point of View	0.553	0.571	0.577	0.523	0.082	0.536	0.543	0.543	0.618	0.689	0.51	0.548	0.524	0.194	0.624	0.528	0.541	0.085	0.585	0.192	0.639
11. Avoid Hurting Someone	0.449	0.609	0.67	0.456	0.169	0.643	0.585	0.473	0.589	0.51	0.994	0.905	0.492	0.259	0.592	0.884	0.79	0.175	0.821	0.273	0.688
12. Avoid deception	0.467	0.53	0.65	0.459	0.258	0.588	0.499	0.501	0.702	0.548	0.905	1.141	0.462	0.398	0.574	0.894	0.93	0.276	1.013	0.428	0.685
13. Stable relations	0.46	0.579	0.543	0.459	0.046	0.551	0.546	0.467	0.525	0.524	0.492	0.462	0.617	0.154	0.556	0.463	0.484	0.042	0.523	0.144	0.537
14. Manipulate	0.143	0.18	0.193	0.153	0.214	0.215	0.112	0.126	0.241	0.194	0.259	0.398	0.154	0.487	0.151	0.249	0.375	0.269	0.407	0.486	0.143
15. Authenticity	0.525	0.605	0.609	0.499	0.07	0.569	0.617	0.531	0.588	0.624	0.592	0.574	0.556	0.151	0.736	0.576	0.555	0.065	0.602	0.136	0.761
16. Sobriety	0.461	0.567	0.624	0.442	0.165	0.588	0.562	0.535	0.642	0.528	0.884	0.894	0.463	0.249	0.576	1.027	0.77	0.17	0.797	0.266	0.687
17. Justice	0.457	0.559	0.635	0.437	0.259	0.615	0.462	0.468	0.718	0.541	0.79	0.93	0.484	0.375	0.555	0.77	1.121	0.275	0.875	0.407	0.642
18. Lie	0.036	0.066	0.092	0.052	0.249	0.11	0.022	0.035	0.134	0.085	0.175	0.276	0.042	0.269	0.065	0.17	0.275	0.291	0.242	0.289	0.088
19. Test someone	0.493	0.581	0.677	0.47	0.23	0.629	0.541	0.546	0.717	0.585	0.821	1.013	0.523	0.407	0.602	0.797	0.875	0.242	1.044	0.433	0.66
20. Get What You Want At Any Cost	0.163	0.169	0.192	0.172	0.235	0.214	0.094	0.143	0.272	0.192	0.273	0.428	0.144	0.486	0.136	0.266	0.407	0.289	0.433	0.576	0.136
21. Tell the truth	0.523	0.604	0.63	0.513	0.096	0.576	0.621	0.548	0.606	0.639	0.688	0.685	0.537	0.143	0.761	0.687	0.642	0.088	0.66	0.136	0.932

Table S7. *Covariance matrix of deviations from person means of EMA goal items.*

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1. Satisfaction With Self	0.914	0.284	0.274	0.358	-0.035	0.28	0.304	0.32	0.347	0.305	0.157	0.129	0.303	0.043	0.29	0.188	0.135	-0.012	0.202	0.111	0.248
2. Make Others Feel Loved	0.284	1.114	0.474	0.325	-0.01	0.52	0.421	0.176	0.232	0.367	0.344	0.114	0.548	0.05	0.394	0.256	0.166	0.005	0.235	0.066	0.315
3. Help Others	0.274	0.474	1.197	0.249	0.005	0.47	0.42	0.234	0.251	0.367	0.333	0.157	0.42	0.085	0.355	0.274	0.191	0.011	0.251	0.08	0.308
4. Happiness	0.358	0.325	0.249	0.871	-0.022	0.332	0.293	0.264	0.267	0.276	0.194	0.123	0.329	0.042	0.286	0.181	0.111	-0.011	0.197	0.08	0.23
5. Hurt someone	-0.035	-0.01	0.005	-0.022	0.613	-0.008	-0.019	-0.018	0.007	0.033	0.033	0.084	-0.013	0.102	0.007	0.042	0.074	0.139	0.046	0.101	0.034
6 .Make Others happy	0.28	0.52	0.47	0.332	-0.008	1.139	0.417	0.192	0.221	0.32	0.36	0.112	0.467	0.058	0.361	0.233	0.158	0.008	0.24	0.05	0.291
7. Respect Others	0.304	0.421	0.42	0.293	-0.019	0.417	1.052	0.266	0.256	0.385	0.317	0.146	0.451	0.071	0.428	0.281	0.153	-0.006	0.252	0.043	0.337
8. Autonomy	0.32	0.176	0.234	0.264	-0.018	0.192	0.266	0.975	0.301	0.244	0.151	0.155	0.216	0.049	0.226	0.161	0.142	-0.006	0.18	0.102	0.189
9. Commit to Ideals	0.347	0.232	0.251	0.267	0.007	0.221	0.256	0.301	1.024	0.34	0.189	0.179	0.262	0.096	0.276	0.194	0.228	0.011	0.238	0.137	0.269
10. Express My Point of View	0.305	0.367	0.367	0.276	0.033	0.32	0.385	0.244	0.34	1.141	0.251	0.239	0.406	0.131	0.439	0.244	0.268	0.007	0.3	0.129	0.412
11. Avoid Hurting Someone	0.157	0.344	0.333	0.194	0.033	0.36	0.317	0.151	0.189	0.251	1.224	0.206	0.293	0.055	0.298	0.312	0.178	0.019	0.255	0.094	0.286
12. Avoid deception	0.129	0.114	0.157	0.123	0.084	0.112	0.146	0.155	0.179	0.239	0.206	0.98	0.153	0.154	0.175	0.202	0.276	0.086	0.34	0.167	0.194
13. Stable relations	0.303	0.548	0.42	0.329	-0.013	0.467	0.451	0.216	0.262	0.406	0.293	0.153	1.082	0.066	0.415	0.254	0.17	-0.011	0.281	0.066	0.326
14. Manipulate	0.043	0.05	0.085	0.042	0.102	0.058	0.071	0.049	0.096	0.131	0.055	0.154	0.066	0.823	0.097	0.083	0.144	0.109	0.143	0.245	0.108
15. Authenticity	0.29	0.394	0.355	0.286	0.007	0.361	0.428	0.226	0.276	0.439	0.298	0.175	0.415	0.097	1.022	0.271	0.208	0.005	0.285	0.095	0.471
16. Sobriety	0.188	0.256	0.274	0.181	0.042	0.233	0.281	0.161	0.194	0.244	0.312	0.202	0.254	0.083	0.271	1.149	0.176	0.042	0.248	0.1	0.255
17. Justice	0.135	0.166	0.191	0.111	0.074	0.158	0.153	0.142	0.228	0.268	0.178	0.276	0.17	0.144	0.208	0.176	0.889	0.068	0.252	0.166	0.216
18. Lie	-0.012	0.005	0.011	-0.011	0.139	0.008	-0.006	-0.006	0.011	0.007	0.019	0.086	-0.011	0.109	0.005	0.042	0.068	0.577	0.043	0.107	0.006
19. Test someone	0.202	0.235	0.251	0.197	0.046	0.24	0.252	0.18	0.238	0.3	0.255	0.34	0.281	0.143	0.285	0.248	0.252	0.043	0.984	0.143	0.266
20. Get What You Want At Any Cost	0.111	0.066	0.08	0.08	0.101	0.05	0.043	0.102	0.137	0.129	0.094	0.167	0.066	0.245	0.095	0.1	0.166	0.107	0.143	0.955	0.113
21. Tell the truth	0.248	0.315	0.308	0.23	0.034	0.291	0.337	0.189	0.269	0.412	0.286	0.194	0.326	0.108	0.471	0.255	0.216	0.006	0.266	0.113	1.069

Table S8. *Estimates (standardized loadings) from the CFA model with two correlated factors at the WP level and saturated BP level*

Latent_Factor Within-Person	Item	Standardized Loading	SE	*z*	*p*
GH	Satisfaction With Self	0.511	0.009	59.624	<.001
GH	Make Others Feel Loved	0.616	0.007	83.297	<.001
GH	Help Others	0.564	0.008	70.691	<.001
GH	Happiness	0.516	0.009	60.670	<.001
GH	Hurt someone	0.581	0.008	74.404	<.001
GH	Make Others happy	0.616	0.007	83.904	<.001
GH	Respect Others	0.400	0.010	41.658	<.001
GH	Autonomy	0.462	0.009	51.052	<.001
GH	Commit to Ideals	0.584	0.008	75.307	<.001
GH	Express My Point of View	0.439	0.009	47.491	<.001
GH	Avoid Hurting Someone	0.319	0.010	31.218	<.001
GH	Avoid deception	0.637	0.007	89.648	<.001
GH	Stable relations	0.627	0.007	86.761	<.001
GH	Manipulate	0.402	0.010	42.115	<.001
GH	Authenticity	0.357	0.010	35.943	<.001
GH	Sobriety	0.460	0.009	50.67	<.001
GH	Justice	0.537	0.008	64.756	<.001
GD	Lie	0.313	0.015	20.582	<.001
GD	Test someone	0.525	0.016	33.746	<.001
GD	Get What You Want At Any Cost	0.332	0.015	21.815	<.001
GD	Tell the truth	0.505	0.015	32.803	<.001

Note: GH – goals of honesty; GD – goals of dishonesty.

Table S9. *Estimates (standardized loadings) from the CFA model with two correlated factors at the BP level and saturated WP level*

Latent_Factor Between-Person	Item	Standardized Loading	SE	*z*	*p*
GH	Satisfaction With Self	0.829	0.024	34.043	<.001
GH	Make Others Feel Loved	0.938	0.011	88.998	<.001
GH	Help Others	0.927	0.012	77.375	<.001
GH	Happiness	0.827	0.024	33.825	<.001
GH	Hurt someone	0.919	0.013	69.934	<.001
GH	Make Others happy	0.922	0.013	72.726	<.001
GH	Respect Others	0.819	0.026	32.014	<.001
GH	Autonomy	0.841	0.023	36.937	<.001
GH	Commit to Ideals	0.907	0.015	61.957	<.001
GH	Express My Point of View	0.837	0.023	35.819	<.001
GH	Avoid Hurting Someone	0.791	0.029	27.529	<.001
GH	Avoid deception	0.895	0.016	55.382	<.001
GH	Stable relations	0.927	0.012	77.520	<.001
GH	Manipulate	0.801	0.027	29.174	<.001
GH	Authenticity	0.767	0.031	24.591	<.001
GH	Sobriety	0.835	0.023	35.676	<.001
GH	Justice	0.862	0.020	42.600	<.001
GD	Lie	0.955	0.009	101.241	<.001
GD	Test someone	0.764	0.033	23.168	<.001
GD	Get What You Want At Any Cost	1.005	0.008	131.651	<.001
GD	Tell the truth	0.761	0.034	22.662	<.001

Note: GH – goals of honesty; GD – goals of dishonesty.

Table S10. *Pearson correlations (EMA and T1)*

	GH_EMA	GD_EMA	H_EMA	T1_GH	T1_GD	T1_H HAS	T1_H HEXACO-60	H_PCA_T1	T1_HAS_E	T1_HAS_X	T1__HAS_A	T1_HAS_C
GH_EMA												
GD_EMA	.41***											
H_EMA	.41***	-.41***										
T1_GH	.35***	-.05	.38***									
T1_GD	.10	.56***	-.33***	-.24***								
T1_H HAS	.21**	-.24***	.53***	.54***	-.42***							
T1_H HEXACO-60	.01	-.49***	.41***	.34***	-.72***	.47***						
H_PCA_T1	.13	-.43***	.54***	.50***	-.68***	.83***	.87***					
T1_HAS Emotionality	-.04	.02	-.05	.04	-.02	-.12	-.03	-.10				
T1_HAS Extraversion	.16*	.00	.07	.27***	-.10	.22**	.07	.16*	-.16*			
T1_HAS Agreeableness	.14	-.01	.17*	.04	-.08	.20**	.04	.16*	-.06	-.12		
T1_HAS Conscientiousness	.16*	.04	.19**	.24***	-.09	.36***	.13	.29***	-.10	-.02	.14	
T1_HAS Openness	.05	.02	-.04	.18*	.01	.05	-.06	.00	.00	.22**	.04	-.01

Note: N = 198 (total sample); EMA variables are aggregated (person means). H = Honesty; GH = Honesty-related goals; GD = Dishonesty-related goals; PCA = Principal Component Analysis.

Table S11. *Pearson correlations (EMA and T2)*

	GH_EMA	GD_EMA	H_EMA	T2_GH	T2_GD	T2_H HAS	T2_H HEXACO-60	H_PCA_T2	T2_HAS_E	T2_HAS_X	T2_HAS_A	T2_HAS_C
GH_EMA												
GD_EMA	.41***											
H_EMA	.41***	-.41***										
T2_GH	.41***	-.07	.46***									
T2_GD	.11	.67***	-.45***	-.38***								
T2_H HAS	.18*	-.40***	.68***	.56***	-.63***							
T2_H HEXACO-60	-.02	-.55***	.45***	.44***	-.75***	.59***						
H_PCA_T2	.07	-.55***	.60***	.53***	-.78***	.87***	.91***					
T2_HAS Emotionality	-.02	.02	-.03	.03	-.04	-.07	.03	-.01				
T2_HAS Emotionality	.24**	.06	.11	.18*	-.08	.23**	-.01	.09	-.14*			
T2_HAS Agreeableness	.18*	-.12	.32***	.22**	-.24***	.43***	.17*	.34***	-.07	.03		
T2_HAS Conscientiousness	.09	-.09	.21**	.28***	-.22**	.36***	.15*	.29***	.00	.01	.21**	
T2_HAS Openness	.04	-.06	.00	.16*	-.08	.13+	.03	.07	.05	.17*	.11	.06

Note: N = 191 (subsample with T2 completed); EMA variables are aggregated (person means). H = Honesty; GH = Honesty-related goals; GD = Dishonesty-related goals; PCA = Principal Component Analysis.

Table S12. *Pearson correlations (T1 and T2)*

	T1_GH	T1_GD	T1_H HAS	T1_H HEXACO-60	H_PCA_T1	T2_GH	T2_GD	T2_H HAS	T2_H HEXACO-60
T1_GH									
T1_GD	-.24***								
T1_H HAS	.54***	-.42***							
T1_H HEXACO-60	.34***	-.72***	.47***						
H_PCA_T1	.50***	-.68***	.83***	.87***					
T2_GH	.65***	-.26***	.44***	.38***	.47***				
T2_GD	-.27***	.74***	-.47***	-.68***	-.68***	-.38***			
T2_H HAS	.49***	-.50***	.77***	.50***	.73***	.56***	-.63***		
T2_H HEXACO-60	.34***	-.72***	.50***	.91***	.83***	.44***	-.75***	.59***	
H_PCA_T2	.44***	-.69***	.68***	.80***	.87***	.53***	-.78***	.87***	.91***

Note: N = 191 (subsample with T2 completed); EMA variables are aggregated (person means). H = Honesty; GH = Honesty-related goals; GD = Dishonesty-related goals; PCA = Principal Component Analysis.

Table S13. *Associations of individual goals with momentary lying behavior (EMA)*

			Spearman's rho correlations (person means)		Individual multilevel models (random-intercept and slope)		Multilevel multiple regression (random-intercept)	
Goal class	***r***	***B***	**SE**	***z***	***R²***	***B***	**SE**	***z***
Satisfaction With Self	-.16*	-0.08	0.035	-2.41*	0.00	-0.06	0.018	-3.08**
Make Others Feel Loved	-.07	0.01	0.032	0.19	0.00	-0.03	0.018	-1.85
Help Others	.07	0.03	0.031	1.11	0.00	0.01	0.017	0.47
Happiness	-.21**	-0.11	0.034	-3.31***	0.01	-0.10	0.018	-5.20***
Hurt someone	.19**	0.19	0.042	4.50***	0.01	0.11	0.016	6.83***
Make Others happy	.03	0.05	0.032	1.42	0.00	0.03	0.018	1.80
Respect Others	-.03	-0.01	0.031	-0.44	0.00	-0.02	0.018	-1.34
Autonomy	-.08	-0.06	0.034	-1.69	0.00	-0.04	0.017	-2.41*
Commit to Ideals	-.08	-0.05	0.033	-1.67	0.00	-0.03	0.018	-1.75
Express My Point of View	-.12	0.05	0.031	1.52	0.00	0.03	0.017	1.77
Avoid Hurting Someone	-.03	0.09	0.029	3.05**	0.01	0.03	0.015	1.83
Avoid deception	.01	0.13	0.032	4.02***	0.01	0.06	0.017	3.50***
Stable relations	-.04	0.03	0.034	0.86	0.00	0.02	0.018	1.30
Manipulate	-.13	0.23	0.030	7.64***	0.02	0.07	0.016	4.23***
Authenticity	-.12	-0.05	0.032	-1.43	0.00	-0.01	0.018	-0.72
Sobriety	.01	0.02	0.029	0.73	0.00	0.01	0.015	0.57
Justice	-.05	0.13	0.032	4.20***	0.01	-0.01	0.018	-0.50
Lie	.16*	0.35	0.039	8.85***	0.03	0.15	0.016	9.79***
Test someone	-.02	0.09	0.032	2.84**	0.01	0.04	0.017	2.41*
Get What You Want At Any Cost	.18**	0.17	0.030	5.53***	0.01	0.07	0.016	4.81***
Tell the truth	.02	-0.07	0.032	-2.28*	0.00	-0.04	0.017	-2.47*
						R^2 = 0.06		

Table S14. *Differences in means of Honesty-Humility facets T1 – T2 and associated paired t-test results.*

	T1		T2					
Facet	**Mean**	**SD**	**Mean**	**SD**	**Difference in means T1-T2**	***t***	**df**	***p***
Sincerity	5.80	0.70	5.85	0.69	-0.05	-1.37	190	0.171
Fairness	5.76	0.66	5.73	0.71	0.04	0.82	190	0.416
Modesty	5.42	0.83	5.45	0.85	-0.03	-0.56	190	0.575
Greed-Avoidance	5.34	0.96	5.44	0.92	-0.10	-1.86	190	0.065
Honesty general	5.95	0.68	5.94	0.69	0.01	0.19	190	0.848

Note: N = 191 (subsample with T2 completed).

Table S15. *Latent growth model –loadings (λ) and intercepts (ν) for T2-T1 honesty trait score (ΔH in Figure 2). Standardized loadings are in parentheses.*

	λ	ν
ΔSinc	1.00 (.620)	-0.131
ΔFair	1.328*** (.690)	-0.276
ΔMod	.952*** (.458)	-0.146
ΔGreeAv	1.10***(.473)	-0.102
ΔHgen	1.192*** (.688)	-0.223

Note: The first loading was constrained to unity to guarantee model identification
*$p < .05$; **$p < .01$; ***$p < .001$

Table S16. *Psi matrix encoding endogenous variances (diagonal elements) and covariances (off-diagonal elements). Standardized estimates are presented in parentheses.*

	I_H	S_H	I_{GH}	S_{GH}	I_{GD}	S_{GD}	Δ_H
I_H	.195*** (.422)						
S_H	-.003(-.181)	.002*** (.510)					
I_{GH}			.576*** (1.000)				
S_{GH}			-.004 (-.114)	.002*** (1.000)			
I_{GD}			.174*** (.405)		.319*** (1.000)		
S_{GD}				.000* (.240)	-.002 (-.073)	0.002*** (1.000)	
Δ_H							.070*** (.734)

Note: I = latent intercept; S = latent slope (growth factor); GH = Honesty-related goals; GD = Dishonesty-related goals; Δ_H = Latent delta score of honesty. Empty cells indicate parameters that were not freely estimated in the model.
*p < .05; **p < .01; ***p < .001

Table S17. *Theta matrix, encoding the residual variances and covariances of observed variables (standardized values in parentheses)*

	H_0 - H_{14}	GH_0 - GH_{14}	GD_0 - GD_{14}	ΔSincerity	ΔFairness	ΔModesty	ΔGreedAvoidance	ΔHgeneric
H_0 - H_{14}	0.10-0.20 (.12-.30)							
GH_0 - GH_{14}	.05 (.27-.47)	0.10-0.19 (.12-.25)						
GD_0 - GD_{14}	-.02 (-.15, -.08)	.03 (.14, .28)	0.09-0.21 (.13-.39)					
ΔSinc				.15 (.62)				
ΔFair					.18 (.52)			
ΔMod						.32(.79)		
ΔGreeAv							.40 (.78)	
ΔHgen								.15 (.53)

Note: Variances of EMA variables are reported as a range. Covariances between EMA observed variables are referred to corresponding timepoints (e.g., between H_0 and GH_0, H_0 and GD_0, and between GH_0 and GD_0). Empty cells indicate parameters that were not freely estimated in the model. All *p*s < .001.

Supplementary Material S18.

ACH scale with instructions

Italian (original): Di seguito, le chiediamo di indicare quanto ciascuno dei seguenti aggettivi (o breve espressione) risulti appropriato a descriverla, usando la seguente scala di risposta. Indichi 1 se l'aggettivo non la descrive per niente, 7 se l'aggettivo la descrive completamente. Usi le risposte 2, 3, 4, 5 e 6 per indicare gradi intermedi di importanza. Sappiamo che in molti casi la risposta dipende dalla situazione specifica. Le chiediamo però di rispondere pensando a quanto ciascun aggettivo la descrive "in media", attraverso diverse situazioni.

English (translated): Below, we ask you to indicate how appropriate each of the following adjectives (or brief expressions) is in describing you, using the following response scale. Indicate 1 if the adjective does not describe you at all, 7 if the adjective describes you completely. Use responses 2, 3, 4, 5, and 6 to indicate intermediate degrees of importance. We understand that in many cases, the answer depends on the specific situation. However, we ask you to respond thinking about how each adjective describes you "on average," across different situations.

1. amante della verità (truth-lover)
2. autentico/a (authentic)
3. avido/a (avid)
4. bugiardo/a (liar)
5. diretto/a (direct)
6. falso/a (false)
7. generoso/a (generous)
8. giusto/a (just)
9. immorale (immoral)
10. ingiusto/a (unfair)
11. insincero/a (insincere)
12. ipocrita (hypocritical)
13. menzognero/a (lying)
14. modesto/a (modest)
15. non avaro/a (not stingy)
16. obiettivo/a (objective)
17. sbruffone/a (boaster)
18. semplice (simple)
19. sincero/a (sincere)
20. vanitoso/a (vain)
21. venale (mercenary)
22. vero/a (truthful)

Scoring:

- Sincerity: 1, 2, 4*, 5, 6*, 11*, 12*, 13*, 19, 22
- Fairness: 8, 9*, 10*, 16
- Modesty: 14, 17*, 18, 20*
- Greed-Avoidance: 3*, 7, 15, 21*

Items marked with a star(*) must be reverse-coded with the formula 8-X, where X is the participants' response. Then, a score for each facet can be obtained by averaging the items.

Note. The scale can be administered together with the HEXACO Adjective Scales (Romano et al., 2023). In this case, items 3, 12, and 19, which are already included in the HAS, should not be administered twice.

Part II

5

Emotional Interdependence in Romantic Relationships: A Couple ESM Study

Galkina, A., Viechtbauer, W., Perugini, M., Costantini, G. (2025), Emotional Interdependence in Romantic Relationships: A Couple ESM Study. In: (This thesis).

Abstract

Emotional interconnection is considered a defining feature of romantic relationships, yet recent research reveals considerable heterogeneity in how partners' emotions interact in daily life. This study investigated both contemporaneous and temporal patterns of emotional interdependence between romantic partners and examined whether being in contact moderates these associations. We collected intensive longitudinal data from 76 heterosexual Italian couples (N = 152 individuals), prompting them five times per day over 30 days. Participants completed assessments of positive affect (PA) and negative affect (NA) and reported whether they had been in contact with their partner since the previous assessment. We employed multilevel modeling within an extended Actor-Partner Interdependence Model framework to estimate concurrent and time-lagged emotional associations while controlling for autoregressive effects. Significant emotional interdependence emerged for PA both contemporaneously and temporally. However, no significant partner effects were found for NA in either concurrent or lagged models. Contact between partners showed a positive main effect on males' PA but did not significantly moderate the strength of emotional interdependence. Substantial between-couple variability in emotional baselines was observed, with partners showing greater similarity in NA than PA. Partners' positive emotions show modest but consistent interdependence in daily life, operating through both immediate and delayed pathways. The absence of NA interdependence and null contact moderation effects suggest that emotional linkage in romantic relationships is more complex and context-dependent than theoretical models often assume, highlighting the need for more nuanced approaches to understanding interpersonal emotion dynamics.

Introduction

Romantic relationships involve a significant level of emotional reliance on each other (Berscheid & Ammazzalorso, 2001; Butler, 2015). Previous studies indicate that individuals are more likely to express their emotions to those they are close to, such as romantic partners, rather than to those with whom they are less connected (D. Y. Liu et al., 2021; Rauers & Riediger, 2023). Partners in romantic relationships continuously influence each other's emotional experiences through daily interactions, creating a dynamic emotional system that extends beyond individual boundaries (Butler, 2011). The ways in which partners' emotions become interconnected — through processes of sharing, responding to, and regulating each other's affective states — have profound implications for both individual well-being and relationship functioning (Randall & Schoebi, 2015; Sels et al., 2020). Understanding how partners' emotions connect and influence each other in everyday life therefore represents a crucial step for comprehending the mechanisms through which romantic relationships shape psychological adjustment and relational outcomes.

The implications of emotional interdependence for relationship functioning and individual well-being are substantial. Greater emotional linkage between partners has been associated with both positive and negative relationship outcomes, depending on the specific pattern of interdependence and the emotions involved (Butler, 2011; Randall et al., 2013). For instance, synchrony in positive emotions tends to predict relationship satisfaction and stability (e.g., Gonzaga et al., 2007), whereas reciprocal escalation of negative emotions characterizes distressed relationships (Gottman & Levenson, 2002). At the individual level, partners' capacity to influence each other's emotional states has implications for psychological adjustment, with evidence suggesting that being receptive to a partner's positive affect (PA) can buffer against psychological distress over time (Randall & Schoebi, 2015). Moreover,

previous research has demonstrated that in couples with strong emotional interdependence, partners reported more life satisfaction and less empathic concern (Sels et al., 2016).

Research has established that emotional interdependence operates through multiple temporal patterns; therefore, it is particularly important to address partners' emotions as a dynamic, interrelated system that changes over time (Boiger & Mesquita, 2012). Emotion dynamics, as a crucial characteristic of dyad functioning, were largely addressed by Butler (2011), who was one of the first to examine them from a temporal interpersonal perspective. Partners' emotions can covary concurrently, such that when one partner experiences elevated PA or negative affect (NA) at a given moment, the other partner tends to experience similar affective states (Butner et al., 2007; Saxbe & Repetti, 2010). Beyond concurrent associations, partners' emotions also exhibit temporal linkage, wherein one partner's emotional state at a prior time point predicts the other partner's subsequent emotional experience (Schoebi, 2008; Thompson & Bolger, 1999).

Capturing the dynamic nature of emotional interdependence in romantic relationships requires methodological approaches that can assess partners' emotions as they naturally unfold in daily life. Experience sampling methodology (ESM), which involves repeated assessments of participants' current emotional states across multiple occasions in their natural environments, has emerged as a particularly valuable tool for this purpose (Bolger et al., 2003; Trull & Ebner-Priemer, 2014). By obtaining multiple measurements from both partners over days or weeks, ESM designs enable researchers to examine within-person and within-dyad emotional processes while minimizing retrospective biases inherent in global self-report measures (Shiffman et al., 2008). The intensive longitudinal data generated by ESM studies provide the temporal resolution necessary to model how partners' emotions covary and influence each other across time.

ESM approaches have revealed important insights into emotional processes in romantic relationships that would not be detectable through cross-sectional or less intensive longitudinal designs. For example, a COVID-19 study found that when romantic partners experienced more stress than usual, they felt less close to their partners (Salo et al., 2022). Research using multiple assessments per day is especially important for studying emotional dynamics between partners, as emotional changes can occur within relatively short time windows. Previous research has established that one partner's morning emotions could predict the other partner's emotions later in the day (Schoebi, 2008). In addition, ESM data allow researchers to examine how contextual factors — such as whether partners are physically together or apart — moderate emotional processes, providing insight into the situational contingencies of emotional linkage (e.g., Sels et al., 2016).

The analysis of ESM data from romantic couples presents unique methodological challenges, as the data structure involves multiple levels of nesting: repeated observations are nested within individuals, and individuals are nested within dyads. Several analytic frameworks have been developed to model this complex data structure appropriately while accounting for the non-independence of partners' responses. The Actor-Partner Interdependence Model (APIM) has become a widely used framework for examining dyadic processes in longitudinal data (Iida et al., 2023; Kenny et al., 2006). The APIM distinguishes between actor effects (the association between an individual's predictor and their own outcome) and partner effects (the association between an individual's predictor and their partner's outcome), thereby allowing researchers to simultaneously examine intrapersonal and interpersonal processes.

A fundamental consideration in the APIM framework concerns whether dyad members can be meaningfully distinguished by some characteristic. For dyads to be considered distinguishable, there should be a characteristic that is theoretically reasonable to be used to distinguish the dyad members and that actually leads to empirical differences on a certain

variable of interest. For example, while heterosexual couples are theoretically distinguishable by gender, the empirical distinguishability depends on whether systematic differences actually exist in the data between the two partner types (Gonzalez & Griffin, 1999). It might happen that, depending on the specific research context, gender might produce meaningful differences, as well as not demonstrate them. Olsen and Kenny (2006) provide a comprehensive framework for evaluating distinguishability within the APIM, examining whether actor and partner effects differ between males and females, whether residual variances in the outcome variables are equivalent across genders, and whether the means and variances of predictor variables show gender-based differences. This empirical assessment determines whether the theoretical distinction translates into meaningful analytical differences that warrant separate parameter estimation for each partner type. An important consideration for APIM models, thus, is not to separate the analysis by actor and partner, which would force researchers to draw separate conclusions for each partner, even when no statistically significant differences exist between them. Therefore, the common approach is to add the distinguishing factor (e.g., gender) as a moderator into the model (Laurenceau & Bolger, 2005).

For ESM data specifically, longitudinal extensions of the APIM (LAPIM) were developed. They can model both concurrent associations between partners' emotions and time-lagged predictions, wherein one partner's emotion at a previous time point predicts the other partner's emotion at a subsequent time point (Iida et al., 2023; Laurenceau & Bolger, 2005). These models can be implemented using multilevel modeling or structural equation modeling frameworks, each offering distinct advantages depending on the research questions and data characteristics (Ledermann & Kenny, 2017). An important consideration in applying these models to emotional interdependence is distinguishing between different patterns of association. Concurrent associations may reflect shared situational influences, emotional contagion, or other processes that produce simultaneous similarity in partners' emotions

(Herrando & Constantinides, 2021). Time-lagged associations more clearly suggest that one partner's emotional state is related to subsequent changes in the other partner's emotions, providing stronger evidence for the interpersonal emotional transmission (Larson & Almeida, 1999).

Despite substantial progress in documenting emotional interdependence in romantic relationships, several important gaps remain in the literature. First, there is still no consistency in findings on whether this interdependence can be typically observed in romantic couples, both in synchronous and time-lagged perspectives. On the one hand, various studies indicate that emotional interconnections between romantic partners do occur, especially those in negative emotions (Bar-Kalifa & Sened, 2020; Butler, 2011; Butner et al., 2007; Larson & Almeida, 1999; Saxbe & Repetti, 2010). On the other hand, recent studies assessing emotional interdependence on a couple-by-couple basis have found some controversy and heterogeneity within the interpersonal emotional connections (Ferrer & Widaman, 2011; Madhyastha et al., 2011; Sels et al., 2016, 2020; Steele et al., 2014). In addition, given the lack of a consistent approach to the question, it remains unclear how these findings relate to one another and how various factors could moderate the levels of emotional interdependence observed among couples.

Contact between partners is one of the most evident contextual factors that could moderate emotional interdependence; however, it has, to date, been examined only to a limited extent. Theoretical perspectives on emotional coregulation emphasize that partners' capacity to influence each other's emotions should depend on opportunities for interaction (Butler & Randall, 2013; Sbarra & Hazan, 2008). Consistent with this idea, some evidence suggests that partners' connectedness is stronger when partners spend more time together (Butner et al., 2007; Papp et al., 2013). However, partners might also influence each other's emotions during periods of separation through anticipation, rumination, or communication at a distance. So far,

very little attention has been paid to whether the presence or absence of contact moderates both concurrent and temporal emotional associations, and whether this moderation differs across positive and negative emotions. Clarifying these moderating effects would provide insight into the interpersonal mechanisms underlying emotional interdependence.

The Current Study

The present study addresses the described research gaps by examining both contemporaneous and temporal patterns of emotional interdependence in romantic couples, and testing whether contact between partners moderates these associations. Using ESM data collected from both partners in romantic relationships, we assessed their emotional experiences at five time points per day across a month. As we examined the relationships between emotional states of the partners bidirectionally and not unidirectionally, and we also added the partner effect of the emotions of the same valence besides those from the opposite valence, our final model represents an extended version of the classic APIM model (see Figure 1 for details). Our analysis included only heterosexual couples to be able to capture the distinguishability of female and male partners within the model, while homosexual couples would require either a different type of analysis based on indistinguishable dyad members or consideration of other characteristics to make dyad members distinguishable (Gistelink & Loyes, 2020; Kenny et al., 2006; Laurenceau & Bolger, 2005; Rostosky & Riggle, 2017). We employed multilevel modeling within this extended version of the APIM framework to estimate concurrent associations (the degree to which partners' emotions covary at the same time point) and time-lagged associations (the degree to which one partner's emotion predicts the other partner's subsequent emotion, adding sex as a moderator). In addition, we tested whether these associations were moderated by whether partners were in contact during the interval between assessments, allowing us to examine whether emotional interdependence depends on opportunities for direct interaction.

We hypothesized that partners' emotions would show both concurrent similarity and time-lagged associations, reflecting both shared experiences and interpersonal emotional transmission. Moreover, we expected that both forms of emotional interdependence would be stronger when partners have been in contact, as it should facilitate emotional contagion and provide opportunities for emotional communication and regulation. We anticipated that these patterns would be evident for both positive and negative emotions, though the magnitude of associations could vary across emotional states. By examining the emotional dynamics within romantic couples and testing the moderating role of contact, this study aims to provide a more comprehensive and nuanced understanding of how partners' emotional lives become intertwined in the context of everyday romantic relationships. This study was preregistered on the OSF platform (https://osf.io/c5j2u/overview). This chapter, however, covers only a part of the analysis described in the preregistration.

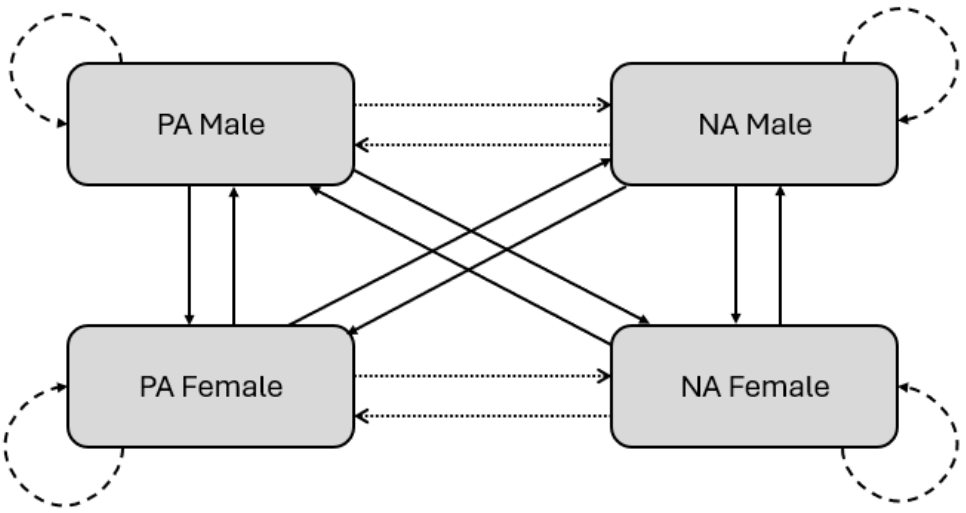


Figure 1. Interpersonal emotion dynamics model used in the current study. The four nodes represent positive (PA) and negative affect (NA) of both partners (male and female). The arrows represent slopes, i.e., the effects of affective states of the partners on each other. Solid arrows correspond to partner effects, dotted arrows to actor effects, and dashed arrows to autoregressive effects.

Methods

Participants

Couples were recruited through social media, posters in university buildings, the internal university platform, and word of mouth. Inclusion criteria were: (1) being over the age of 18 and less than the age of 40[1], (2) having been in a relationship for at least six months, (3) both partners' willingness to participate, (4) both partners staying in the same time zone for the whole duration of the study, and (5) compatibility of smartphones with the application used in the study. All participants of the study had Italian nationality. Following the preregistered criteria, five couples were excluded from the study because they had fewer than 20 ESM observations in common (i.e., at the same measurement occasion). The full demographic characteristics of the final sample are presented in Table 1.

[1] Though we specified the inclusion criteria in the information letter for the participants, there was one participant, aged 43, who enrolled in the study. We decided to include the data of the relative couple, as the age difference from the study's age limit is small.

Table 1. *Descriptive statistics of the final sample.*

	Total sample	**Males**	**Females**
N sample initial phase/ESM	150/152 (76 couples)	74/76[2]	76/76
Age: mean (SD, range)	24.04 (4.23, 18-43)	24.68 (4.47, 18-43)	23.42 (3.90, 18-39)
N (%) by level of Education			
Bachelor's degree or higher	77 (51%)	33 (45%)	44 (58%)
High school diploma	68 (45%)	37 (50%)	31 (41%)
Lower than high school	5 (3%)	4 (5%)	1 (1%)
ESM observations per participant: mean (SD, range)	128.13 (22.80, 31 - 149)	123.71 (26.89, 31 - 149)	132.55 (16.85, 74 - 149)
ESM observations in common per couple: mean (SD, range)	112 (28.8, 27 - 146)		
Relationship length in months[3]: mean (SD, range)	44.96 (34.81, 6 - 214)		
Married, N (%)	3 couples (4%)		
With children, N (%)	2 couples (3%)		
Living together:			
Yes	8 couples (11%)		
No	58 couples (76%)		
Partially	7 couples (9%)		
Couples with incoherent responses (male - partially, female - no)	3 couples (4%)		

[2] We were unable to connect the ESM data of two male participants with the demographic questionnaires completed in the initial phase.

[3] Five couples were excluded from the calculation of relationship length, as the partners within these couples indicated different dates of the start of their relationships.

Procedure

First, participants were asked to read the information about the study on a dedicated study website (https://couplestudy.flazio.com/about). If they agreed to participate in the study, they followed a link from the website to the initial questionnaire, which was implemented on the Qualtrics platform. The initial questionnaire contained questions that captured demographic information (e.g., age and relationship duration), as well as several assessments that were not analyzed in the current study. There, they received more formal standardized information about the study, gave informed consent, and completed a battery of questionnaires. Additionally, participants were asked how they preferred to receive briefing information about the ESM phase: in person at the university office, via video call, or through a WhatsApp chat[4]. Most of the couples (N = 51) requested a briefing via WhatsApp. During the briefing, all participants were provided with written instructions on using the SEMA3 smartphone application (O'Brien et al., 2024) and were guided through its installation and necessary setup.

In the ESM phase, which began the day after app installation and lasted 30 days, participants received app notifications to complete the questionnaires five times a day at fixed times (10:00, 13:00, 16:00, 19:00, 22:00) for a total of 150 prompts. Each ESM prompt contained the items described in the *Materials* section (see below), along with some other ESM measures not analyzed in this study. In total, participants were prompted with 16 question items if they had contact with their partner, and 11 items if they were not in contact. They had an hour to respond to the prompt, during which they received three more notifications (after every 15 minutes) if no response was given. Participants were instructed to respond individually, not sharing their responses with their partners. Additionally, they were asked to report on all cases of app malfunctioning to the research team through the dedicated WhatsApp contact or email.

[4] For the purposes of the study, a dedicated WhatsApp business account was used, to which all members of the research team had access, but which did not correspond to anyone's personal contact. After the end of the data collection, all the contacts of the participants, as well as the chat history, were deleted.

These beeps were not counted as missed at the end of the study for the reimbursement. In total, we received 176 notifications from 86 participants about the app malfunctioning.

After the end of the ESM phase, participants received a Qualtrics link to complete the final questionnaire battery, which was identical to the initial one, apart from the demographic questions. Upon completing the final questionnaire, all participants received information about their compliance rate and instructions on how to receive the relevant reimbursement within one week. A compliance rate of 90% and more corresponded to a €20 compensation, 80% and more to a €15 compensation, and 70% and more to a €10 compensation. Partners within a couple could receive different compensation amounts based on their individual compliance rates.

The study was approved and deemed to involve minimal risk for participants by the Committee for Research in Psychology (CRIP) of the University of Milan-Bicocca (protocol number RM-2023-741, as of December 4, 2023).

Statistical Analyses

To evaluate the pre-registered hypothesis concerning the contemporaneous association between positive and negative emotions of the partners, we employed a multilevel modeling approach with random intercepts at both the individual and couple levels (resulting in a three-level structure). We estimated two separate models with PA and NA as the outcome variable, including sex as a moderator. Each affect state was predicted by the opposite valence affect state of the same person (actor effect), and from both affect states of their partner (partner effects) (see Figure 1). We accounted for autocorrelation by adding the lagged outcome variable as a predictor into the model. All predictors were person-mean centered.

To estimate emotional linkage of partners (i.e., the emotional interconnection between lagged observations of partners' affective states), we used a similar multilevel model, predicting PA and NA from lag-1 values of PA and NA of both partners. Finally, we included contact as a moderator in the contemporaneous effects model to assess if emotional

interconnections between partners differed if partners were in contact prior to measurement occasions. As we considered couples as having or not having been in contact only if they both agreed on this (as in Sels et al., 2016), this resulted in including only 70% of measurements in the moderation analysis. This means that those observations, on which partners gave incoherent responses (around 30% of measurements) were excluded from this analysis.

The analyses were conducted using the *nlme* R package (Pinheiro et al., 2021). To mitigate the bias from standard error estimation resulting from omitted random slopes due to convergence issues[5], we employed cluster-robust variance estimation methods from the *clubSandwich* R package (Pustejovsky & Tipton, 2016).

Materials

Affect States in Daily Life

At each prompt, participants were asked to indicate the extent to which they felt cheerful, relaxed, and satisfied, with these items averaged to reflect PA, and insecure, lonely, anxious, irritated, down, and guilty, with these items averaged to reflect NA. Each item was rated on a seven-point Likert scale (from 1 - *not at all* to 7 - *very much*). These adjectives were derived from the Profile of Mood States, a self-report questionnaire for assessing current mood (McNair et al., 1971), and based on psychometric properties of the items revealed in previous studies (Jacobs et al., 2013; Myin-Germeys et al., 2001). CFA confirmed the two-factor structure (PA, NA) of affect states at both the within-person (WP) and between-person (BP) levels. The full results of the CFA analysis can be found in the Appendix. Reliability was estimated for both the within and between components of the measure within a multilevel factor analysis framework (Geldhof et al., 2014) using the *lavaan* (version 0.6-17, Rosseel, 2012) and *semTools* R packages (version 0.5-6., Jorgensen et al., 2012). They were $\alpha_{BP} = 0.95$ and $\alpha_{WP} =$

[5] The convergence issues were related to singular fit for partner effects at the individual level, which means that these effects do not show sufficient variance between the subjects for the random effects to be estimated.

0.75 for PA, and $\alpha_{BP} = 0.94$ and $\alpha_{WP} = 0.75$ for NA. The intraclass correlation coefficient was ICC = 0.43 for PA and ICC = 0.55 for NA.

Contact Between Partners

Each assessment prompt included a question about partner contact since the previous signal. The question read as follows: "Have you interacted with your partner since the last prompt? Indicate all the modalities in which you have interacted with your partner or respond that you have not interacted". The question contained multiple-choice options with the following choices: in person, by phone, via video call, by message (e.g., SMS/WhatsApp/chat or other), another modality, or no interaction. Contact was coded as present whenever partners interacted, irrespective of whether the interaction occurred face-to-face or via mediated channels (e.g., phone, video call, or messaging). This operationalization explains why contact could be reported even during typical work or study hours. Individual participants reported partner contact during 87% of the assessment moments on average. However, a partner agreement occurred only in 70% of the assessment moments (7,575 observations out of 10,896 beeps answered), calculated as the proportion of instances where both partners reported contact or both reported no contact, divided by the total number of responses. In the current study, we consider couples as having or not having been in contact only if they both agree on this, similar to Sels et al. (2016). The majority of observations, when partners agreed (approximately 93%), referred to moments with contact, while only around 7% referred to moments without contact.

Results

Contemporaneous Effects

The contemporaneous models (Table 2) revealed significant actor and partner effects in emotional interdependence between romantic partners. Both for PA and NA, the analysis

demonstrated temporal autocorrelation effects, indicating inertia in affective states over time. Actor effects demonstrated a robust pattern of intra-individual emotional regulation, where actor NA negatively predicted contemporaneous PA, and vice versa. Regarding partner effects, significant positive emotional spillover emerged: partner PA positively predicted actor PA. However, partner NA effects on actor PA did not reach statistical significance, although the direction of the relationship was in the expected direction (negative). In contrast, for NA, the results did not reveal any significant partner effect. Moderation analysis on sex showed that the only significant difference between males and females was in the prediction of their NA by their PA (actor effect): for males, this negative predictive relationship was weaker ($B = 0.08$, SE = 0.028, $t = 3.10$, $p < 0.005$).

In terms of individual variability (Table 4), random intercept estimates revealed that there was more variability between couples than at the individual level (i.e., male or female deviation from the couple level). Additionally, males varied in their levels of PA and NA more than females. Partners within couples demonstrated more similarity (as indicated by the correlation of the random intercepts) in their levels of NA ($\rho = 0.63$) than in their levels of PA ($\rho = 0.37$).

Table 2. *Contemporaneous models.*

	B	**SE**	**t**	***df***	***B***	**SE**	**t**	***df***
Positive Affect								
Female					**Male**			
Intercept (SD, 95% Prediction interval)	**4.46 (0.13, [**3.85-5.13**])**	0.092	48.69***	**75**	**4.49 (0.33, [**4.20-4.71**])**	0.110	40.98***	75
Lag-1 Actor PA	**0.26**	0.022	11.92***	**12288**	**0.31**	0.025	12.36***	**12288**
Actor NA	**-0.74**	0.043	-17.00***	**12288**	**-0.66**	0.041	-16.08***	**12288**
Partner PA	**0.10**	0.016	6.11***	**12288**	**0.09**	0.015	6.44***	**12288**
Partner NA	-0.04	0.022	-1.67	**12288**	-0.01	0.019	-0.36	**12288**
Negative Affect								
Female					**Male**			
Intercept (SD, 95% Prediction interval)	**2.12 (0.17,** 1.79-2.47])	0.090	23.78***	**75**	**2.05 (0.27, [**1.53-2.57])	0.098	20.84***	75
Lag-1Actor NA	**0.32**	0.027	11.94***	**12297**	**0.35**	0.024	14.61***	**12297**
Actor PA	**-0.40**	0.022	-18.51***	**12297**	**-0.32**	0.026	-12.45***	**12297**
Partner NA	0.02	0.019	1.09	**12297**	0.01	0.013	0.93	**12297**
Partner PA	-0.00	0.011	-0.05	**12297**	-0.01	0.011	-1.35	**12297**

Note: 12,447 observations for PA models, 12,456 observations for NA models (the difference is due to missing lagged observations).
The values correspond to simple effects for females and males. Two separate models were estimated for PA and NA with sex as a moderator.
PA, positive affect; NA, negative affect; *B*, Unstandardized regression estimates; SE, standard error.
*p<0.05; **p<0.01; ***p<0.001.

Lagged Effects

The lagged models (Table 3) examined the emotional interconnection between partners across temporally displaced measurements. Significant autocorrelation for both affective states was also observed in this model.

Cross-lagged actor effects demonstrated intra-individual emotional regulation over time: lagged NA negatively predicted PA at the subsequent observation, and lagged PA

negatively predicted NA.

Significantly, a lagged partner effect emerged for PA: the partner's lag-1 PA positively predicted the actor's PA at a subsequent observation, suggesting a persistent inter-partner emotional connection over time. No significant difference was observed between males and females in the lagged model. Random intercept estimates showed a pattern similar to the contemporaneous model (Table 4), with a stronger correlation between male and female random intercepts in NA levels ($\rho = 0.55$) than in PA levels ($\rho = 0.25$).

Table 3. *Lagged models.*

	B	**SE**	**t**	***df***	***B***	**SE**	**t**	**df**
				Positive Affect				
	Female				**Male**			
Intercept (SD, 95% Prediction interval)	**4.47 (0.13, [**4.22-4.73])	0.092	48.46***	**75**	**4.52 (0.34, [**3.86-5.17])	0.111	40.68***	75
Lag-1 Actor PA	**0.41**	0.026	16.14***	**12105**	**0.43**	0.024	17.57***	**12105**
Lag-1 Actor NA	**-0.09**	0.033	-2.89**	**12105**	**-0.15**	0.032	-4.56***	**12105**
Lag-1 Partner PA	**0.05**	0.018	2.60*	**12105**	**0.04**	0.016	2.37*	**12105**
Lag-1 Partner NA	-0.02	0.027	-0.73	**12105**	0.01	0.017	0.55	**12105**
				Negative Affect				
	Female				**Male**			
Intercept (SD, 95% Prediction interval)	**2.12 (0.18, [**1.78-2.47])	0.089	23.95***	**75**	**2.04 (0.27, [**1.50-2.58])	0.098	20.87***	75
Lag-1 Actor NA	**0.47**	0.027	17.14***	**12118**	**0.46**	0.030	15.60***	**12118**
Lag-1 Actor PA	**-0.04**	0.015	-3.09**	**12118**	**-0.07**	0.015	-4.43***	**12118**
Lag-1 Partner NA	0.02	0.016	1.07	**12118**	-0.02	0.013	-1.51	**12118**
Lag-1 Partner PA	-0.01	0.013	-0.73	**12118**	-0.02	0.012	-1.68	**12118**

Note: 12,264 observations for PA models, 12,277 observations for NA models.

PA, positive affect; NA, negative affect; *B*, Unstandardized regression estimates; SE, standard error.
The values correspond to simple effects for females and males. Two separate models were estimated for PA and NA with sex as a moderator.

*p<0.05; **p<0.01; ***p<0.001.

The Role of Contact between Partners in Contemporaneous Emotional Interconnection

The contact moderation analysis (Table 4) revealed substantial differences in emotional effects based on the presence of partner interaction.

As shown in Table 4, the positive contact effect on PA was significant for males, but not for females, while no significant effect was found on NA. On the other hand, contact did not show any significant interaction effect for either PA or NA. Table 4 includes the estimates from a two-intercept model. Thus, significance is specified separately for each estimate, but it is not possible to directly determine whether two coefficients differ significantly. As a consequence, the differences between the coefficients for interactions between contact and PA or NA were tested in a reparametrized (single-intercept) model. The results showed that these differences did not reach statistical significance (see Table A1 in Appendix). For both affect types, actor effects were similarly intense with and without contact with one's partner. Actor PA was significantly predicted by the partner's PA in the contact condition for both males and females, but remained significant without contact only for males (see Table 4). For NA, no difference was found between contact and no-contact conditions. The random structure estimates were almost identical to those of the original contemporaneous model (Table 4), with a stronger correlation between male and female random intercepts in NA levels ($\rho = 0.60$) than in PA levels ($\rho = 0.35$).

Table 4. *Contemporaneous models with contact between partners as a moderator.*

	B	**SE**	**t**	***df***	***B***	**SE**	**t**	**df**
					Positive Affect			
	Female					**Male**		
Intercept (SD, 95% Prediction interval)	**4.42 (0.14,** [4.15-4.69])	0.105	41.95***	**75**	**4.35 (0.33,** [3.71-4.99])	0.125	34.82***	75
Lag-1 Actor PA	**0.26**	0.022	11.98***	**10923**	**0.31**	0.027	11.50***	**10923**
Contact	0.04	0.054	0.82	**10923**	**0.16**	0.056	2.89**	**10923**
Actor NA with contact	**-0.73**	0.049	-15.11***	**10923**	**-0.67**	0.045	-14.75***	**10923**
Actor NA w/o contact	**-0.74**	0.056	-13.19***	**10923**	**-0.65**	0.096	-6.75***	**10923**
Partner PA with contact	**0.11**	0.018	6.29***	**10923**	**0.10**	0.017	5.75***	**10923**
Partner PA w/o contact	0.08	0.055	1.39	**10923**	**0.16**	0.039	4.04**	**10923**
Partner NA with contact	-0.03	0.023	-1.46	**10923**	-0.01	0.019	-0.66	**10923**
Partner NA w/o contact	0.05	0.106	0.49	**10923**	**0.14**	0.056	2.44*	**10923**
					Negative Affect			
	Female					**Male**		
Intercept (SD, 95% Prediction interval)	**2.14 (0.17,** [1.80-2.48])	0.105	20.37***	**75**	**2.02 (0.26,** [1.51-2.53])	0.103	19.74***	75
Lag-1Actor NA	**0.33**	0.027	12.34***	**10932**	**0.35**	0.026	13.38***	**10932**
Contact	-0.01	0.051	-0.12	**10932**	0.02	0.032	0.74	**10932**
Actor PA with contact	**-0.40**	0.019	-20.76***	**10932**	**-0.32**	0.028	-11.42***	**10932**
Actor PA w/o contact	**-0.41**	0.073	-5.67***	**10932**	**-0.43**	0.038	-11.28***	**10932**
Partner NA with contact	0.03	0.020	1.37	**10932**	0.02	0.013	1.21	**10932**
Partner NA w/o contact	0.10	0.073	1.32	**10932**	0.09	0.060	1.47	**10932**
Partner PA with contact	0.00	0.012	0.03	**10932**	-0.01	0.011	-1.26	**10932**
Partner PA w/o contact	0.07	0.045	1.56	**10932**	0.06	0.051	1.26	**10932**

Note: 11,152 observations for PA models: 10,341 - with contact, 749 - without contact; 11,099 observations for NA models: 10,351 - with contact, 748 - without contact.

PA, positive affect; NA, negative affect; *B*, Unstandardized regression estimates; SE, standard error.

*p<0.05; **p<0.01; ***p<0.001

Discussion

The present study investigated emotional interdependence in romantic relationships using ESM methodology to examine both contemporaneous and temporal patterns of emotional interdependence between partners, while testing the distinguishability by sex and the moderating role of contact. Our findings provide compelling evidence that partners' emotional dynamics are indeed interconnected in daily life, with significant associations emerging both within the same time points and across temporal lags. These results contribute significantly to our understanding of the mechanisms underlying interpersonal emotion dynamics in close relationships and address ongoing debates about the prevalence and nature of emotional interdependence in romantic couples.

The present study builds upon and extends the important foundational work of Sels et al. (2020), who systematically examined emotional interdependence in romantic couples across multiple timescales and contexts. While Sels et al. provided crucial evidence that emotional connections between partners may be less pervasive than theoretical models have traditionally assumed—with significant interdependence observed in a minority of couples depending on the measure and context—the present investigation contributes several methodological refinements that allow for more comprehensive assessment of daily emotional dynamics. First, our 30-day ESM protocol provides extended temporal coverage, allowing us to capture emotional patterns across a wider range of naturalistic circumstances, life events, and relationship contexts that unfold over longer timeframes, while also increasing statistical power to detect partner effects that may be modest in magnitude but meaningful in cumulative impact. In addition, testing of the role of contact in our study was conducted more naturally—sampling participants five times per day and accounting for different types of contact, rather than sampling mainly when the couples were expected to be together, as in the previous study. In this way, we aimed at achieving more ecologically valid results.

Evidence for Emotional Interdependence

The central finding of this study is the evidence for emotional interdependence between romantic partners operating through multiple temporal pathways. Partners' PA levels demonstrated significant concurrent associations, such that when one partner experienced elevated PA at a given moment, the other partner tended to experience higher PA as well. This contemporaneous interdependence for positive emotions aligns with some previous research documenting emotional covariation in couples (Butner et al., 2007; Saxbe & Repetti, 2010), though the effect sizes we observed were modest.

Importantly, positive emotional interdependence extended beyond concurrent associations to show temporal persistence. One partner's PA at a previous time point significantly predicted the other partner's PA at a subsequent observation, even after controlling for autoregressive effects. Though these lagged partner effects were small, the existence of this temporal linkage provides evidence that partners serve as external regulators of each other's emotional trajectories and that emotional influences in romantic relationships have lasting effects that extend beyond the moment of interaction (Butler & Randall, 2013).

We observed only limited differences between males and females in their emotional dynamics. Only the actor effect of PA on NA in the contemporaneous models was significantly different between sexes. From this, we can conclude that the dyads are partially distinguishable in the context of our study. However, a more in-depth test of indistinguishability would be needed in further studies (Gistelink et al., 2018). It is particularly interesting that Sels et al. (2016) found differences in partner effects between males and females. These discrepancies between our findings could be potentially explained by the idiographic approach of their study, as they analyzed all couples separately, while we used a nomothetic approach in our work, and thus, we might lose some information from couple-specific differences.

For NA, we found no significant partner effects either contemporaneously or

temporally. This absence of negative emotional interdependence contrasts with some earlier literature emphasizing negative emotional transmission in romantic relationships (Larson & Almeida, 1999; Thompson & Bolger, 1999), but is more in line with recent studies' findings of opposite effects (Randall & Schoebi, 2015) or heterogeneity in emotional linkage patterns (Madhyastha et al., 2011; Sels et al., 2016). These null findings, however, require cautious interpretation, as they could reflect not only a genuine absence of negative emotional interdependence in our sample, but also insufficient statistical power to detect small effects, measurement limitations, or artifacts due to considerable skewness of the NA distribution, which may lead to violations of assumptions of a linear model used in the analysis (Dejonckheere et al., 2024; Mestdagh et al., 2018).

The difference between PA and NA patterns found in the current work requires, though, careful interpretation. As PA and NA were assessed within separate models, the results do not permit any conclusions about whether the observed differences are significant. The valence difference may reflect purely methodological artifacts. Our NA composite included six items (insecure, lonely, anxious, irritated, down, guilty) compared to three items for PA (cheerful, relaxed, satisfied), and these different emotions may have distinct interpersonal dynamics. Research has shown that different discrete emotions have distinguishable interpersonal dynamics, which does not necessarily represent pure contagion (Bar-Kalifa & Sened, 2020). For instance, sadness may elicit support-seeking and caregiving responses (Luginbuehl & Schoebi, 2020), while guilt may prompt positive feelings in one’s partner (Overall et al., 2014). Our composite approach may have averaged across these distinct patterns, potentially obscuring emotion-specific effects.

If, however, the differences between PA and NA interpersonal dynamics are indeed significant, this asymmetry could reflect several processes that cannot be examined within the current study and would require further investigation. For instance, one possibility is that

couples had developed effective strategies for containing negative emotional transmission while remaining open to positive emotional influence. Butler and Randall (2013) proposed that healthy coregulation involves dampening negative emotional oscillations while maintaining connection, which could manifest as the pattern we observed. Additionally, previous studies have linked increase of positive emotions in couples with increased intimacy, which is an important factor of satisfying romantic relationships (Laurenceau et al., 2005). At the same time, negative emotions were associated with couple conflicts. If this is the case, characteristics of the present sample may have contributed to the observed asymmetry. Because participation in the study required both partners to volunteer for an intensive ESM protocol, the sample likely consisted predominantly of relatively well-functioning couples, potentially restricting variability and synchronization in conflict-related emotional exchanges. Future research could examine this interpretation more directly by including more heterogeneous samples (e.g., by oversampling distressed couples) or by combining ESM with event-contingent reports of disagreements. Such designs would allow testing whether negative affect interdependence becomes more pronounced for less well-functioning couples and in conditions of relational strain. Valence effects may also be highly context-dependent and moderated by factors such as the attachment styles of partners (Butner et al., 2007), stress levels (Neff & Karney, 2007), or cultural norms (Schoebi et al., 2010).

Our analyses highlight the important role of individual emotional processes in the overall couple's emotional dynamics. The autoregressive effects for both PA and NA were noticeable and larger than any partner effects we observed. This pattern indicates substantial emotional inertia, where individuals' emotional states show strong continuity across time (Houben et al., 2015; Kuppens et al., 2010). Importantly, not only autoregressive effects, but also cross-valence actor effects, were observed, where higher NA predicted lower PA, further highlighting the primacy of individual regulatory processes. What is even more interesting is

that this was the only effect where findings supported differences between sexes: for males, their regulatory effect of PA on NA was weaker than for females. This result could be related to the more extensive general engagement of women in emotion regulation strategies compared to men (Nolen-Hoeksema, 2012).

The substantial variability revealed in random intercept estimates provides important insights into the individual differences that average effects may obscure. Between-couple variance in intercepts exceeded individual-level variance for both PA and NA, indicating that couples differ more from each other than participants differ from each other in the extent to which they deviate from their couples' means. This pattern suggests that couples develop shared emotional baselines, consistent with research on emotional convergence, showing that partners become more similar over time (Anderson et al., 2003).

However, the presence of substantial between-couple variability also implies that the average partner effects may not characterize all couples equally. Some couples may show much stronger emotional interdependence than others, while some may show no interdependence or even negative associations. This interpretation is supported by the studies, which emphasize that emotional interdependence is heterogeneous across couples (Madhyastha et al., 2011; Sels et al., 2016). To quantify the heterogeneity in partner effects, a model with random slopes should be estimated; therefore, a further study including such a model is needed. However, the large between-couple variance in intercepts suggests the heterogeneity in effects likely exists as well.

The correlation between male and female random intercepts was moderate for NA ($r = 0.63$) but weaker for PA ($r = 0.37$), suggesting that partners are more aligned in their baseline NA than PA. These results corroborate the findings of the previous cross-sectional work, which showed that romantic partners were more similar to each other in their NA level than in their PA level (Orth et al., 2018). These findings, though, must be interpreted with caution due to

the noticeable skewness of the NA observations. Moreover, contextual factors, which could be potentially important for interpreting these results, were not taken into account in the current study.

Males showed greater variability in both PA and NA compared to females within our models. It should be noted that this variability is interindividual, while most existing findings in the literature refer to the intraindividual variability (e.g., Charles & Pasupathi, 2003), thus, making them not comparable to each other. Few studies report on interindividual variability; for example, the one by Thomsen et al. (2005) observed larger variability in baseline affect in young females rather than males. However, their results are based on a single questionnaire, while ours correspond to the average of multiple observations across a month of study.

Effects of Contact

Our investigation of contact as a moderator of couple emotional dynamics revealed a pattern that did not support our initial hypotheses. Contact between partners showed a significant positive main effect on males' PA but not females', suggesting that being in contact with one's romantic partner enhances positive emotional experience for men but not necessarily for women. This difference between sexes may reflect different relational needs or different meanings of contact for men versus women, though our data cannot address the mechanisms underlying this difference. In contrast to our expectations, however, contact did not significantly moderate the strength of emotional interdependence. This absence of moderation is particularly noteworthy, given that previous works have emphasized that emotional coregulation should depend on opportunities for interaction (Butler & Randall, 2013; Sbarra & Hazan, 2008). Butner et al. (2007) found that emotional covariation was higher on days when partners spent more time together, and Papp et al. (2013) found stronger physiological linkage during periods of greater proximity. However, these studies employed different operationalizations of contact (e.g., time spent together, physical proximity) and varied time

scales, making direct comparisons challenging. This discrepancy of our findings from previous research could be also attributed to the substantial disbalance between the number of observations that included contact (more than 10,000), and those without contact (less than 1,000), as well as to the exclusion of almost a third of observations from the moderation analysis due to the incongruence between partners' responses about the contact. Moreover, the binary measure of contact may not have been able to capture all relevant distinctions. Differentiating between positive and negative interactions, and between different modalities of contact (in-person, phone, text) could give a more comprehensive view of the role of contact, although relying on smaller sample sizes for each specific type of contact. Our findings suggest that the moderating role of contact may be more complex or context-dependent than a simple presence versus absence.

Limitations, Future Directions, and Implications

This study employed several methodological approaches that represent advances for dyadic emotion dynamics research while also facing several limitations. Our extended APIM framework simultaneously examined bidirectional partner effects while controlling for autoregressive processes and cross-valence actor effects, providing a more comprehensive model than many previous studies. The three-level random intercept structure appropriately modeled dependencies at the individual and couple levels, though our inability to estimate random slopes (due to convergence failures) limited our capacity to examine between-couple heterogeneity in partner effects. Future research with larger samples of couples might achieve convergence with random slopes. An alternative solution could be a re-analysis of the study data with different software, such as SAS (SAS Institute Inc.), which offers certain advantages in fitting a random-effects model compared to R (West & Galecki, 2011).

Certain characteristics of our sample, namely, young age, high level of education, predominance of couples without children, and heterosexual relationships, make our findings

less generalizable to other populations. However, research across diverse populations is essential. Future studies, which include couples of diverse ages, socio-demographic statuses, sexual orientations, and numbers of children, will need to be undertaken. Cross-cultural research would be particularly valuable, given evidence that cultural norms shape emotional processes (Boiger & Mesquita, 2012).

Additional factors not controlled in our study include the partners' personalities and the context of their daily lives, both of which play a non-negligible role in romantic relationships and alignment between partners (Neff & Karney, 2017; Van Scheppingen, 2025). Shared or separate activities, environmental stressors, working hours, financial issues, or other factors could produce apparent emotional interdependence or mask the existing one, regardless of a genuine partner effect. For instance, if couples typically engage in pleasant activities together in the evening, both partners' elevated PA at evening assessments might reflect these shared activities rather than emotional transmission per se. On the other hand, similar personalities of partners could also be related to similar emotional reactions to shared context. A natural progression of the current work would thus be to analyze the emotional dynamics of couples, considering their personality traits and the contextual factors of their everyday lives.

Overall, the evidence from this study suggests that romantic partners' emotional experiences in daily life are not independent of each other. This is supported mainly by PA interdependence. Second, the absence of NA interdependence in our sample suggests that not all emotional states transmit equally between partners, and that couples may vary considerably in whether and how different emotions are shared. The difference between couples is also highlighted by a considerable variability of random effects. Such heterogeneity implies that one-size-fits-all intervention approaches in couples and family therapy may be inappropriate, and that understanding individual couples' specific patterns of emotional interdependence may be necessary for effective interventions.

Conclusions

This study provides evidence that romantic partners' emotions are interconnected in daily life, though this interdependence is more limited and complex than some theoretical models suggest. We found reliable evidence of emotional interdependence for positive affect, operating both contemporaneously and across temporal lags, indicating that when one partner experiences positive emotions, the other partner tends to experience similar positive states both simultaneously and several hours later. This pattern supports theories of emotional contagion and interpersonal emotion regulation in close relationships (Butler & Randall, 2013; Herrando & Constantinides, 2021).

However, several findings moderate conclusions about the strength and universality of emotional interdependence. First, partner effects were modest in magnitude compared to WP processes. Individuals' own prior emotional states and internal emotional dynamics were substantially stronger predictors of their subsequent emotions than their partners' emotions, consistent with recent research emphasizing the primacy of intrapersonal emotion dynamics (Houben et al., 2015; Kuppens et al., 2010). Second, we found no significant interdependence for negative affect, which contrasts with some earlier literature but aligns with recent studies revealing heterogeneity in emotional linkage patterns (Sels et al., 2016). Third, contact between partners enhanced positive affect in male partners but did not moderate interdependence strength in couples, suggesting that emotional connection may operate through mechanisms beyond the mere fact of being in contact.

The substantial between-couple variability observed in our random effects indicates that emotional interdependence is not uniform across relationships. This heterogeneity supports growing recognition that couples exhibit diverse patterns of emotional dynamics (Madhyastha et al., 2011; Sels et al., 2016) and suggests that nomothetic approaches may obscure important individual differences. Future research should employ approaches that can model this

heterogeneity, examine discrete emotions rather than broad composites, incorporate richer measures of contact and context, and test these processes across diverse populations. Understanding the conditions under which emotional interdependence occurs and its implications for relationship functioning remains an important priority for relationship science and the development of couple and family interventions.

Appendix

Table A1. *Contemporaneous models with contact between partners as a moderator.*

	B	**SE**	**t**	***p*-value**
	Positive Affect			
Intercept	**4.37**	0.100	43.50***	<0.001
Sex	0.03	0.101	0.34	0.74
Lag-1 Actor PA	**0.28**	0.019	15.09***	<0.001
Contact	**0.10**	0.041	2.44*	0.02
Actor NA	**-0.70**	0.064	-11.01***	<0.001
Partner NA	0.09	0.070	1.30	0.21
Partner PA	**0.11**	0.046	2.51*	0.02
Actor NA:Contact	-0.01	0.067	-0.10	0.92
Partner NA:Contact	-0.11	0.070	-1.60	0.13
Partner PA: Contact	-0.01	0.048	-0.22	0.83
	Negative Affect			
Intercept	**2.12**	0.096	22.08***	<0.001
Sex	-0.09	0.097	-0.88	0.38
Lag-1Actor NA	**0.34**	0.018	19.34***	<0.001
Contact	0.01	0.035	0.27	0.79
Actor PA	**-0.42**	0.044	-9.44***	<0.001
Partner PA	0.07	0.040	1.66	0.12
Partner NA	0.09	0.062	1.49	0.16
Actor PA:Contact	0.06	0.043	1.32	0.20
Partner PA:Contact	-0.07	0.040	-1.82	0.09
Partner NA: Contact	-0.07	0.061	-1.16	0.26

Note: 11,152 observations for PA models, 11,099 observations for NA models.
PA, positive affect; NA, negative affect; *B*, Unstandardized regression estimates; SE, standard error.
*p<0.05; **p<0.01; ***p<0.001
Standard errors were corrected with cluster-robust variance estimation method from the clubSandwich R package (Pustejovsky & Tipton, 2016).

References:

Anderson, C., Keltner, D., & John, O. P. (2003). Emotional convergence between people over time. *Journal of Personality and Social Psychology*, *84*(5), 1054–1068. https://doi.org/10.1037/0022-3514.84.5.1054

Bar-Kalifa, E., & Sened, H. (2020). Using Network Analysis for Examining Interpersonal Emotion Dynamics. *Multivariate Behavioral Research*, *55*(2), 211–230. https://doi.org/10.1080/00273171.2019.1624147

Berscheid, E., & Ammazzalorso, H. (2001). Emotional Experience in Close Relationships. In G. J. O. Fletcher & M. S. Clark (Eds.), *Interpersonal processes*. Blackwell Publishers.

Boiger, M., & Mesquita, B. (2012). The Construction of Emotion in Interactions, Relationships, and Cultures. *Emotion Review*, *4*(3), 221–229. https://doi.org/10.1177/1754073912439765

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary Methods: Capturing Life as it is Lived. *Annual Review of Psychology*, *54*(1), 579–616. https://doi.org/10.1146/annurev.psych.54.101601.145030

Butler, E. A. (2011). Temporal Interpersonal Emotion Systems: The "TIES" That Form Relationships. *Personality and Social Psychology Review*, *15*(4), 367–393. https://doi.org/10.1177/1088868311411164

Butler, E. A. (2015). Interpersonal Affect Dynamics: It Takes Two (and Time) to Tango. *Emotion Review*, *7*(4), 336–341. https://doi.org/10.1177/1754073915590622

Butler, E. A., & Randall, A. K. (2013). Emotional Coregulation in Close Relationships. *Emotion Review*, *5*(2), 202–210. https://doi.org/10.1177/1754073912451630

Butner, J., Diamond, L. M., & Hicks, A. M. (2007). Attachment style and two forms of affect coregulation between romantic partners. *Personal Relationships*, *14*(3), 431–455. https://doi.org/10.1111/j.1475-6811.2007.00164.x

Charles, S. T., & Pasupathi, M. (2003). Age-related patterns of variability in self-descriptions: Implications for everyday affective experience. *Psychology and Aging*, *18*(3), 524–536. https://doi.org/10.1037/0882-7974.18.3.524

Dejonckheere, E., Penne, I., Briels, L., & Mestdagh, M. (2024). For Better or for Worse? Visualizing Previous Intensity Levels Improves Emotion (Dynamic) Measurement in Experience Sampling. *Psychological Assessment*, *36*(3), 215–234. https://doi.org/10.1037/pas0001296

Ferrer, E., & Widaman, K. F. (2011). Dynamic Factor Analysis of Dyadic Affective Processes With Intergroup Differences. In N. A. Card, J. P. Selig, & T. Little (Eds.), *Modeling Dyadic and Interdependent Data in the Developmental and Behavioral Sciences* (0 ed., pp. 117–148). Routledge. https://doi.org/10.4324/9780203892763-10

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72–91. https://doi.org/10.1037/a0032138

Gonzaga, G. C., Campos, B., & Bradbury, T. (2007). Similarity, convergence, and relationship satisfaction in dating and married couples. *Journal of Personality and Social Psychology*, *93*(1), 34–48. https://doi.org/10.1037/0022-3514.93.1.34

Gonzalez, R., & Griffin, D. (1999). The correlational analysis of dyad-level data in the distinguishable case. Personal Relationships, 6(4), 449–469. https://doi.org/10.1111/j.1475-6811.1999.tb00203.x

Gottman, J. M., & Levenson, R. W. (2002). A Two-Factor Model for Predicting When a Couple Will Divorce: Exploratory Analyses Using 14-Year Longitudinal Data. *Family Process*, *41*(1), 83–96. https://doi.org/10.1111/j.1545-5300.2002.40102000083.x

Herrando, C., & Constantinides, E. (2021). Emotional Contagion: A Brief Overview and Future Directions. *Frontiers in Psychology*, *12*, 712606. https://doi.org/10.3389/fpsyg.2021.712606

Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin*, *141*(4), 901–930. https://doi.org/10.1037/a0038822

Iida, M., Savord, A., & Ledermann, T. (2023). Dyadic longitudinal models: A critical review. *Personal Relationships*, *30*(2), 356–378. https://doi.org/10.1111/pere.12468

Jacobs, N., Menne-Lothmann, C., Derom, C., Thiery, E., Van Os, J., & Wichers, M. (2013). Deconstructing the familiality of variability in momentary negative and positive affect. *Acta Psychiatrica Scandinavica*, *127*(4), 318–327. https://doi.org/10.1111/j.1600-0447.2012.01924.x

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2012). *semTools: Useful Tools for Structural Equation Modeling* (p. 0.5-7) [Dataset]. https://doi.org/10.32614/CRAN.package.semTools

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic Data Analysis*. Guilford Publications.

Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional Inertia and Psychological Maladjustment. *Psychological Science*, *21*(7), 984–991. https://doi.org/10.1177/0956797610372634

Larson, R. W., & Almeida, D. M. (1999). Emotional Transmission in the Daily Lives of Families: A New Paradigm for Studying Family Process. *Journal of Marriage and Family*, *61*(1), 5–20. https://doi.org/10.2307/353879

Laurenceau, J.-P., & Bolger, N. (2005). Using Diary Methods to Study Marital and Family Processes. *Journal of Family Psychology*, *19*(1), 86–97. https://doi.org/10.1037/0893-3200.19.1.86

Laurenceau, J.-P., Troy, A. B., & Carver, C. S. (2005). Two Distinct Emotional Experiences in Romantic Relationships: Effects of Perceptions Regarding Approach of Intimacy and Avoidance of Conflict. *Personality and Social Psychology Bulletin*, *31*(8), 1123–1133. https://doi.org/10.1177/0146167205274447

Ledermann, T., & Kenny, D. A. (2017). Analyzing dyadic data with multilevel modeling versus structural equation modeling: A tale of two methods. *Journal of Family Psychology*, *31*(4), 442–452. https://doi.org/10.1037/fam0000290

Liu, D. Y., Strube, M. J., & Thompson, R. J. (2021). Interpersonal Emotion Regulation: An Experience Sampling Study. *Affective Science*, *2*(3), 273–288. https://doi.org/10.1007/s42761-021-00044-y

Luginbuehl, T., & Schoebi, D. (2020). Emotion dynamics and responsiveness in intimate relationships. *Emotion*, *20*(2), 133–148. https://doi.org/10.1037/emo0000540

Madhyastha, T. M., Hamaker, E. L., & Gottman, J. M. (2011). Investigating spousal influence using moment-to-moment affect data from marital conflict. *Journal of Family Psychology*, *25*(2), 292–300. https://doi.org/10.1037/a0023028

McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Manual: Profile of Mood States, Educational and Industrial Testing Service*.

Mestdagh, M., Pe, M., Pestman, W., Verdonck, S., Kuppens, P., & Tuerlinckx, F. (2018). Sidelining the mean: The relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychological Methods*, *23*(4), 690–707. https://doi.org/10.1037/met0000153

Myin-Germeys, I., Van Os, J., Schwartz, J. E., Stone, A. A., & Delespaul, P. A. (2001). Emotional Reactivity to Daily Life Stress in Psychosis. *Archives of General Psychiatry*, *58*(12), 1137. https://doi.org/10.1001/archpsyc.58.12.1137

Neff, L. A., & Karney, B. R. (2007). Stress Crossover in Newlywed Marriage: A Longitudinal and Dyadic Perspective. *Journal of Marriage and Family*, *69*(3), 594–607. https://doi.org/10.1111/j.1741-3737.2007.00394.x

Neff, L. A., & Karney, B. R. (2017). Acknowledging the elephant in the room: How stressful environmental contexts shape relationship dynamics. *Current Opinion in Psychology*, *13*, 107–110. https://doi.org/10.1016/j.copsyc.2016.05.013

Nolen-Hoeksema, S. (2012). Emotion Regulation and Psychopathology: The Role of Gender. *Annual Review of Clinical Psychology*, *8*(1), 161–187. https://doi.org/10.1146/annurev-clinpsy-032511-143109

O'Brien, S. T., Dozo, N., Hinton, J. D. X., Moeck, E. K., Susanto, R., Jayaputera, G. T., Sinnott, R. O., Vu, D., Alvarez-Jimenez, M., Gleeson, J., & Koval, P. (2024). SEMA3: A free smartphone platform for daily life surveys. *Behavior Research Methods*, *56*(7), 7691–7706. https://doi.org/10.3758/s13428-024-02445-w

Olsen, J. A., & Kenny, D. A. (2006). Structural equation modeling with interchangeable dyads. Psychological Methods, 11(2), 127–141. https://doi.org/10.1037/1082-989X.11.2.127

Orth, U., Erol, R. Y., Ledermann, T., & Grob, A. (2018). Codevelopment of well-being and self-esteem in romantic partners: Disentangling the effects of mutual influence and shared environment. *Developmental Psychology*, *54*(1), 151–166. https://doi.org/10.1037/dev0000400

Overall, N. C., Girme, Y. U., Lemay, E. P., & Hammond, M. D. (2014). Attachment anxiety and reactions to relationship threat: The benefits and costs of inducing guilt in romantic partners. *Journal of Personality and Social Psychology*, *106*(2), 235–256. https://doi.org/10.1037/a0034371

Papp, L. M., Pendry, P., Simon, C. D., & Adam, E. K. (2013). Spouses' Cortisol Associations and Moderators: Testing Physiological Synchrony and Connectedness in Everyday Life. *Family Process*, *52*(2), 284–298. https://doi.org/10.1111/j.1545-5300.2012.01413.x

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2021). *nlme: Linear and Nonlinear Mixed Effects Models.* [Computer software]. https://CRAN.R-project.org/package=nlme

Pustejovsky, J. E., & Tipton, E. (2016). Small sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, *36*(4), 672–683. https://doi.org/10.1080/07350015.2016.1247004

Randall, A. K., Post, J. H., Reed, R. G., & Butler, E. A. (2013). Cooperating with your romantic partner: Associations with interpersonal emotion coordination. *Journal of Social and Personal Relationships*, *30*(8), 1072–1095. https://doi.org/10.1177/0265407513481864

Randall, A. K., & Schoebi, D. (2015). Lean on me: Susceptibility to partner affect attenuates psychological distress over a 12-month period. *Emotion*, *15*(2), 201–210. https://doi.org/10.1037/emo0000043

Rauers, A., & Riediger, M. (2023). Ease of Mind or Ties That Bind? Costs and Benefits of Disclosing Daily Hassles in Partnerships. *Social Psychological and Personality Science*, *14*(5), 551–561. https://doi.org/10.1177/19485506221112252

Rosseel, Y. (2012). lavaan: An *R* Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Salo, K. I., Pauw, L. S., Echterhoff, G., & Milek, A. (2022). Daily stress, closeness, and coping in romantic relationships during COVID-19-related lockdown: An experience-sampling study. *Couple and Family Psychology: Research and Practice*, *11*(3), 217–231. https://doi.org/10.1037/cfp0000228

Saxbe, D., & Repetti, R. L. (2010). For better or worse? Coregulation of couples' cortisol levels and mood states. *Journal of Personality and Social Psychology*, *98*(1), 92–103. https://doi.org/10.1037/a0016959

Sbarra, D. A., & Hazan, C. (2008). Coregulation, Dysregulation, Self-Regulation: An Integrative Analysis and Empirical Agenda for Understanding Adult Attachment, Separation, Loss, and Recovery. *Personality and Social Psychology Review*, *12*(2), 141–167. https://doi.org/10.1177/1088868308315702

Schoebi, D. (2008). The coregulation of daily affect in marital relationships. *Journal of Family Psychology*, *22*(4), 595–604. https://doi.org/10.1037/0893-3200.22.3.595

Schoebi, D., Wang, Z., Ababkov, V., & Perrez, M. (2010). Affective interdependence in married couples' daily lives: Are there cultural differences in partner effects of anger? *Family Science*, *1*(2), 83–92. https://doi.org/10.1080/19424620903471681

Sels, L., Cabrieto, J., Butler, E., Reis, H., Ceulemans, E., & Kuppens, P. (2020). The occurrence and correlates of emotional interdependence in romantic relationships. *Journal of Personality and Social Psychology*, *119*(1), 136–158. https://doi.org/10.1037/pspi0000212

Sels, L., Ceulemans, E., Bulteel, K., & Kuppens, P. (2016). Emotional Interdependence and

Well-Being in Close Relationships. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00283

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological Momentary Assessment. *Annual Review of Clinical Psychology*, *4*(1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Steele, J. S., Ferrer, E., & Nesselroade, J. R. (2014). An Idiographic Approach to Estimating Models of Dyadic Interactions with Differential Equations. *Psychometrika*, *79*(4), 675–700. https://doi.org/10.1007/s11336-013-9366-9

Thompson, A., & Bolger, N. (1999). Emotional Transmission in Couples Under Stress. *Journal of Marriage and the Family*, *61*(1), 38. https://doi.org/10.2307/353881

Thomsen, D. K., Mehlsen, M. Y., Viidik, A., Sommerlund, B., & Zachariae, R. (2005). Age and gender differences in negative affect—Is there a role for emotion regulation? *Personality and Individual Differences*, *38*(8), 1935–1946. https://doi.org/10.1016/j.paid.2004.12.001

Trull, T. J., & Ebner-Priemer, U. (2014). The Role of Ambulatory Assessment in Psychological Science. *Current Directions in Psychological Science*, *23*(6), 466–470. https://doi.org/10.1177/0963721414550706

Van Scheppingen, M. A., Olaru, G., & Leopold, T. (2025). Personality trait similarity in recently cohabiting couples: Partner choice, convergence, or selective breakup? *Journal of Personality and Social Psychology*, *128*(4), 887–904. https://doi.org/10.1037/pspp0000527

West, B. T., & Galecki, A. T. (2011). An Overview of Current Software Procedures for Fitting Linear Mixed Models. *The American Statistician*, *65*(4), 274–282. https://doi.org/10.1198/tas.2011.11077

Appendix. *Model fit and comparisons*

Model	CFI	RMSEA	SRMR	SRMR WP	df	AIC	BIC	χ^2	*p*	Diff. χ^2	Df diff.	*p* diff.
Model 1: Single factor (Affect) at WP and BP levels	0.896	0.071	0.163	0.050	54	502119	502473	5320.764	<0.001	2375.3	2	<0.001
Model 2: Two factors (PA and NA) at WP and BP levels	**0.943**	**0.053**	**0.076**	**0.038**	**52**	**499747**	**500118**	**2945.420**	**<0.001**			**<0.001**
Model 3: Single factor (Affect) at WP level and two factors (PA and NA) at BP level	0.902	0.069	0.088	0.050	53	501827	502189	5027.068	<0.001	2081.7	1	<0.001
Model 4: Two factors (PA and NA) at WP level and one factor (Affect) at BP level	0.937	0.056	0.150	0.038	53	500037	500399	3236.698	<0.001	291.28	1	<0.001

Note: Model comparisons were conducted with respect to Model 2 fit at baseline. BP – between-person; CFI - Comparative Fit Index; RMSEA - Root Mean Square Error of Approximation; SRMR – Root Mean Square Residual; WP – within-person.

Part III

6

Applying Mixed-Effects Tobit Models to Right-Skewed ESM/EMA Data

Galkina, A., Viechtbauer, W. (2025), Applying Mixed-Effects Tobit Models to Right-Skewed ESM/EMA Data. (Under review).

Abstract

Intensive longitudinal data collection methods like the Experience Sampling Method (ESM) and Ecological Momentary Assessment (EMA) are used to repeatedly assess people's emotional states, behaviors, and context to examine their associations. Many variables of interest, such as negative affect and psychopathological symptoms, are measured with bounded scales that can produce right-skewed data. Standard linear mixed-effects models do not account for this intricacy, potentially producing biased parameter estimates.

This study proposes to conceptualize skewed observations as censored manifestations of underlying latent variables, using mixed-effects tobit models as an alternative analytical approach. Using data from N = 322 participants from three mental health groups (controls, depressed, psychotic) who completed mood assessments ten times daily for six days, we examined parameter recovery capabilities of the standard linear and tobit mixed-effects models by inducing increasing skewness into an originally roughly normally distributed outcome variable (positive affect). Results indicated that the tobit model provided relatively stable parameter estimates under progressive censoring conditions, while the standard linear model showed deterioration in slope estimates approaching zero.

When applied to a naturally skewed outcome variable (negative affect), the tobit model again estimated larger associations across all groups, with the greatest differences observed in controls, where floor effects were most pronounced. Tobit models also revealed greater between-person variability and reduced intercept-slope correlations.

While computational complexity presents implementation challenges, these findings suggest that tobit models offer important advantages for analyzing bounded psychological constructs in ESM research. However, further research across diverse contexts is needed to establish implementation guidelines.

Introduction

Intensive longitudinal data collection methods like the Experience Sampling Method (ESM) (also called Ecological Momentary Assessment; EMA) have gained significant popularity in psychological research in recent years (Bolger & Laurenceau, 2013; Hamaker & Wichers, 2017; Trull & Ebner-Priemer, 2014). These methods enable researchers to collect data about variables of interest in everyday life through brief, frequent assessments that participants typically complete on their smartphones throughout the day. ESM serves as an effective method for capturing participants' feelings, behaviors, experiences, and life circumstances with strong ecological validity while minimizing recall bias (Lucas et al., 2021).

Building upon its capacity to capture temporal fluctuations, ESM is frequently used to examine psychological phenomena characterized by dynamic properties, including emotional processes (e.g., Tammilehto et al., 2022; Trampe et al., 2015), affective states (e.g., Schoevers et al., 2021), and symptoms associated with psychopathological conditions (e.g., Telford et al., 2012; R. Wang et al., 2016). However, one of the major issues with certain types of data, such as symptom manifestations, negative affect, or other emotional states, is that they can take on different distributional forms and are often right-skewed (Haslbeck et al., 2023). This is usually referred to in the literature as a floor effect (Mestdagh & Dejonckheere, 2021; Siepe et al., 2025), while its opposite, the ceiling effect, though less common, is also sometimes observed (L. Wang et al., 2008).

Potential explanations for data skewness in ESM studies include scale boundary constraints (Dejonckheere et al., 2024), item formulation considerations (Mestdagh & Dejonckheere, 2021), scales characteristics (Haslbeck et al., 2023), temporal sampling parameters, including the number of observations and measurement frequency (Haslbeck et al., 2023), and insufficient construct relevance to participants' experiences (Siepe et al., 2025; Zelenski & Larsen, 2000). Ignoring floor and ceiling effects can lead to serious

misinterpretations in psychological research (Van Den Oord & Van Der Ark, 1997). For instance, Kalokerinos et al. (2020) have suggested that in previous studies, neuroticism was wrongfully associated with emotional variability, while this relation could at least partially be a result of its measurement using bounded scales. In fact, the skewness of the distribution of negative emotions does not allow the detection of high variability, which induces a dependency between the mean and the variability levels of a variable. Thus, when controlling for the mean level of negative emotions, the association between neuroticism and emotional variability was no longer observed (Kalokerinos et al., 2020).

As a result, there has been a growing concern about the statistical implications of substantial data skewness observed in many ESM studies and the potential for bias in the results. First, the variable affected by the floor or ceiling effect becomes less sensitive in detecting changes in the latent variable that it is supposed to measure (Šimkovic & Träuble, 2019). Second, as discussed above, bounded rating scales create an artificial relationship between a person's mean and the maximum possible variability, which can lead to potentially spurious associations when this structural dependency is not accounted for in the analyses (Baird et al., 2006; Dejonckheere et al., 2024). Therefore, conclusions drawn from studies using traditional variability measures could be incorrect or misleading, as any association found between variability measures and another variable may be confounded by the mean level (Mestdagh et al., 2018). Similarly, if the study sample is divided into subsamples based on a skewed variable, the resulting subgroups can differ in their variance on the outcome variable of interest (Terluin et al., 2016). For instance, if the splitting variable is some measure of psychopathology, the subgroup of participants who fall into the group with high levels of psychopathology will also tend to report more extreme levels of negative emotions, which would be characterized by larger variance in this variable. This could lead to potential bias in the results of any statistical analysis that is based on estimating the amount of shared variance

between negative affect and other variables (Terluin et al., 2016), which includes traditional models used to analyze ESM data, such as the linear mixed-effects (multilevel) regression model. Moreover, floor and ceiling effects may lead to a violation of the basic statistical assumptions underlying these models, that is, normality of the error and random-effects distributions, homoscedasticity of the errors, and linearity of relationships (Von Klipstein et al., 2023).

Various solutions have been proposed to mitigate the problem of floor and ceiling effects in ESM research. Some of them are related to conceptual changes in the study design, such as limiting the sample to a clinical population when measuring psychopathology or, in the case of emotions, prompting people in emotionally intense situations (Kalokerinos et al., 2020). Other suggestions are related to the measurement of the variables themselves, such as reframing assessments in a relative way, for example, by asking participants to compare their current state with the previous one (e.g., "Compared to the previous measurement, how sad do you feel right now?") (Dejonckheere et al., 2024) or with the usual one (e.g., "How sad do you feel right now compared to how you usually feel?") (Kalokerinos et al., 2020). Finally, some solutions pertain to the statistical methods themselves. For instance, Mestdagh et al. (2018) suggested a new variability measure, a relative variability index, which corrects for the confound of variability and mean levels by scaling the observed variability measure against the theoretical maximum variability at a given mean level, ensuring independence from the mean. Terluin et al. (2016) proposed to use a generalized linear mixed-effects model with an inverse Gaussian distribution for the analysis of ESM data, which better accounts for the right skewness of the dependent variable. Similarly, Weijers et al. (2018) used a two-part gamma mixed-effect model for predicting negative affect and psychotic experiences. Others have ultimately circumvented the issue by dichotomizing the outcome variable for the analysis (e.g., into symptoms present

versus absent) and using logistic mixed-effects models for the analysis (e.g., Bailen et al., 2022; Tschacher & Lienhard, 2021).

In other fields of research, floor and ceiling effects are also frequently encountered (e.g., Ferreira et al., 2012; Gasser & Rousson, 2004; Spriensma et al., 2018). For example, in clinical research, this may happen when using laboratory tests with detection limits (e.g., blood parameters, virus loads) and with patient self-reports (e.g., in rehabilitation, scoring higher than the maximum is impossible, though patients might still be improving). Consequently, some researchers have proposed viewing these situations as involving an unrestricted latent variable that is observed through a measurement instrument that cannot fully reflect the entire latent continuum, creating a form of censoring (Lin & Cheng, 2011; Spriensma et al., 2018; Twisk & Rijmen, 2009). Floor and ceiling effects, thus, could alternatively be described as lower and upper censoring. In the ESM literature, a few authors have also alluded to a similar idea, raising the question whether participants who consistently reported the lowest or highest possible scores were truly experiencing identical states at each time point, or if subtle variations existed below or above the scale's threshold but were not captured by the measurement scale (Dejonckheere et al., 2024; Kalokerinos et al., 2020; Von Klipstein et al., 2023). This conceptualization leads to the idea of viewing skewed ESM data as censored by measurement and testing properties (L. Wang et al., 2008).

One of the most commonly used methods for analyzing censored outcomes is the so-called *tobit* model (e.g., Austin et al., 2000; Bowen & Wiersema, 2005; Cunha et al., 2022; D'Angelo et al., 2013; Ferraro et al., 2002; Marshall et al., 2003). Tobit models were introduced by Tobin (1958) to deal with cases where the data distribution was characterized by a boundary value (either a minimum or maximum) where a significant portion of the responses are clustered, while the remaining data points are spread across a range of values on the opposite side of this boundary. The author argued that for such distributions, there was a need for a

model that could unite *probit analysis* to separate boundary from non-boundary observations, with multiple regression to estimate associations of the outcome variable with other predictor variables, while accounting for the concentration of values at the boundary (Tobin, 1958). As a result, tobit models, sometimes referred to as censored regression models, are capable of simultaneously addressing both censoring and outcome variability across the non-censored range.

As longitudinal studies require using methods that account for the within-person clustering of observations as well, in the presence of floor and ceiling effects, the mixed-effects tobit model could be applied. The standard mixed-effects model is defined as

$$y_{ij} = x'_{ij}\beta + z'_{ij}b_i + \epsilon_{ij}$$

where:

- y_{ij} is the observed value for observation j for person i
- x'_{ij} is the vector of covariates for the fixed effects
- β is the vector of fixed effect coefficients
- z'_{ij} is the vector of covariates for the random effects
- $b_i \sim N(0, D)$ are person-specific random effects
- $\epsilon_{ij} \sim N(0, \sigma^2)$ is the idiosyncratic error term.

The difference between tobit mixed-effects models and the standard linear one lies in the fact that the predicted variable becomes a latent variable y^*_{ij}, while the observed variable y_{ij} is defined through the following censoring mechanism (Wang & Griswold, 2016):

$$y_{ij} = \begin{cases} L \; if \; y_{ij}^{*} \leq L \; (left \; censoring) \\ y_{ij}^{*} \; if \; L \leq y_{ij}^{*} \leq U \\ U \; if \; y_{ij}^{*} \geq U \; (right \; censoring) \end{cases}$$

Despite being proven to perform better than traditional multilevel linear models (see Dagne & Huang, 2012; Spriensma et al., 2018; Twisk & Rijmen, 2009), there are only a few examples of studies actually using this approach. However, as discussed earlier, the latter can introduce significant bias in the results in the case of a skewed outcome variable, as they estimate the relationship between predictors and the censored observed variable, instead of the latent variable. Using a mixed-effects tobit model could provide an alternative that potentially reduces this bias. To the best of our knowledge, the application of mixed-effects tobit models to ESM data exhibiting floor or ceiling effects remains unexplored in the psychological literature. In the present study, we examine the potential utility of this approach and compare its performance to the conventional linear mixed-effects model when analyzing skewed ESM data.

Methods

Data

The current work makes use of the data from an ESM study comprising 328 participants who completed questionnaires assessing mood and various contextual variables ten times per day across six consecutive days. The sample was stratified into three groups according to mental health status: 112 healthy controls, 109 individuals with a lifetime history of depression and current residual depressive symptoms, and 107 participants diagnosed with a psychotic disorder (predominantly schizophrenia).

Participants received questionnaire prompts at semi-randomized intervals within 90-minute time blocks, commencing at 7:30 and concluding at 22:30 each day (with the initial daily prompt delivered between 7:30-9:00, the second between 9:00-10:30, and so forth).

Prompt timing was algorithmically determined to ensure inter-prompt intervals of a minimum of 20 and a maximum of 160 minutes, with a mean inter-prompt interval of approximately 90 minutes (SD approximately 30 minutes). Questionnaire prompts were transmitted via wristwatch devices, and participants recorded responses in portable booklets carried throughout the study period.

Aside from a number of other variables, at each prompt, participants were asked to rate their current emotional state using three positive affect (PA) items rated 1 (*not at all*) to 7 (*very much*) (cheerful, relaxed, satisfied) and six negative affect (NA) items using identical scaling (irritated, anxious, down, guilty, insecure, lonely). In addition, participants provided an event pleasantness (EP) valence rating from -3 (*very unpleasant*) to +3 (*very pleasant*) of the most significant inter-prompt event.

The dataset is freely available at https://osf.io/x2rhw/. The dataset has undergone minor modifications from the original for didactic purposes and to further anonymize the participants. However, the reported data structure and findings correspond to those observed in the original sample.

Statistical Analysis

PA and NA scores at each prompt were computed by averaging the three PA and the six NA items, respectively. We checked if there were any participants with no variance in any of the three variables of interest (i.e., PA, NA, and EP). There were four participants with constant values of NA and two participants with constant values of EP. These participants were excluded from the analyses, resulting in 109 participants in the depressed and control groups and 104 participants in the psychotic group (N = 322). This subset of participants provided a total of (109 + 109 + 104) × 60 = 19,320 observations. After excluding missing data (due to prompts not responded to or skipped items) in any of the three variables of interest, the final

dataset for the analyses included 13,706 observations with complete data for all three variables of interest.

As a prototypical example of analyses that are commonly carried out with ESM data, we examined if affect could be predicted by event pleasantness. Analyses along these lines have been conducted in a number of previous studies (e.g., Bylsma et al., 2011; Pacheco-Romero et al., 2024; Peeters et al., 2003) to examine how sensitive people's emotional state is to the occurrence of pleasant / unpleasant events that are happening in their daily lives. Such analyses are typically conducted using a standard linear mixed-effects regression model with event pleasantness as predictor, PA or NA as outcome, and with (correlated) random intercepts and slopes at the subject level, to account for subject-level differences in the overall level of affect and in the strength of the association between the predictor and the outcome (e.g., Bolger & Laurenceau, 2013; Viechtbauer, 2022).

For PA, such an analysis approach raises little concerns. The variable is fairly symmetrically distributed and extreme values at the bounds of the scale (i.e., 1 or 7) are rare: out of the 13,706 observations, 275 values (2.0%) were exactly equal to 1 and 583 values (4.3%) were exactly equal to 7. On the other hand, NA showed a rather right-skewed distribution, with 5,243 values (38.3%) exactly equal to 1 and only 8 observations (0.1%) exactly equal to 7. This pattern is particularly notable, as since PA had only 3 items and NA had 6, all else being equal, such extreme values are more likely to occur for PA than for NA, because the probability of someone answering at least one item in a non-extreme way arguably increases with the number of items. However, the observed distributions show the opposite tendency. Considering these distributional characteristics, if we treat the values of NA at the bounds as censored observations of a latent affect continuum, then the standard mixed-effects model could underestimate the true association between event pleasantness and NA. As an

alternative analysis approach, one could therefore consider the use of a mixed-effects tobit model, again with random intercepts and slopes at the subject level as described above.

To examine how much such censoring might affect the results of the standard model and to what extent the tobit model can compensate for this problem, we first fitted both models with PA as the outcome. The results obtained from this model were expected to be similar and were considered the "true" (or original) estimates of the association with an essentially uncensored dependent variable. We then induced increasing levels of censoring in the outcome by subtracting a constant value from the original PA scores, assigning the minimum value of 1 to all observations that fell below it (a similar artificial censoring approach was used by Y. Liu et al., 2018, to estimate high or low quantiles of a distribution, allowing "zooming" into its tails). For the purposes of our analysis, we used 20 different values that were subtracted from the original scores, ranging from 0 to 3, not going beyond to preserve some variability in the data.

Our aim was to compare the results from the standard model with those from the tobit model in terms of their capability to recover the original estimates of the model fitted to the original uncensored dependent variable. We compared the following parameter estimates from the two models: the fixed effect estimates (i.e., the estimated average intercept and slope), the corresponding standard errors (SEs), the estimated standard deviations (SDs) of the random intercepts and slopes, the estimated error SD, and the estimated correlation between the intercept and slope random effects. Finally, to compare results across a truly skewed outcome variable, we examined the association between event pleasantness and NA using the standard mixed-effects model and the corresponding tobit model.

We used the R (R Core Team, 2025) package *nlme* (version 3.1-166; Pinheiro et al., 2021) to fit the standard linear mixed-effects model and the *brms* package (version 2.22.0;

Bürkner, 2018) to fit the corresponding tobit model. While *brms* uses a Bayesian framework for model fitting, the default priors used are fairly vague, and the dataset is quite large, such that the resulting estimates (i.e., the means of the posterior distributions) can be treated essentially as maximum likelihood estimates. We used the default settings for the number of chains (4), but increased the number of burn-in iterations (to 2000), and the number of samples extracted from each chain (to 8000) to obtain stable results. All analyses were carried out separately within each mental health status group as the overall levels of affect differed across these subgroups.

Results

Data Description

Figure 1 shows the distribution of PA (Figure 1a) and NA (Figure 1e) and the distribution of PA with increasing censoring induced by subtracting values of 1, 2, and 3 from the original data (Figure 1b-d). It is apparent from the graphs that this form of subtracting is gradually shifting PA from an initially fairly symmetrically distributed variable to a right-skewed distribution which resembles the one for NA when a large value is subtracted.

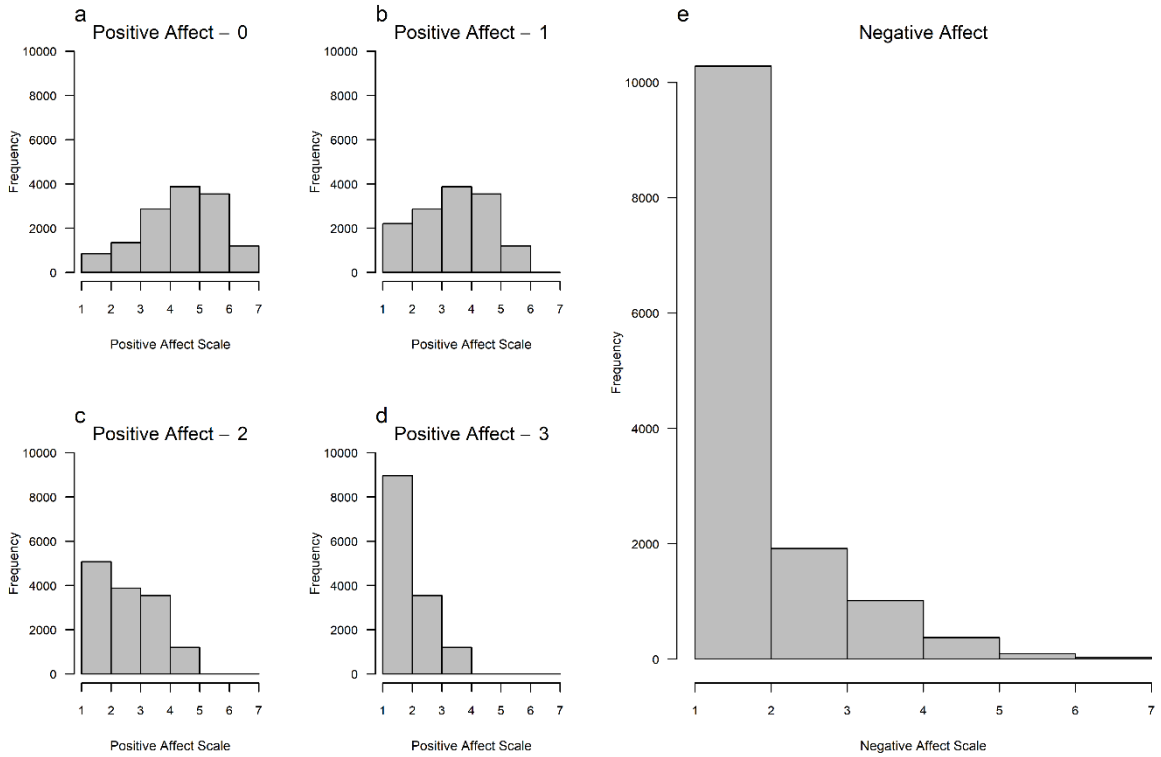


Figure 1. Distributions of the variables: positive affect and the artificially censored positive affect variable with subtracted values of 1, 2, and 3, and negative affect. The x-axis represents the variable values, the y-axis represents the frequencies of the measurements.

Linear and tobit mixed-effects models for artificially censored positive affect

The results of the comparison of the standard linear mixed-effects model and the tobit mixed-effects model are illustrated in Figure 2 for the control group, in Figure 3 for the depressed group, and in Figure 4 for the psychotic group. The figures compare the parameter estimates obtained from the two models: the intercept and slope estimates, their SEs, the SDs of the random effects and the errors, and the intercept-slope correlation. The horizontal axis corresponds to the value subtracted from the original PA scores. The vertical axis corresponds to the estimated value of a given parameter.

The estimates shown on the y-axis for an x-axis value of zero represent the original values of the estimates without essentially any censoring. Thus, we evaluate the performance of each model based on how stable (close to the original one) each estimate is as increasingly

large values are subtracted from PA. The main coefficient of interest is the slope estimate, as it represents the strength of the relationship between the predictor (EP) and the outcome (PA). It is apparent from Figures 2-4 that in all three groups of participants, the slope coefficient was better (i.e., more stably) estimated by the tobit model, even after subtracting up to three points from the original PA scores. By contrast, the standard model gradually produced smaller and smaller slope estimates with increasing subtraction values. This indicates that the standard model was unable to recover the true relationship between the predictor and the artificially censored outcome variable, while the tobit model showed rather stable results, more or less recovering the original value of the estimate. There was also a slight decrease in the slope estimate of the tobit model in the control and psychotic groups, but this decrease was much less pronounced than the one observed for the standard model.

A potential limitation of the tobit model is that the precision of the slope estimate may decrease as the outcome variable becomes more skewed. However, an examination of the plot of the slope SEs indicates that this was not the case. Either the SE remained fairly stable (in the depression group) or even decreased slightly (in the control and psychotic groups). In contrast, the SEs obtained from the standard model showed a pronounced monotonic decrease with the subtraction value. This may initially appear to be a good thing, but given the corresponding decrease in the slope estimate, this implies that we are instilling increased confidence in an increasingly biased estimate.

The SDs of the random effects and the error SD showed analogous patterns. In all three groups, there was a slight uptick in the estimates from the tobit model for the most extreme subtraction value of 3. This might be an indication that the model may start to behave somewhat erratically for the rather skewed distribution that this subtraction value implies (see Figure 1d), although in the control and psychotic groups, the corresponding estimates were then actually closer to the estimates from the original PA variable. The estimated intercept-slope correlation

was also more stable and closer to the original values for the tobit model, while the standard model overestimated the correlation as the subtraction value increased.

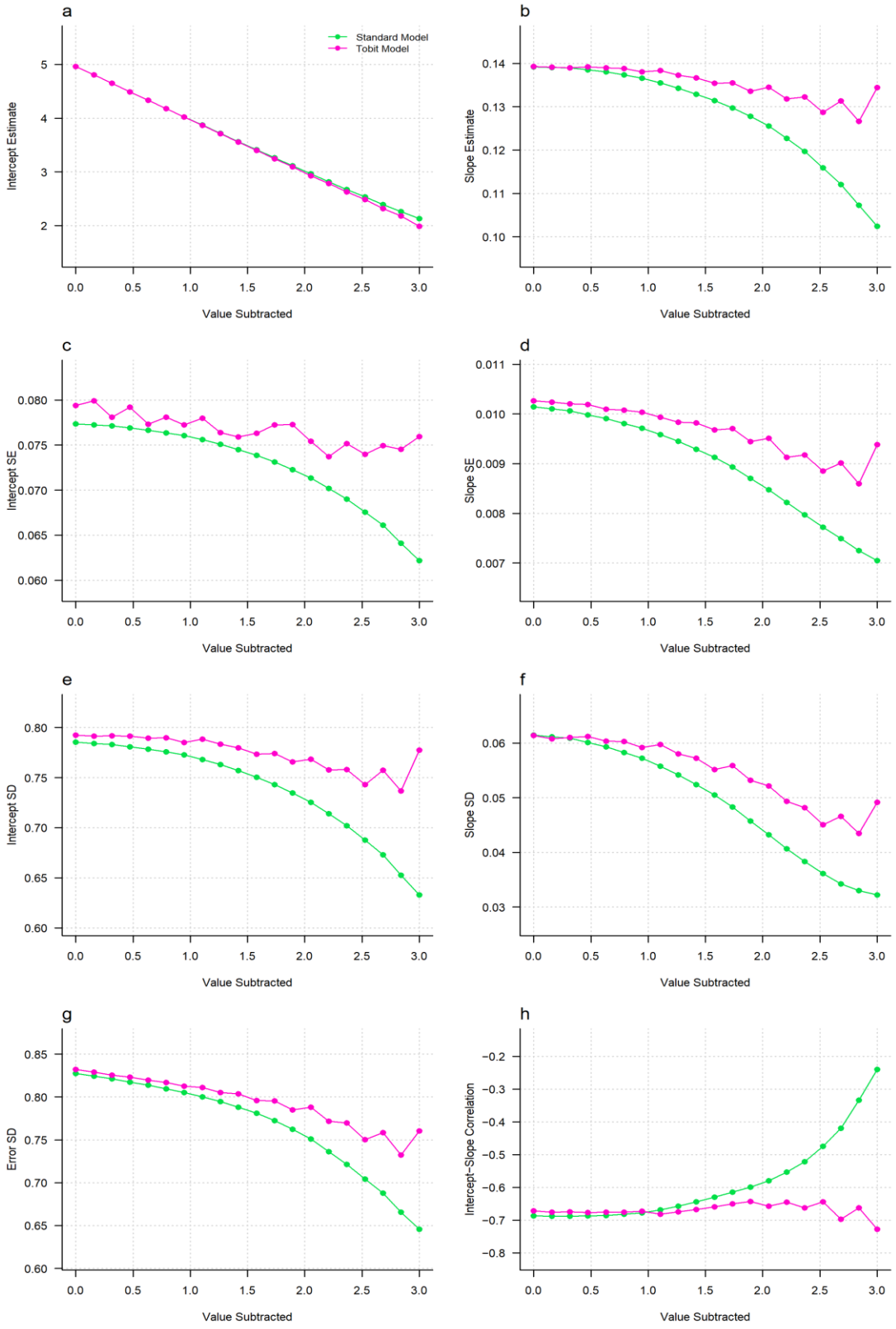


Figure 2. Control group: Comparison of parameter estimates of the standard linear mixed-effects model and the tobit mixed-effects model predicting positive affect from event pleasantness, increasingly subtracting values from 0 to 3 from the original positive affect value (artificial censoring). Abbreviations: SE: standard error; SD: standard deviation.

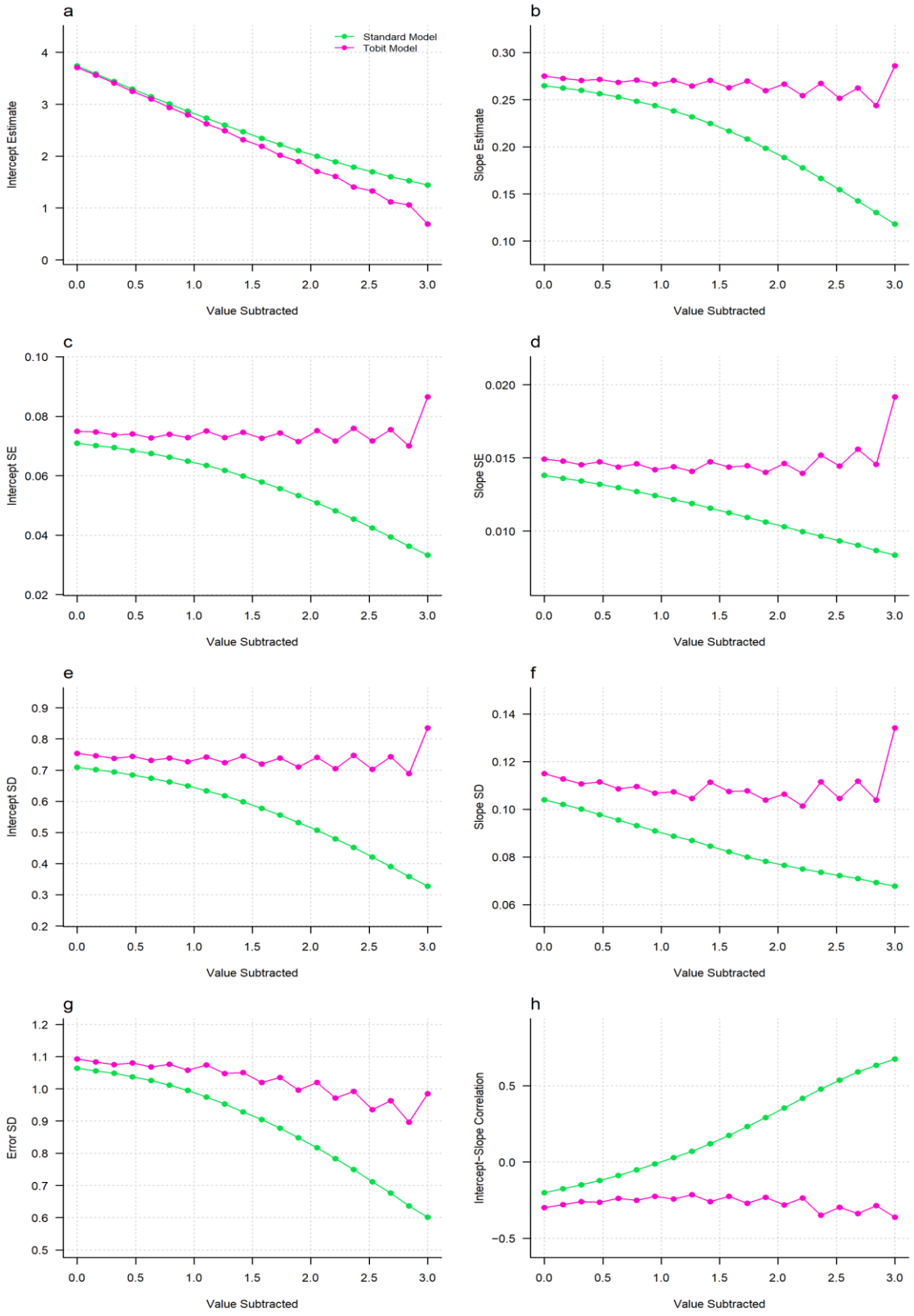


Figure 3. Depressed group: Comparison of parameter estimates of the standard linear mixed-effects model and the tobit mixed-effects model predicting positive affect from event pleasantness, increasingly subtracting values from 0 to 3 from the original positive affect value (artificial censoring). Abbreviations: SE: standard error; SD: standard deviation.

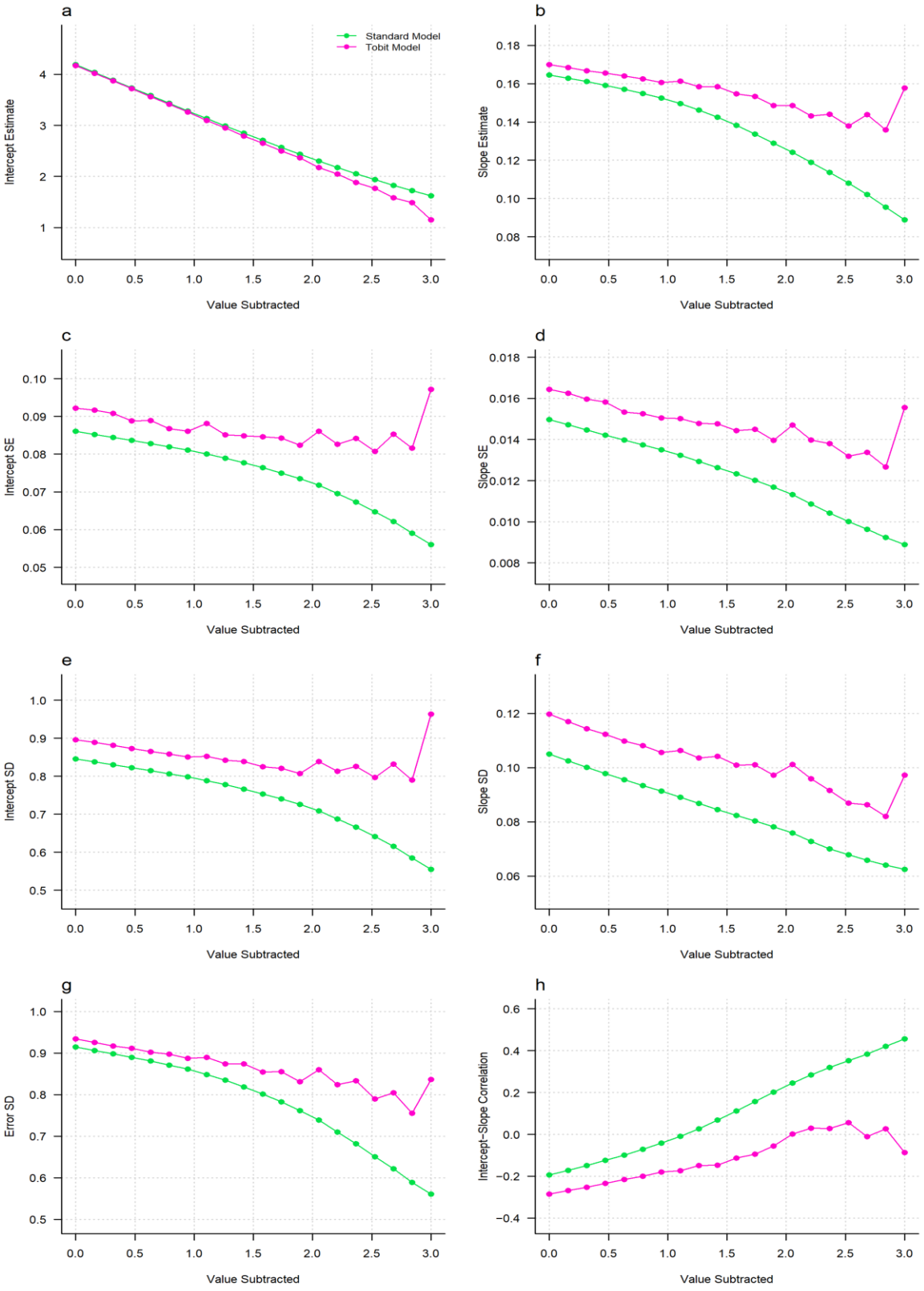


Figure 4. Psychotic group: Comparison of parameter estimates of the standard linear mixed-effects model and the tobit mixed-effects model predicting positive affect from event pleasantness, increasingly subtracting values from 0 to 3 from the original positive affect value (artificial censoring). Abbreviations: SE: standard error; SD: standard deviation.

Linear and tobit mixed-effects models for negative affect

Table 1 compares the results of the standard linear mixed-effects model to those from the tobit model when predicting NA from EP across the three groups of participants. Note that NA is naturally quite skewed in this sample (Figure 1e) and therefore some of the biases we observed in the previous section would be expected to affect the standard model to a certain extent.

Table 1. *Comparison of parameter estimates of the standard linear mixed-effects model and the tobit mixed-effects model predicting negative affect from event pleasantness by participant group.*

Model	**Intercept**	**Slope**	**Intercept SE**	**Slope SE**	**Intercept SD**	**Slope SD**	**Error SD**	**Intercept-slope correlation**
Control group								
Standard	1.33	-0.05	0.037	0.006	0.38	0.05	0.39	-0.78
Tobit	0.85	-0.11	0.067	0.010	0.66	0.04	0.77	-0.35
Depressed group								
Standard	2.30	-0.15	0.070	0.011	0.72	0.08	0.79	-0.58
Tobit	2.19	-0.18	0.082	0.012	0.85	0.08	0.92	-0.26
Psychotic group								
Standard	1.93	-0.09	0.079	0.012	0.79	0.09	0.63	-0.61
Tobit	1.69	-0.13	0.107	0.016	1.06	0.11	0.82	-0.27

Several observations can be made with respect to these results. First, the slope estimates were larger for the tobit model compared to the standard one. This was especially prominent in the control group, in which the mean NA value is particularly low compared to the depressed and psychotic groups. In this group, the slope was more than twice as large in the tobit model compared to the standard one. However, in the latter two groups, the estimated relationship was also stronger for the tobit model. At the same time, the slope SEs and the SDs of the slope random effect were slightly larger for the tobit model.

On the other hand, the intercept-slope correlations were more negative (i.e., stronger) for the standard model compared to the tobit one, that is, participants with high average NA values tended to have more negative relationships between EP and NA. This can be seen in Figure 5 (left column), which shows scatterplots of the estimated subject-specific intercepts and slopes based on the standard model (i.e., adding together the fixed effect estimates – as indicated by the dotted lines – and the best linear unbiased predictions of the random effects). This finding might suggest that the affective state of participants who report higher levels of NA on average tended to be more sensitive to the occurrence of (un)pleasant events in their daily lives. However, this finding might, at least in part, be an artifact of the censored distribution of NA. For participants with low average NA, the relationship can simply not be as strong because even for very pleasant events, the observed value of NA cannot dip below 1. As a result, participants with intercepts close to 1 tended to have slopes close to 0. However, in the tobit model, the slope reflects the relationship between EP and the latent NA continuum, which is not constrained by the lower bound of 1 of the censored observed variable. Figure 5 (right column) shows the relationship between the subject-specific intercepts and slopes for this model. Here, the correlations were weaker, suggesting less of an association between the average NA levels of the participants and their affect sensitivity. In addition, it is worth noting that the tobit model produced higher intercept SDs across all participant groups, which reflects

more inter-individual variability in the mean levels of negative effect. This can also be seen in Figure 5 (right column) in the increased range of the subject-specific intercepts from the tobit model.

Overall, these results indicate that there are considerable differences in key parameter values estimated by the standard and tobit mixed-effects models, suggesting stronger relationships between EP and NA overall, but a weaker relationship between the average NA levels of the participants and their affect sensitivity.

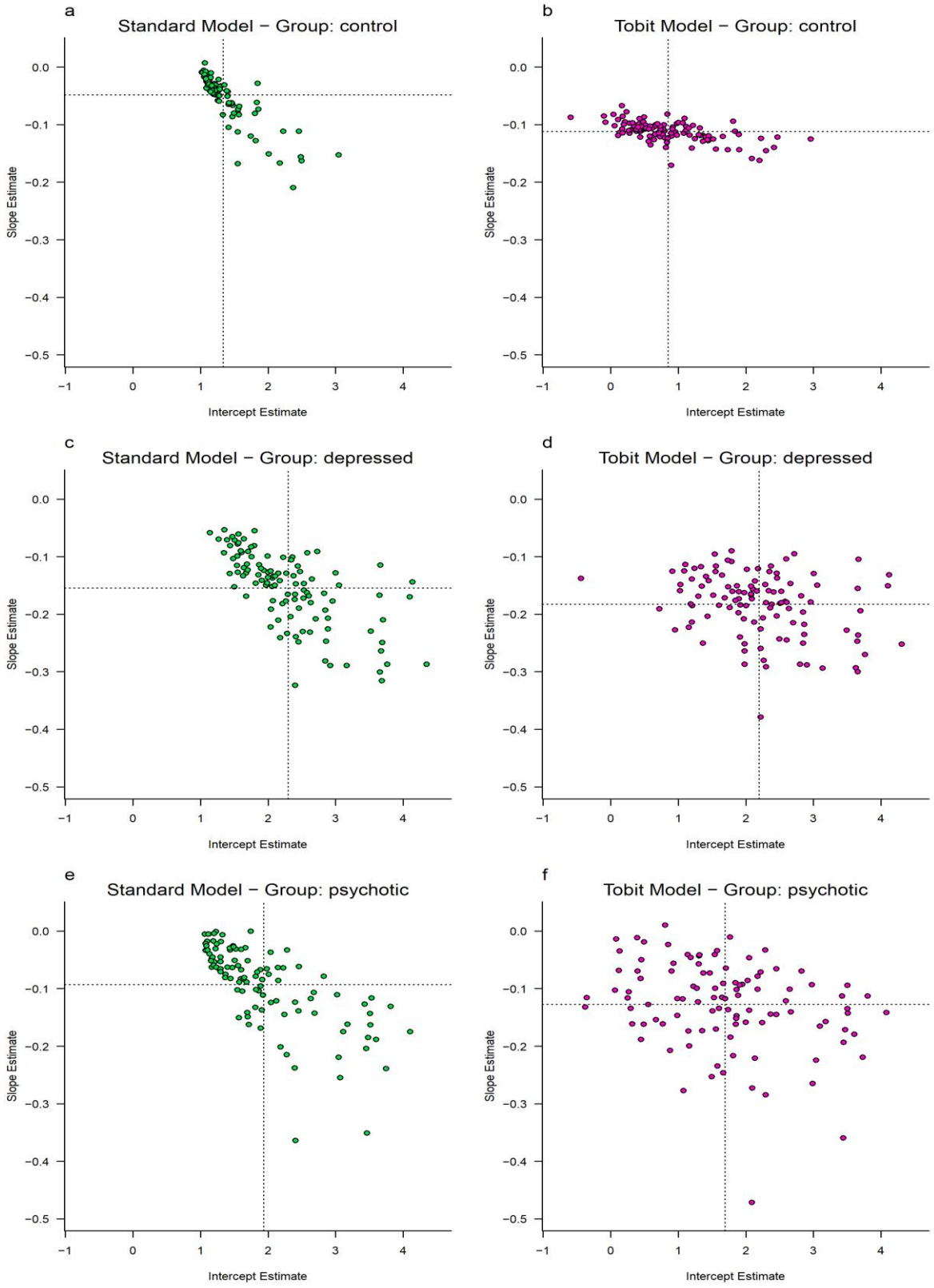


Figure 5: Scatterplots of the subject-specific intercepts and slopes (obtained by adding together the fixed effect estimates – as indicated by the dotted lines – and the best linear unbiased predictions of the random effects) by group, estimated by the standard linear mixed-effects model (left column) and the tobit mixed-effects model (right column).

Discussion

The current study addresses one of the fundamental methodological challenges in ESM research – the mismatch between the statistical assumptions of the models used and the real data characteristics. From the ESM literature, it is known that many variables of interest exhibit highly-skewed distributions with floor or ceiling effects (Mestdagh & Dejonckheere, 2021; Siepe et al., 2025; L. Wang et al., 2008), which can lead to potential bias in the results of the analyses (Kalokerinos et al., 2020; Terluin et al., 2016). In this paper, we propose to conceptualize such data to be affected by censoring, where true values may extend beyond the observable range, and thus the variable of interest represents a latent construct observed through a limited value scale. Based on this perspective, our suggestion is to use a tobit mixed-effects model for analyzing such skewed data (Dagne & Huang, 2012; Tobin, 1958). To demonstrate the improved performance of this model compared to the standard linear mixed-effects model, we compared the models' estimates for artificially censored data with the original values obtained from the essentially uncensored variable. Tobit model coefficients should be interpreted as the expected change in the latent continuous variable for a one-unit change in the predictor, recognizing that this latent change may not fully manifest in observed scores for individuals whose observations are collocated near boundaries.

The artificial censoring of the PA values demonstrated that the mixed-effects tobit model provided relatively stable parameter estimates as the data were progressively more censored. While the standard linear mixed-effects model showed a tendency for declining slope estimates with increased censoring, the tobit model appeared to better preserve the original parameter estimates. This finding suggests that tobit models have a better capability to recover the real effect size of a relationship when the data are skewed. The given results also agree with an earlier study, which showed that removing skewness from data through transformations can lead to larger correlations between variables (Dunlap et al., 1995). The stability of the tobit

model extends beyond the slope estimates to encompass other critical parameters, including standard errors, the standard deviation of the random effects and the errors, and the intercept-slope correlation. Notably, the lower intercept-slope correlations observed in the tobit model may indicate that it does not artificially inflate the association between individual differences in mean levels and relationship patterns, as the standard model does. Another important aspect is that the tobit model showed quite stable performance across different participant groups, demonstrating its robustness and generalizability across diverse psychological samples.

When applied to the naturally skewed NA variable, the tobit model estimated larger effect sizes compared to the standard model, in line with the findings obtained with the artificially censored data. This pattern was consistent across all groups, suggesting that the standard model may systematically underestimate the true strength of relationships when dealing with skewed outcomes. The improved sensitivity to detect relationships makes tobit models particularly valuable for studies examining subtle psychological states, where effect sizes may be small but theoretically meaningful. Moreover, the tobit models revealed substantially higher intercept standard deviations across all groups, indicating greater inter-individual variability in mean levels of negative affect than suggested by the standard model. This enhanced detection of individual differences has important implications for understanding the heterogeneity of psychological experiences and may lead to more nuanced modeling of psychological phenomena. Overall, these findings suggest that tobit models offer certain advantages when analyzing data characterized by skewness compared to the standard linear model.It is worth mentioning that the adoption of tobit models for skewed ESM data does not merely represent a technical solution to a statistical problem but reflects a fundamental conceptual shift in how we understand the nature of these data. While various statistical approaches have been proposed in the past to address skewed data distributions – such as using inverse Gaussian distributions (Terluin et al., 2016), gamma distributions (Weijers et al., 2018),

or dichotomizing outcomes (Bailen et al., 2022; Tschacher & Lienhard, 2021) – the tobit approach follows naturally from seeing the observations at the extremes of a scale as representing censored manifestations of an underlying latent continuum. This theoretical perspective acknowledges that when participants consistently report the minimum (or maximum) score on a scale (e.g., 1 for negative affect on a 1-7 Likert scale), there may still exist subtle but real differences in their latent emotional states that the measurement instrument fails to capture. In contrast, dichotomizing the outcome leads to a loss of valuable information, eliminating variability that could be clinically or theoretically meaningful.

Such a conceptual view appears to be reasonable, as people may not be sufficiently sensitive in their introspection to distinguish subtle differences in variables like negative affect at the lower end of the emotional spectrum, similar to how there are well-known limits on our introspection to higher-order cognitive processes (Nisbett & Wilson, 1977). Moreover, as was pointed out in the literature, the absolute nature of bounded scales requires participants to anticipate future rating scenarios for proper calibration (Dejonckheere et al., 2024). To illustrate this concept, the authors give the following example: selecting the maximum score on a sadness scale following an argument with a friend leaves no room to reflect a more extreme emotional state should a tragic event subsequently occur, such as learning of that friend's death in an accident. Thus, the minimum or maximum value on a bounded scale can actually reflect different latent states and have meaningful consequences and therefore should not be treated as equivalent in the analysis. The tobit model, therefore, not only provides a more appropriate statistical solution but also reflects the phenomenological complexity of real-life experiences by recognizing the existence of psychological states that extend beyond the artificial boundaries of measurement tools.

While the theoretical and empirical aspects of the application of tobit mixed-effects models are reassuring, several practical considerations deserve attention. The computational complexity of tobit models, particularly when implemented in a Bayesian framework using packages like *brms* (Bürkner, 2018), may present certain technical challenges. These challenges are related to the fact that the model requires the estimation of a large number of parameters: at a minimum, the fixed effects, the variance and correlation among the random effects, and the error variance. This is particularly complex in a Bayesian framework, which requires thousands of iterations, warm-up phases, and complex computations at each iteration. As a consequence, the computations have a high memory demand and may require a considerable amount of processing time. With complex models and especially in smaller datasets, technical expertise is also needed to assess model convergence. A future study could assess maximum likelihood estimation methods for fitting tobit mixed-effects models.

Several other considerations about study limitations and potential future developments need to be acknowledged. In our study, we focused on contemporaneous relationships between two variables only. A natural progression of this work would be to assess the performance of the tobit model when estimating temporal relationships and when models include multiple predictor variables. Finally, this study focused primarily on affect and related constructs; thus, the question about the generalizability of the results arises. Future research should examine the tobit model performance across a broader range of naturally occurring skewed distributions and different types of psychological constructs, including psychopathology symptoms, which are widely studied in clinical research and often considerably skewed.

Conclusions

This study provides first evidence for the potential advantages of tobit mixed-effects models over conventional linear mixed-effects models when analyzing right-skewed ESM data. The enhanced parameter recovery, improved effect size estimation, and greater stability across

varying degrees of skewness make tobit models a valuable methodological tool for researchers studying psychological phenomena prone to floor and ceiling effects. We recommend that ESM researchers routinely assess the distributional characteristics of their outcome variables and consider censored regression approaches when substantial floor or ceiling effects are present.

As ESM continues to evolve as a primary method for studying psychological processes in naturalistic settings, the adoption of appropriate statistical methods that account for the unique characteristics of intensive longitudinal data becomes increasingly critical. The improved performance of the tobit mixed-effects model compared to the standard linear one when applied to skewed data represents an important advancement in ESM methodology, offering researchers a principled approach to analyzing bounded psychological data, which often characterizes ESM studies.

The implications of this work extend beyond methodological considerations to potentially reshape our understanding of psychological phenomena previously studied using traditional approaches. By providing more accurate estimates of relationships and individual differences, tobit models may reveal previously undetectable patterns in psychological data and contribute to more nuanced theories of human experience.

References

Austin, P. C., Escobar, M., & Kopec, J. A. (2000). The use of the Tobit model for analyzing measures of health status. *Quality of Life Research*, *9*, 901–910. https://doi.org/10.1023/a:1008938326604

Bailen, N. H., Koval, P., Strube, M., Haslam, N., & Thompson, R. J. (2022). Negative emotion and nonacceptance of emotion in daily life. *Emotion*, *22*(5), 992–1003. https://doi.org/10.1037/emo0000898

Baird, B. M., Le, K., & Lucas, R. E. (2006). On the nature of intraindividual personality variability: Reliability, validity, and associations with well-being. *Journal of Personality and Social Psychology*, *90*(3), 512–527. https://doi.org/10.1037/0022-3514.90.3.512

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research.* The Guilford Press.

Bowen, H. P., & Wiersema, M. F. (2005). Foreign-based competition and corporate diversification strategy. *Strategic Management Journal*, *26*(12), 1087–1181. https://doi.org/10.1002/smj.499

Bürkner, P.-C. (2018). Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, *10*(1), 395. https://doi.org/10.32614/RJ-2018-017

Bylsma, L. M., Taylor-Clift, A., & Rottenberg, J. (2011). Emotional reactivity to daily events in major and minor depression. *Journal of Abnormal Psychology*, *120*(1), 155–167. https://doi.org/10.1037/a0021662

Cunha, D. R., Divino, J. A., & Saulo, H. (2022). On a log-symmetric quantile tobit model applied to female labor supply data. *Journal of Applied Statistics*, *49*(16), 4225–4253. https://doi.org/10.1080/02664763.2021.1976120

Dagne, G., & Huang, Y. (2012). Bayesian inference for a nonlinear mixed-effects Tobit model with multivariate skew-t distributions: Application to AIDS studies. *The International Journal of Biostatistics*, *8*(1). https://doi.org/10.1515/1557-4679.1387

D'Angelo, A., Majocchi, A., Zucchella, A., & Buck, T. (2013). Geographical pathways for SME internationalization: Insights from an Italian sample. *International Marketing Review*, *30*(2), 80–105. https://doi.org/10.1108/02651331311314538

Dejonckheere, E., Penne, I., Briels, L., & Mestdagh, M. (2024). For Better or for Worse? Visualizing Previous Intensity Levels Improves Emotion (Dynamic) Measurement in Experience Sampling. *Psychological Assessment*, *36*(3), 215–234.

https://doi.org/10.1037/pas0001296

Dunlap, W. P., Burke, M. J., & Greer, T. (1995). The Effect of Skew on the Magnitude of Product-Moment Correlations. *The Journal of General Psychology*, *122*(4), 365–377.

Ferraro, K. F., Su, Y., Gretebeck, R. J., Black, D. R., & Badylak, S. F. (2002). Body Mass Index and Disability in Adulthood: A 20-Year Panel Study. *American Journal of Public Health*, *92*(5), 834–840. https://doi.org/10.2105/AJPH.92.5.834

Ferreira, L. N., Ferreira, P. L., Pereira, L. N., & Rowen, D. (2012). Reducing the Floor Effect in the SF-6D: A Feasibility Study. *Applied Research in Quality of Life*, *7*(2), 193–208. https://doi.org/10.1007/s11482-011-9149-3

Gasser, T., & Rousson, V. (2004). Modelling neuromotor ratings with floor-effects. *Statistics in Medicine*, *23*, 3641–3653. https://doi.org/10.1002/sim.1914

Hamaker, E. L., & Wichers, M. (2017). No Time Like the Present: Discovering the Hidden Dynamics in Intensive Longitudinal Data. *Current Directions in Psychological Science*, *26*(1), 10–15. https://doi.org/10.1177/0963721416666518

Haslbeck, J., Ryan, O., & Dablander, F. (2023). Multimodality and skewness in emotion time series. *Emotion*, *23*(8), 2117–2141. https://doi.org/10.1037/emo0001218

Kalokerinos, E. K., Murphy, S. C., Koval, P., Bailen, N. H., Crombez, G., Hollenstein, T., Gleeson, J., Thompson, R. J., Van Ryckeghem, D. M. L., Kuppens, P., & Bastian, B. (2020). Neuroticism may not reflect emotional variability. *Proceedings of the National Academy of Sciences*, *117*(17), 9270–9276. https://doi.org/10.1073/pnas.1919934117

Lin, K.-C., & Cheng, S.-F. (2011). Tobit Model for Outcome Variable Is Limited by Censoring in Nursing Research. *Nursing Research*, *60*(5), 354–360. https://doi.org/10.1097/NNR.0b013e318226a091

Liu, Y., Salibián-Barrera, M., Zamar, R. H., & Zidek, J. V. (2018). Using Artificial Censoring to Improve Extreme Tail Quantile Estimates. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *67*(4), 791–812. https://doi.org/10.1111/rssc.12262

Lucas, R. E., Wallsworth, C., Anusic, I., & Donnellan, M. B. (2021). A direct comparison of the day reconstruction method (DRM) and the experience sampling method (ESM). *Journal of Personality and Social Psychology*, *120*(3), 816–835. https://doi.org/10.1037/pspp0000289

Marshall, T. A., Levy, S. M., Broffitt, B., Warren, J. J., Eichenberger-Gilmore, J. M., Burns, T. L., & Stumbo, P. J. (2003). Dental Caries and Beverage Consumption in Young Children. *Pediatrics*, *112*(3), e184–e191. https://doi.org/10.1542/peds.112.3.e184

Mestdagh, M., & Dejonckheere, E. (2021). Ambulatory assessment in psychopathology

research: Current achievements and future ambitions. *Current Opinion in Psychology*, *41*, 1–8. https://doi.org/10.1016/j.copsyc.2021.01.004

Mestdagh, M., Pe, M., Pestman, W., Verdonck, S., Kuppens, P., & Tuerlinckx, F. (2018). Sidelining the mean: The relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychological Methods*, *23*(4), 690–707. https://doi.org/10.1037/met0000153

Nisbett, R. E., & Wilson, T. D. (1977). Telling More Than We Can Know: Verbal Reports on Mental Processes. *Psychological Review*, *84*(3), 231–259. https://doi.org/10.1037/0033-295x.84.3.231

Pacheco-Romero, A. M., Martín-García, Ó., Rey-Sáez, R., Boemo, T., Blanco, I., Vázquez, C., & Sánchez-López, Á. (2024). An integrative analysis of potential mechanisms of reduced positive affect in daily life in depression: An ESM study. *Cognition and Emotion*, *38*(4), 587–604. https://doi.org/10.1080/02699931.2024.2314066

Peeters, F., Nicolson, N. A., Berkhof, J., Delespaul, P., & deVries, M. (2003). Effects of daily events on mood states in major depressive disorder. *Journal of Abnormal Psychology*, *112*(2), 203–211. https://doi.org/10.1037/0021-843X.112.2.203

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2021). *nlme: Linear and Nonlinear Mixed Effects Models.* [Computer software]. https://CRAN.R-project.org/package=nlme

Schoevers, R. A., Van Borkulo, C. D., Lamers, F., Servaas, M. N., Bastiaansen, J. A., Beekman, A. T. F., Van Hemert, A. M., Smit, J. H., Penninx, B. W. J. H., & Riese, H. (2021). Affect fluctuations examined with ecological momentary assessment in patients with current or remitted depression and anxiety disorders. *Psychological Medicine*, *51*(11), 1906–1915. https://doi.org/10.1017/S0033291720000689

Siepe, B. S., Rieble, C. L., Tutunji, R., Rimpler, A., März, J., Proppert, R. K. K., & Fried, E. I. (2025). Understanding Ecological-Momentary-Assessment Data: A Tutorial on Exploring Item Performance in Ecological-Momentary-Assessment Data. *Advances in Methods and Practices in Psychological Science*, *8*(1), 25152459241286877. https://doi.org/10.1177/25152459241286877

Šimkovic, M., & Träuble, B. (2019). Robustness of statistical methods when measure is affected by ceiling and/or floor effect. *PLOS ONE*, *14*(8), e0220889. https://doi.org/10.1371/journal.pone.0220889

Spriensma, A. ., Eekhout, I., De Boer, M. R., Luime, J. J., De Jong, P. H., Bahçecitapar, M. K., Heymans, M. W., & Twisk, J. W. R. (2018). Analysing outcome variables with floor

effects due to censoring: A simulation study with longitudinal trial data. *Epidemiology, Biostatistics, and Public Health*, *15*(2). https://doi.org/10.2427/12850

Tammilehto, J., Bosmans, G., Kuppens, P., Flykt, M., Peltonen, K., Kerns, K. A., & Lindblom, J. (2022). Dynamics of attachment and emotion regulation in daily life: Uni- and bidirectional associations. *Cognition and Emotion*, *36*(6), 1109–1131. https://doi.org/10.1080/02699931.2022.2081534

Telford, C., McCarthy-Jones, S., Corcoran, R., & Rowse, G. (2012). Experience Sampling Methodology studies of depression: The state of the art. *Psychological Medicine*, *42*(6), 1119–1129. https://doi.org/10.1017/S0033291711002200

Terluin, B., De Boer, M. R., & De Vet, H. C. W. (2016). Differences in Connection Strength between Mental Symptoms Might Be Explained by Differences in Variance: Reanalysis of Network Data Did Not Confirm Staging. *PLOS ONE*, *11*(11), e0155205. https://doi.org/10.1371/journal.pone.0155205

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables. *Econometrica*, *26*(1), 24. https://doi.org/10.2307/1907382

Trampe, D., Quoidbach, J., & Taquet, M. (2015). Emotions in Everyday Life. *PLOS ONE*, *10*(12), e0145450. https://doi.org/10.1371/journal.pone.0145450

Trull, T. J., & Ebner-Priemer, U. (2014). The Role of Ambulatory Assessment in Psychological Science. *Current Directions in Psychological Science*, *23*(6), 466–470. https://doi.org/10.1177/0963721414550706

Tschacher, W., & Lienhard, N. (2021). Mindfulness Is Linked with Affectivity in Daily Life: An Experience-Sampling Study with Meditators. *Mindfulness*, *12*(6), 1459–1472. https://doi.org/10.1007/s12671-021-01615-7

Twisk, J., & Rijmen, F. (2009). Longitudinal tobit regression: A new approach to analyze outcome variables with floor or ceiling effects. *Journal of Clinical Epidemiology*, *62*(9), 953–958. https://doi.org/10.1016/j.jclinepi.2008.10.003

Van Den Oord, E. J. C. G., & Van Der Ark, L. A. (1997). A note on the use of the Tobit approach for tests scores with floor or ceiling effects. *British Journal of Mathematical and Statistical Psychology*, *50*(2), 351–364. https://doi.org/10.1111/j.2044-8317.1997.tb01150.x

Viechtbauer, W. (2022). Statistical methods for ESM data. In *I. Myin-Germeys & P. Kuppens (Eds.), The open handbook of Experience Sampling Methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies* (2nd ed., pp. 153–183). Center for Research on Experience Sampling and Ambulatory Methods Leuven.

Von Klipstein, L., Servaas, M. N., Lamers, F., Schoevers, R. A., Wardenaar, K. J., & Riese, H. (2023). Increased affective reactivity among depressed individuals can be explained by floor effects: An experience sampling study. *Journal of Affective Disorders*, *334*, 370–381. https://doi.org/10.1016/j.jad.2023.04.118

Wang, L., Zhang, Z., McArdle, J. J., & Salthouse, T. A. (2008). Investigating Ceiling Effects in Longitudinal Data Analysis. *Multivariate Behavioral Research*, *43*(3), 476–496. https://doi.org/10.1080/00273170802285941

Wang, R., Aung, M. S. H., Abdullah, S., Brian, R., Campbell, A. T., Choudhury, T., Hauser, M., Kane, J., Merrill, M., Scherer, E. A., Tseng, V. W. S., & Ben-Zeev, D. (2016). CrossCheck: Toward passive sensing and detection of mental health changes in people with schizophrenia. *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 886–897. https://doi.org/10.1145/2971648.2971740

Wang, W., & Griswold, M. E. (2016). Estimating overall exposure effects for the clustered and censored outcome using random effect Tobit regression models. *Statistics in Medicine*, *35*(27), 4948–4960. https://doi.org/10.1002/sim.7045

Weijers, J., Viechtbauer, W., Eurelings-Bontekoe, E., & Selten, J.-P. (2018). Reported Childhood Abuse and Stress Reactivity in Psychosis: A Conceptual Replication and Exploration of Statistical Approaches. *Frontiers in Psychiatry*, *9*, 639. https://doi.org/10.3389/fpsyt.2018.00639

Zelenski, J. M., & Larsen, R. J. (2000). The Distribution of Basic Emotions in Everyday Life: A State and Trait Perspective from Experience Sampling Data. *Journal of Research in Personality*, *34*(2), 178–197. https://doi.org/10.1006/jrpe.1999.2275

7

General Discussion

Contributions and Implications

This thesis addressed fundamental methodological and applied challenges in ESM research, contributing to systematic approaches to ESM measurement development, analytical techniques for complex data structures, and solutions for common distributional problems. These methodological contributions provide both theoretical frameworks and practical tools that enhance the rigor and accuracy of ESM research. Throughout this work, several themes emerged: the importance of distinguishing between-person differences from within-person processes; the value of multi-method approaches combining self-reports, informant-reports, and behavioral observations; the need for transparent, systematic procedures in measurement development; and the recognition of the interplay between stable and dynamic processes, requiring frameworks that can accommodate both structure and process. The discussion that follows focuses on the main findings of the current work and their implications within the broader context of the ESM methodology, as well as on suggestions for future study directions.

Systematic Bottom-Up Measurement Development

One of the methodological innovations of this thesis lies in the systematic, bottom-up approach to developing a new ESM-compatible measurement scale. Traditional psychometric instruments, designed for retrospective trait assessment, are often incompatible with the demands of intensive longitudinal research, which requires brief, ecologically valid measures sensitive to within-person fluctuations (Wright & Zimmermann, 2019).

The measurement development process demonstrated across **Chapters 2-4** represents a paradigm shift toward empirically grounded, participant-centered construction. The development of the ESM questionnaire initially involved constructing a retrospective measure through a systematic lexical approach, which included extensive item generation from multiple sources, the use of a natural language processing algorithm, expert rater classifications, and rigorous validation both through psychometric tools and behavioral measures. Subsequently,

Chapter 4 describes the creation and validation of the ESM measurement scale through a systematic examination of the factor structure at both the within-person (WP) and between-person (BP) levels, using multilevel confirmatory factor analysis. This addresses a fundamental gap in ESM research, where most studies fail to report appropriate psychometric indices for intensive longitudinal measurements (Hall et al., 2021). The reliability estimation framework, which computes separate coefficients for the WP and BP components using multilevel factor analysis (Geldhof et al., 2014), provides a template for future ESM measure validation. In addition, the findings of our study showed that reliability can differ substantially across levels, which demonstrates the importance of level-specific evaluation, which is also highlighted in the literature (Leertouwer et al., 2025).Our validation approach extended beyond traditional cross-sectional methods to examine dynamic relationships between the measured constructs and the predicted behavior at the momentary and lagged levels. We demonstrated that momentary goal importance related to honesty was able to predict both contemporaneous and lagged honesty states, and that changes in goals during the ESM phase predicted trait-level change from the initial to the final assessment. An analytical framework that employed both multilevel modeling with explicit disentangling the between WP and BP components (Wang & Maxwell, 2015) and latent growth modeling represents a comprehensive validation strategy, capturing the temporal dynamics essential to intensive longitudinal research.

The emphasis on cross-sectional measurement development in **Chapters 2** and **3** may initially appear tangential to the thesis's overarching focus on advances in experience sampling methodology. However, this sequential approach reflects a fundamental but often overlooked prerequisite for rigorous ESM research: the necessity of establishing psychometrically sound retrospective measures before adapting constructs to momentary assessment formats. The ESM literature has increasingly recognized that directly translating trait-level items to state-level formats without systematic validation can introduce substantial measurement artifacts,

including altered factorial structures, unstable psychometric properties across temporal resolutions, and conceptual ambiguities regarding what respondents are actually reporting when asked about "right now" versus "in general" (Finnigan & Vazire, 2018; Horstmann & Ziegler, 2020). By first developing comprehensive cross-sectional measures through transparent, bottom-up procedures (**Chapters 2-3**) before systematically adapting them to ESM contexts (**Chapter 4**), this thesis addresses a critical methodological gap: ensuring that momentary measures capture theoretically meaningful within-person variation rather than merely reflecting measurement error or respondent confusion about temporal framing. This progression from stable trait structure to motivational architecture to dynamic state fluctuations demonstrates that advances in intensive longitudinal assessment depend fundamentally on clarity regarding what is being measured at the cross-sectional level.

The bottom-up, participant-centered methodology employed throughout **Chapters 2-4** offers a generalizable template for ESM instrument development that could be applied to other constructs: systematic lexical or goal elicitation to ensure ecological validity, computational text analysis combined with expert human judgment to classify content domains, hierarchical psychometric validation across temporal scales, and behavioral criterion validation to confirm that momentary measures capture functionally meaningful variation. In this sense, while Honesty-Humility serves as the empirical vehicle, the methodological contributions extend well beyond this specific trait domain to address fundamental questions about how personality constructs should be operationalized for intensive longitudinal research. The integrated measurement framework demonstrated across these chapters thus represents an advance in ESM methodology itself—providing researchers with a systematic procedure for developing theoretically grounded, ecologically valid, and behaviorally predictive momentary measures rather than relying on ad hoc item adaptation or unvalidated trait-to-state translations.

Latent Growth Curve Modeling: The application of latent growth modeling (**Chapter 4**) to examine personality change processes demonstrates sophisticated SEM approaches for ESM data. The specification of intercept and slope factors from daily aggregated scores, combined with mediation modeling to examine pathways from goal changes to trait changes through state changes, provides a methodological template for examining bottom-up developmental processes. The finding of mediation (changes in goals predicted trait change through changes in states) demonstrates the importance of examining multiple pathways in developmental models. This approach has sometimes been seen as problematic, as it was claimed that in many EMA studies the constructs of interest may exhibit largely unsystematic "noise-like" fluctuations that might be poorly suited for growth curve modeling approaches (e.g., Losardo et al., 2024). However, our study, together with a relatively small but growing body of literature on growth modeling applied to ESM studies (e.g., Burgess-Hull et al., 2021; Di Sarno et al., 2023; Runyan et al., 2024; Wessel et al., 2019), has shown that this methodology provides particularly interesting insights into the process of change, especially when measures are taken to reduce the contribution of momentary nuisance, in our case by creating daily composite scores.

Multi-Level Understanding of Honesty-Humility: Integrating Traits, Motivations, and States

Chapters 2, 3, and **4** collectively demonstrate that Honesty-Humility operates as a multi-level personality construct with distinct manifestations across different temporal scales, from stable trait dispositions to dynamic momentary states. This integrated examination reveals that a comprehensive understanding of honest and dishonest tendencies requires simultaneous consideration of what individuals are like on average (trait level), what goals they pursue (motivational level), and how they fluctuate from moment to moment in daily life (state level). The lexical foundation established in **Chapter 2** through the Adjective Checklist of Honesty

(ACH) provided crucial conceptual clarification by empirically demonstrating that Truthfulness and Sincerity are functionally indistinguishable in natural language use, resolving a theoretical ambiguity that had persisted in prior HEXACO conceptualizations. Furthermore, the robust four-facet structure of ACH (Sincerity, Fairness, Modesty, Greed-Avoidance) demonstrates that even within a brief, adjective-based format, the theoretical complexity of Honesty-Humility can be preserved. **Chapter 3'**s motivational analysis extended this understanding by revealing that the psychological architecture underlying honest and dishonest behavior is not merely bipolar—individuals do not simply possess more or less motivation for honesty—but rather involves two partially independent motivational systems. The moderate negative correlation between Goals for Honesty and Goals for Dishonesty indicates that people can simultaneously endorse both approach-oriented strivings toward honest conduct and avoidance-oriented aversions to dishonest behavior, or conversely, may lack strong motivation in either direction. This finding has profound implications for understanding moral decision-making: it suggests that enhancing honest motivation does not automatically diminish dishonest impulses, and vice versa, which carries important consequences for intervention design and behavioral prediction.

From a contextual perspective, the 17 classes of Goals for Honesty encompass a remarkably diverse motivational landscape that extends well beyond simple truth-telling or rule-following. Several goals reflect internalized moral commitments and identity concerns—including authenticity, being true to one's ideals and principles, and being satisfied with oneself—suggesting that honest behavior is partly motivated by self-concept maintenance and the psychological need for integrity between values and actions. Other goals reveal interpersonal and prosocial orientations, such as helping others, making others feel loved, making others happy, and maintaining stability of relationships, indicating that honesty serves not merely as a constraint against wrongdoing but as an active facilitation of positive social bonds and others'

well-being. Notably, several goals reflect protective or defensive motivations—avoiding being deceived, avoiding hurting someone, and testing someone—which highlight that honest conduct involves not only expressing truth but also vigilance against interpersonal exploitation and harm prevention. In contrast, the four classes of Goals for Dishonesty form a more tightly organized cluster centered on instrumental gain and interpersonal antagonism: getting what you want at any cost (exploitation and advantage-seeking), lying (active deception for benefit or self-protection), hurting someone (intentional harm), and manipulating others (strategic influence for personal convenience). The substantive content of these goals challenges traditional conceptualizations of Honesty-Humility that emphasize humility, modesty, and fairness at the trait level, revealing that at the motivational level, honesty is equally driven by authenticity, autonomy, prosocial concern, and relationship investment. This finding has important implications for understanding how stable dispositions translate into behavioral tendencies: individuals high in trait Honesty-Humility may arrive at honest behavior through diverse motivational pathways.

The transition from cross-sectional trait and motivational measures (**Chapters 2-3**) to intensive longitudinal state assessment (**Chapter 4**) revealed critical insights into the temporal dynamics and contextual sensitivity of Honesty-Humility that would remain invisible in traditional research designs. **Chapter 4** demonstrated that momentary honest and dishonest states show substantial within-person variability across time and situations, with individuals fluctuating considerably around their own average levels depending on contextual demands, social interactions, and momentary goal activation. The successful adaptation of both trait-descriptive adjectives and motivational goals to the ESM format confirmed that the constructs identified through bottom-up elicitation methods possess sufficient ecological validity to capture meaningful variation in everyday experience. Importantly, the multilevel structure of ESM data allowed for empirical decomposition of between-person differences (stable

individual differences in average honesty levels) from within-person processes (how individuals deviate from their own typical levels), addressing a fundamental limitation of cross-sectional research that conflates these analytically distinct sources of variation. The finding that momentary dishonest states predicted subsequent dishonest behavior within relatively short time windows, with trait measures preserving limited but still predictive value even accounting for momentary states, has theoretical implications for an important personality-related debate: rather than asking whether traits or situations better predict behavior, the multi-level framework demonstrates that trait-like stability and state-like variability coexist within the same individuals.

Together, these three chapters provide empirical support for integrative personality frameworks—particularly Whole Trait Theory (Fleeson & Jayawickreme, 2015) and the density distribution approach to personality description—by demonstrating that traits are best understood not as categorical essences but as probability distributions of state occurrences shaped by stable individual differences in density parameters. The systematic progression from lexical trait structure (**Chapter 2**) to motivational mechanisms (**Chapter 3**) to momentary state and motivational dynamics (**Chapter 4**) operationalizes the theoretical proposition that personality traits represent explanatory abstractions that summarize regularities in lower-level processes rather than causal forces that directly produce behavior. The specific case of Honesty-Humility proved particularly instructive because honest and dishonest behaviors carry clear moral and social consequences, creating strong situational pressures that interact with individual dispositions in ways that are less pronounced for other traits. Uncovering that Goals for Dishonesty predicted actual cheating behavior beyond trait measures, while Goals for Honesty more strongly predicted authentic self-presentation, suggests that the motivational level of analysis captures intention-behavior links that trait descriptions miss. Moreover, the demonstration that bottom-up, participant-generated content can successfully capture

personality constructs challenges the dominance of theory-driven, top-down instrument development in personality assessment. The convergence of lexically derived traits, goal-elicited motivations, and ESM-adapted states and goals around a coherent Honesty-Humility construct suggests that personality structure emerges from natural language and lived experience in ways that align with—but are not wholly determined by—purely theoretical models, supporting a more empirically grounded approach to personality theory construction.

Advanced Multilevel Modeling for Complex Data Structures

The analytical approaches demonstrated throughout this thesis showcase sophisticated multilevel modeling techniques specifically adapted for ESM data, including temporal dependencies, dyadic interdependence, and complex nesting structures. **Chapter 5** introduced methodological innovations for dyadic ESM data through an extended APIM (or even more precisely, L-APIM) framework (Gistelinck & Loeys, 2019). Key innovations of this chapter included: bidirectional modeling of partner effects while controlling for autoregressive processes and cross-valence actor effects, systematic examination of both contemporaneous and lagged effects with appropriate temporal modeling (excluding overnight lags, handling missed assessments), and robust variance estimation (Pustejovsky & Tipton, 2016) to address biased standard errors when random slopes cannot be estimated due to model convergence issues.

We implemented our extended L-APIM within a three-level hierarchy (occasions nested within individuals, individuals nested within couples). The appropriate number of levels for dyadic ESM models remains debated in the literature. Laurenceau and Bolger (2005) argued that distinguishable dyads require only two levels—observations and couples—because the individual level becomes saturated when between-person variability is fully captured by the role-defining variable within the couple (in our case, sex). Similarly, Iida et al. (2023) recommended treating person and time as crossed rather than nested factors in longitudinal

dyadic frameworks. Conversely, other researchers have emphasized the importance of explicitly modeling individual-level variability within dyadic contexts (del Rosario & West, 2025; Gaudreau et al., 2010).

To capture all relevant sources of variability (within-person, between-person, and between-couple) while avoiding overparameterization that would result from estimating within-couple random effects with only two distinguishable partners, we specified a random structure allowing couples to differ in their means while permitting individuals to vary in their deviations from their couple's mean. In our study, between-couple variability exceeded between-person variability in deviations from the couple mean, although the reverse pattern could emerge in different cases. Estimating these random components provides valuable information about variance sources and should not be omitted in dyadic intensive longitudinal analyses.

Innovative Solutions for Distributional Challenges

Chapter 6 addressed one of the major methodological challenges in ESM research: the analysis of skewed data exhibiting floor or ceiling effects. Standard analytic approaches (mainly linear mixed-effects models) applied to such data risk misrepresenting variability, inflating or obscuring associations, and undermining results (Dejonckheere et al., 2024; Terluin et al., 2016; Von Klipstein et al., 2023). Proposed in **Chapter 6**, our approach introduces tobit models as a principal solution that both addresses statistical distortions and reframes how skewed ESM data are understood. Though tobit models were previously applied in the social sciences, this was mainly done in cross-sectional studies and, thus, not within a mixed-effects modeling framework (e.g., McBee et al., 2010), or in intensive longitudinal studies applied to outcome variables of a different (non-psychological) nature (e.g., time) (Petersen et al., 2015; Sonnenberg et al., 2012). Only a very limited number of studies have actually applied mixed-effects tobit models to psychological ESM data (Pries et al., 2019; Savla et al., 2008), but even

those studies do not provide enough rationale for using the model, nor demonstrate its performance advantages compared to the standard linear model.

Several strategies have been suggested before to manage floor and ceiling effects in ESM. These include modifying the study design (e.g., focusing on clinical samples or prompting during intense experiences; Kalokerinos et al., 2020), altering measurement approaches (relative assessments; Dejonckheere et al., 2024), or applying statistical workarounds such as variability indices (Mestdagh et al., 2018), generalized linear models with alternative distributions (Terluin et al., 2016), or dichotomization for logistic regression (Bailen et al., 2022; Tschacher & Lienhard, 2021). However, these solutions either restrict generalizability, discard information, or bypass rather than resolve the underlying problem of bounded scales. Unlike dichotomization, which sacrifices meaningful variation, tobit models retain sensitivity to subtle changes, allowing the detection of clinical patterns that were previously obscured. Moreover, the tobit-model approach provides not only a technical improvement but also deeper theoretical insight into symptom dynamics.

The conceptualization of skewed ESM data as censored manifestations of underlying latent variables represents a novel approach in treating skewed data, recognizing that behind minimum or maximum scores, there might be subtle differences in latent psychological states that bounded measurement scales cannot capture. The artificial censoring approach applied in our study — progressively subtracting constants from normally distributed positive affect to examine parameter recovery — provided a rigorous framework for evaluating analytical approaches under controlled conditions. Results demonstrated that mixed-effects tobit models maintained stable parameter estimates while standard linear models showed deteriorating performance, with slope estimates approaching zero under severe censoring conditions. When applied to naturally skewed negative affect data, tobit models consistently estimated larger

effect sizes, greater between-person variability, and weaker intercept-slope correlations compared to standard linear models. These findings suggest that many ESM studies using standard approaches may systematically underestimate relationship strengths and individual differences, particularly in constructs prone to floor effects. The tobit framework can be naturally extended beyond negative affect to many ESM constructs, including psychopathological symptoms, specific emotions, and behavioral tendencies, as well as other areas where observations are likely to exhibit a skewed distribution.

The recommendation of mixed-effects tobit models for handling skewed ESM data requires, however, their comparison to other analytical strategies that researchers have employed to address distributional challenges. Several alternative statistical approaches exist, each with distinct assumptions and limitations. Log-transformation and generalized linear models with log-normal distributions attempt to normalize skewed data but fundamentally alter the scale and interpretation of effects, making coefficients difficult to interpret in terms of the original metric and potentially introducing bias when back-transforming predictions (Manning & Mullahy, 2001). Moreover, log-transformations cannot accommodate zero values without arbitrary constant additions. Beta regression (Ferrari & Cribari-Neto, 2004) offers an alternative for bounded continuous data by modeling outcomes on the (0,1) interval, but it explicitly excludes boundary values and thus cannot handle the floor and ceiling effects that are precisely the problem in skewed ESM data—participants who report the minimum scale value must be artificially recoded or excluded, discarding potentially meaningful information. Hurdle or two-part models (Cragg, 1971) decompose the data-generating process into a binary component (whether the outcome exceeds the boundary) and a continuous component (the magnitude conditional on exceeding the boundary), which can be appropriate when zero or minimum values reflect a qualitatively different process (e.g., true absence of an experience) than non-zero values. However, for many psychological constructs assessed via

ESM—such as negative affect, where a minimum rating may reflect very low but not absent distress—the hurdle framework imposes a categorical distinction that may be psychologically artificial. In contrast, the tobit framework assumes a continuous latent variable that is simply observed imperfectly due to measurement censoring, which aligns more naturally with the theoretical understanding that rating scales provide bounded windows into underlying psychological continua. The tobit approach preserves all observations without transformation or exclusion, maintains the original metric for interpretation, and explicitly models the censoring mechanism that generates the distributional distortion.

Limitations

Several design and sample-related constraints limit the generalizability of the findings of this work. All studies relied on demographically homogeneous samples — predominantly young, highly educated Italian participants — which restricts external validity to diverse populations varying in age, education, cultural background, and socioeconomic status. **Chapter 5,** examining emotional interdependence, was particularly constrained by a relatively small sample of 76 heterosexual couples, which could have limited statistical power to detect modest partner effects and precluded estimation of complex random effects structures due to convergence issues. The measurement development studies (**Chapters 2-4**), while employing rigorous bottom-up procedures, were conducted exclusively in an Italian population, raising questions about the cross-cultural and cross-linguistic validity of the lexical findings. Additionally, the relatively short observation window (15 days of ESM study) for assessing the trait and motivation change (**Chapter 4**) may be insufficient to capture longer-term developmental processes or more gradual personality changes. The reliance on self-report measures does not allow for ruling out shared method variance and social desirability biases, particularly for evaluative constructs like Honesty-Humility, though some validation did include behavioral measures (**Chapter 3**).

The use of fixed-interval prompting schedules in **Chapters 4** and **5**, rather than random or event-contingent designs, warrants explicit justification given ongoing debates about optimal sampling strategies in ESM research. Our decision to employ fixed assessment times was driven by multiple methodological and practical considerations. First, the theoretical focus on lagged within-person processes—particularly the temporal dynamics of honest and dishonest states in **Chapter 4** and emotional interdependence in **Chapter 5**—required temporally comparable intervals between assessments to ensure that autoregressive and cross-lagged effects reflected equivalent time lags rather than varying intervals that could confound temporal patterns with interval heterogeneity (Hamaker & Wichers, 2017). Second, fixed-time schedules have been shown to enhance compliance rates compared to random-interval designs, particularly in studies requiring sustained participation over multiple weeks (Vachon et al., 2019), which was critical for our extended 15-day and 30-day protocols. Third, the dyadic design in **Chapter 5** necessitated simultaneous completion by both partners to enable meaningful assessment of contemporaneous emotional interdependence, a requirement that fixed schedules fulfill more reliably than randomized approaches, where partners might receive prompts at different times (e.g., Bar-Kalifa & Sened, 2020). While fixed schedules potentially introduce predictability that could theoretically increase reactivity or anticipatory responses, empirical evidence suggests that such reactivity effects are generally modest and do not invalidate EMA data quality or validity (Hufford et al., 2002). Nevertheless, future research could benefit from exploring whether the patterns observed in our studies replicate under alternative sampling schemes, particularly for constructs where anticipatory effects might be more pronounced, such as dishonest behavior.

Study-specific limitations further constrain the interpretation of the findings. The studies were carried out in Italy, which posed some additional challenges. For example, in validating the Adjective Checklist of Honesty (**Chapter 2**), we had to rely on self-report

criterion measures for which an official Italian translation was not available, and we therefore implemented a rigorous back-translation procedure. Although this approach adheres to recommended practices, the lack of formally validated Italian versions should be acknowledged as a limitation. In the goals questionnaire development (**Chapter 3**), a substantial portion of participants had to be excluded from the behavioral validation task because they correctly identified the deception, significantly reducing the effective sample size and statistical power for behavioral analyses. In the dyadic study (**Chapter 5**), we had to exclude approximately one-third of the measurements from the moderation analysis due to discrepancies between the partners' responses regarding their contact with each other. This study faced additional measurement challenges, including a contact indicator that produced severe imbalance (over 90% "in-contact" observations) and failed to capture nuanced interaction qualities that might moderate emotional dynamics. The analysis of affect as broad composites (positive and negative affect) likely obscured discrete emotion dynamics, while the marked skewness of negative affect complicated standard model assumptions.

From a methodological validation perspective, the proposed innovations require more extensive testing before they can be widely adopted. More specifically, the mixed-effects tobit modeling approach (**Chapter 6**) was demonstrated on only a single ESM dataset and has not been validated against external criteria, benchmarked against alternative approaches for skewed data (such as gamma or inverse-Gaussian generalized linear mixed models, or hurdle models), or assessed within a simulation study on data with known properties, which would provide more definitive evidence of model accuracy and robustness. The computational complexity of mixed-effects tobit models implemented within the Bayesian framework requires thousands of iterations and substantial memory, and thus might present practical implementation barriers.

Future research should address these limitations through replication across diverse

samples and cultural contexts, validation with simulated data of known properties, systematic comparison with alternative modeling approaches, experimental manipulation of goals to establish causal mechanisms, and development of more accessible computational implementations.

Future Methodological Directions

Several methodological avenues warrant continued development based on the innovations and limitations identified in this thesis.

First, the convergence challenges encountered with complex multilevel models highlight the need for improved algorithms, alternative software implementations, and clearer guidelines for model specification. Future work should explore alternative software solutions for estimating multilevel dyadic models accounting for all random effects, including random slopes, as well as maximum likelihood implementations of tobit mixed-effects models to reduce computational demands while maintaining analytical advantages.

Second, while this thesis advanced multilevel psychometric evaluation, the field needs continued development in dynamic measurement models, including methods for assessing reliability across different time scales, quantifying measurement invariance over time, and developing comprehensive guidelines for providing validity evidence.

As ESM research incorporates passive sensing, physiological monitoring, and other technological innovations (Mehl et al., 2024), the methodological frameworks developed should be extended to handle multimodal data streams, sensor fusion approaches, and more complex measurement models combining active and passive indicators.

Finally, the future of ESM methodology lies in adaptive, personalized designs. Machine learning approaches for detecting non-relevant items, optimizing sampling frequency based on

individual patterns, and tailoring assessment content to maximize information while minimizing burden represent important methodological frontiers (see e.g., Khanshan et al., 2026).

Conclusions

This thesis makes substantial methodological contributions to ESM research, addressing fundamental challenges in measurement, analysis, and interpretation. The integrated methodological approach demonstrated — combining systematic measurement development, complex multilevel modeling, and innovative solutions to distributional challenges — provides a foundation for more accurate, nuanced, and replicable ESM research. Through continued methodological innovation and rigorous application, ESM research can achieve its potential for revealing the dynamic processes underlying human psychology in naturalistic contexts.

References:

Bailen, N. H., Koval, P., Strube, M., Haslam, N., & Thompson, R. J. (2022). Negative emotion and nonacceptance of emotion in daily life. *Emotion*, *22*(5), 992–1003. https://doi.org/10.1037/emo0000898

Bar-Kalifa, E., & Sened, H. (2020). Using Network Analysis for Examining Interpersonal Emotion Dynamics. Multivariate Behavioral Research, 55(2), 211–230. https://doi.org/10.1080/00273171.2019.1624147

Burgess-Hull, A. J., Smith, K. E., Schriefer, D., Panlilio, L. V., Epstein, D. H., & Preston, K. L. (2021). Longitudinal patterns of momentary stress during outpatient opioid agonist treatment: A growth-mixture-model approach to classifying patients. *Drug and Alcohol Dependence*, *226*, 108884. https://doi.org/10.1016/j.drugalcdep.2021.108884

Cragg, J. G. (1971). Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. *Econometrica*, *39*(5), 829. https://doi.org/10.2307/1909582

Dejonckheere, E., Penne, I., Briels, L., & Mestdagh, M. (2024). For Better or for Worse? Visualizing Previous Intensity Levels Improves Emotion (Dynamic) Measurement in Experience Sampling. *Psychological Assessment*, *36*(3), 215–234. https://doi.org/10.1037/pas0001296

del Rosario, K. S., & West, T. V. (n.d.). A Practical Guide to Specifying Random Effects in Longitudinal Dyadic Multilevel Modeling. *Advances in Methods and Practices in Psychological Science*, *8*(3), 1–36. https://doi.org/10.1177/2515245925135

Di Sarno, M., Costantini, G., Richetin, J., Preti, E., & Perugini, M. (2023). Why are you (un)conscientious? The dynamic interplay of goals, states, and traits in everyday life. *Journal of Personality*, *91*(4), 977–991. https://doi.org/10.1111/jopy.12701

Ferrari, S., & Cribari-Neto, F. (2004). Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, *31*(7), 799–815. https://doi.org/10.1080/0266476042000214501

Finnigan, K. M., & Vazire, S. (2018). The incremental validity of average state self-reports over global self-reports of personality. *Journal of Personality and Social Psychology*, *115*(2), 321–337. https://doi.org/10.1037/pspp0000136

Fleeson, W., & Jayawickreme, E. (2015). Whole Trait Theory. *Journal of Research in Personality*, *56*, 82–92. https://doi.org/10.1016/j.jrp.2014.10.009

Gaudreau, P., Fecteau, M.-C., & Perreault, S. (2010). Multi-Level Modeling of Dyadic Data in

Sport Sciences: Conceptual, Statistical, and Practical Issues. *Measurement in Physical Education and Exercise Science*, *14*(1), 29–50. https://doi.org/10.1080/10913670903455017

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72–91. https://doi.org/10.1037/a0032138

Gistelinck, F., & Loeys, T. (2019). The Actor–Partner Interdependence Model for Longitudinal Dyadic Data: An Implementation in the SEM Framework. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(3), 329–347. https://doi.org/10.1080/10705511.2018.1527223

Hall, M., Scherner, P. V., Kreidel, Y., & Rubel, J. A. (2021). A Systematic Review of Momentary Assessment Designs for Mood and Anxiety Symptoms. *Frontiers in Psychology*, *12*, 642044. https://doi.org/10.3389/fpsyg.2021.642044

Hamaker, E. L., & Wichers, M. (2017). No Time Like the Present: Discovering the Hidden Dynamics in Intensive Longitudinal Data. *Current Directions in Psychological Science*, *26*(1), 10–15. https://doi.org/10.1177/0963721416666518

Horstmann, K. T., & Ziegler, M. (2020). Assessing Personality States: What to Consider when Constructing Personality State Measures. *European Journal of Personality*, *34*(6), 1037–1059. https://doi.org/10.1002/per.2266

Hufford, Michael R., Alan L. Shields, Saul Shiffman, Jean Paty, and Mark Balabanis. "Reactivity to Ecological Momentary Assessment: An Example Using Undergraduate Problem Drinkers." Psychology of Addictive Behaviors 16, no. 3 (2002): 205–11. 2002-17901-003. https://doi.org/10.1037/0893-164X.16.3.205

Iida, M., Savord, A., & Ledermann, T. (2023). Dyadic longitudinal models: A critical review. *Personal Relationships*, *30*(2), 356–378. https://doi.org/10.1111/pere.12468

Kalokerinos, E. K., Murphy, S. C., Koval, P., Bailen, N. H., Crombez, G., Hollenstein, T., Gleeson, J., Thompson, R. J., Van Ryckeghem, D. M. L., Kuppens, P., & Bastian, B. (2020). Neuroticism may not reflect emotional variability. *Proceedings of the National Academy of Sciences*, *117*(17), 9270–9276. https://doi.org/10.1073/pnas.1919934117

Khanshan, A., Van Gorp, P., & Markopoulos, P. (2026). An Exploration to Enhance Response Rate and Sampling Density During ESM Experiments by Online Supervised Learning and Edge Computing on Smartwatches. In *C. Ardito et al. (Eds.): INTERACT 2025, LNCS 16109* (pp. 71–86). Springer. https://doi.org/10.1007/978-3-032-05002-1_4

Laurenceau, J.-P., & Bolger, N. (2005). Using Diary Methods to Study Marital and Family

Processes. *Journal of Family Psychology*, *19*(1), 86–97. https://doi.org/10.1037/0893-3200.19.1.86

Leertouwer, Ij., Keijsers, L., van Roekel, E., & Schuurman, N. (2025). A Practical Guide to Estimating Reliability of Intensive Longitudinal Data. *Preprint*.

Losardo, D., Chow, S.-M., Panter, A. T., Burkley, M., & Burkley, E. (2024). Ecological Momentary Assessment (EMA) Designs with Planned Missingness. In *M. Stemmler et al. (Eds.), Dependent Data in Social Sciences Research* (2nd ed.). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-56318-8_26

Manning, W. G., & Mullahy, J. (2001). Estimating log models: To transform or not to transform? *Journal of Health Economics*, *20*(4), 461–494. https://doi.org/10.1016/S0167-6296(01)00086-8

McBee, M. (2010). Modeling Outcomes With Floor or Ceiling Effects: An Introduction to the Tobit Model. *Gifted Child Quarterly*, *54*(4), 314–320. https://doi.org/10.1177/0016986210379095

Mehl, M. R., Eid, M., Wrzus, C., Harari, G. M., & Ebner-Priemer, U. W. (2024). Mobile sensing in psychology: Where we are and where we might go from here. In *Mobile sensing in psychology: Methods and applications* (pp. 737–748). The Guilford Press.

Mestdagh, M., Pe, M., Pestman, W., Verdonck, S., Kuppens, P., & Tuerlinckx, F. (2018). Sidelining the mean: The relative variability index as a generic mean-corrected variability measure for bounded variables. *Psychological Methods*, *23*(4), 690–707. https://doi.org/10.1037/met0000153

Petersen, J., Austin, D., Mattek, N., & Kaye, J. (2015). Time Out-of-Home and Cognitive, Physical, and Emotional Wellbeing of Older Adults: A Longitudinal Mixed Effects Model. *PLOS ONE*, *10*(10), e0139643. https://doi.org/10.1371/journal.pone.0139643

Pries, L.-K., Klingenberg, B., Menne-Lothmann, C., Decoster, J., van Winkel, R., Collip, D., Delespaul, P., De Hert, M., Derom, C., Thiery, E., Jacobs, N., Wichers, M., Cinar, O., Lin, B. D., Luykx, J. J., Rutten, B. P. F., van Os, J., & Guloksuz, S. (2020). Polygenic liability for schizophrenia and childhood adversity influences daily-life emotion dysregulation and psychosis proneness. *Acta Psychiatrica Scandinavica*, *141*(5), 465–475. https://doi.org/10.1111/acps.13158

Pustejovsky, J. E., & Tipton, E. (2016). Small sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, *36*(4), 672–683. https://doi.org/10.1080/07350015.2016.1247004

Runyan, J. D., Vermilya, S., St. Pierre, M., Brooks, N. W., Fowler, A., & Brewer, T. (2024).

A mixed methods experience sampling study of a posttraumatic growth model for addiction recovery. *Scientific Reports*, *14*(1), 3511. https://doi.org/10.1038/s41598-024-53740-7

Savla, J., Almeida, D. M., Davey, A., & Zarit, S. H. (2008). Routine Assistance to Parents: Effects on Daily Mood and Other Stressors. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *63*(3), S154–S161. https://doi.org/10.1093/geronb/63.3.S154

Sonnenberg, B., Riediger, M., Wrzus, C., & Wagner, G. G. (2012). Measuring time use in surveys – Concordance of survey and experience sampling measures. *Social Science Research*, *41*(5), 1037–1052. https://doi.org/10.1016/j.ssresearch.2012.03.013

Terluin, B., De Boer, M. R., & De Vet, H. C. W. (2016). Differences in Connection Strength between Mental Symptoms Might Be Explained by Differences in Variance: Reanalysis of Network Data Did Not Confirm Staging. *PLOS ONE*, *11*(11), e0155205. https://doi.org/10.1371/journal.pone.0155205

Tschacher, W., & Lienhard, N. (2021). Mindfulness Is Linked with Affectivity in Daily Life: An Experience-Sampling Study with Meditators. *Mindfulness*, *12*(6), 1459–1472. https://doi.org/10.1007/s12671-021-01615-7

Vachon, H., Viechtbauer, W., Rintala, A., & Myin-Germeys, I. (2019). Compliance and Retention With the Experience Sampling Method Over the Continuum of Severe Mental Disorders: Meta-Analysis and Recommendations. Journal of Medical Internet Research, 21(12), e14475. https://doi.org/10.2196/14475

Von Klipstein, L., Servaas, M. N., Lamers, F., Schoevers, R. A., Wardenaar, K. J., & Riese, H. (2023). Increased affective reactivity among depressed individuals can be explained by floor effects: An experience sampling study. *Journal of Affective Disorders*, *334*, 370–381. https://doi.org/10.1016/j.jad.2023.04.118

Wang, L. (Peggy), & Maxwell, S. E. (2015). On disaggregating between-person and within-person effects with longitudinal data using multilevel models. *Psychological Methods*, *20*(1), 63–83. https://doi.org/10.1037/met0000030

Wessel, J., Bradley, G. L., & Hood, M. (2019). Comparing effects of active and passive procrastination: A field study of behavioral delay. *Personality and Individual Differences*, *139*, 152–157. https://doi.org/10.1016/j.paid.2018.11.020

Wright, A. G. C., & Zimmermann, J. (2019). Applied ambulatory assessment: Integrating idiographic and nomothetic principles of measurement. *Psychological Assessment*, *31*(12), 1467–1480. https://doi.org/10.1037/pas0000685

Addendum

Summary

Experience Sampling Methodology and Its Methodological Advances

The work presented in this thesis aims at exploring, applying, and expanding the methodological foundations of Experience Sampling Method (ESM) research through systematic measurement development, advanced analytical techniques, and innovative solutions for common distributional challenges. ESM, also known as Ecological Momentary Assessment (EMA), has gained increasing prominence in psychological research by enabling the collection of real-time data on variables of interest as they naturally occur in participants' daily lives. However, the methodology presents unique analytical challenges involving complex nested data structures, measurement development needs and a lack of common practice, and data characteristics that necessitate specialized approaches to ensure accurate parameter estimation and valid conclusions.

Following an introduction to the ESM methodology and its challenges, the thesis is divided into three main parts. **Part I** describes the systematic bottom-up development of Honesty-Humility measures across trait, motivation, and state levels. **Part II** explores interpersonal emotion dynamics in romantic relationships using advanced dyadic ESM approaches. **Part III** addresses fundamental methodological issues in ESM data analysis, particularly the problem of censored distributions and bounded psychological constructs.

Systematic Measurement Development

Chapters 2, 3, and 4 form an integrated sequence focused on developing and validating measures of Honesty-Humility suitable for both cross-sectional trait assessment and intensive longitudinal research. These chapters demonstrate a systematic, bottom-up approach that minimizes researcher bias while ensuring ecological validity and psychometric rigor.

Chapter 2 presents the development of the Adjective Checklist of Honesty (ACH), a comprehensive 22-item adjective-based measure capturing the four main facets of Honesty-Humility: Sincerity, Fairness, Modesty, and Greed-Avoidance. Using an extensive corpus of Italian adjectives refined through independent expert raters, this systematic approach yielded important theoretical insights, particularly that Truthfulness and Sincerity are conceptually indistinguishable in natural language use. The scale demonstrated good factorial validity and showed meaningful associations with theoretically relevant constructs, while revealing interesting overlap with conscientiousness in workplace ethics domains.

Chapter 3 extends this approach to uncover the motivational core of Honesty-Humility, developing the Goals for Honesty and Goals for Dishonesty (GH/GD) questionnaire. Through participant-generated goal elicitation from over 9,000 textual responses, classified by experts with the help of natural language processing algorithms, researchers identified 48 goal classes potentially related to Honesty-Humility. The final 78-item questionnaire revealed that Goals for Honesty (GH) and Goals for Dishonesty (GD) represent two distinct, only moderately negatively correlated motivational orientations. Critically, GD consistently predicted actual dishonest behavior in an incentivized cheating task beyond traditional trait measures, while GH related more strongly to authentic living and self-concept integrity.

Chapter 4 completes this measurement sequence by adapting the GH/GD questionnaire for intensive longitudinal research (GH/GD-EMA). In a 15-day ESM/EMA study involving 198 participants who received five daily prompts, multilevel confirmatory factor analyses confirmed the two-factor structure at both the within- and between-person levels. Most importantly, momentary goal importance robustly predicted Honesty-Humility states both contemporaneously and temporally. Using latent growth curve modeling, the study demonstrated that changes in goals during the ESM period predicted trait-level change from

baseline to follow-up through their effects on state changes, providing evidence for bottom-up personality development processes.

Interpersonal Emotion Dynamics

Chapter 5 shifts focus to interpersonal processes, examining emotional interdependence in romantic relationships through an extended Longitudinal Actor-Partner Interdependence Model (L-APIM) framework. Using intensive longitudinal data from 76 couples completing assessments five times daily over 30 days, the study investigated both contemporaneous and temporal patterns of emotional interdependence between partners while controlling for autoregressive effects and cross-valence actor effects.

Results revealed significant emotional interdependence for positive affect both contemporaneously and temporally, indicating that partners' positive emotions show a modest but consistent linkage in daily life. However, no significant partner effects emerged for negative affect, suggesting that emotional transmission patterns might differ by valence. Contact between partners showed a positive main effect on males' positive affect, but did not significantly moderate the strength of emotional interdependence. Substantial between-couple variability in emotional baselines was observed, with partners showing greater similarity in negative affect than positive affect baselines. These findings highlight that emotional interdependence is more complex and heterogeneous than theoretical models often assume, requiring more nuanced, couple-specific approaches.

Methodological Advances

Chapter 6 addresses a fundamental challenge in ESM research: analyzing skewed data that exhibits floor or ceiling effects. Many variables of interest in ESM studies, particularly negative affect and psychopathological symptoms, show highly skewed distributions that can violate assumptions of commonly used statistical models, like linear

mixed-effects models, and lead to biased parameter estimates.

The study proposes conceptualizing skewed observations as censored manifestations of underlying latent variables and demonstrates that mixed-effects tobit models provide superior parameter recovery compared to standard linear mixed-effects models. Through artificial censoring of initially normally distributed positive affect data, the results showed that tobit models maintained stable parameter estimates under progressive censoring conditions, whereas standard models exhibited deteriorating slope estimates that approached zero. When applied to naturally skewed negative affect data, tobit models consistently estimated larger effect sizes, greater between-person variability, and more plausible intercept-slope correlations across all participant groups. These findings suggest that standard approaches may introduce systematic bias in the relationship strengths and individual differences in bounded psychological constructs.

New Horizons

Chapter 7 discusses the main findings and their implications for the future of ESM research. The thesis demonstrates that assessing psychological variability through intensive longitudinal methods is crucial to understanding the dynamic processes that occur in daily life. The systematic measurement development procedures, advanced multilevel modeling techniques, and innovative analytical solutions provide a foundation for more accurate, nuanced, and replicable ESM research. Future developments will likely focus on multimodal data integration, personalized assessment approaches, and continued refinement of analytical techniques for increasingly complex psychological phenomena.

Samenvatting

Experience Sampling Methodology en haar Methodologische Vooruitgang

Het werk gepresenteerd in dit proefschrift heeft als doel de methodologische fundamenten van Experience Sampling Method (ESM) onderzoek te verkennen, toe te passen en uit te breiden door middel van systematische meetontwikkeling, geavanceerde analysetechnieken en innovatieve oplossingen voor veelvoorkomende problemen gerelateerd aan verspreiding. ESM, ook wel bekend als Ecological Momentary Assessment (EMA), is steeds populairder geworden in psychologisch onderzoek vanwege de mogelijkheid van real-time gegevensverzameling over variabelen van belang zoals deze natuurlijk voorkomen in het dagelijks leven van deelnemers. Deze methodologie brengt echter ook unieke analytische uitdagingen met zich mee, zoals complexe, ingenestelde datastructuren. Daarnaast onstaat er een nieuwe behoefte aan meetontwikkeling en standardizatie. Ook is er bij ESM sprake van data-eigenschappen die gespecialiseerde benaderingen vereisen voor een nauwkeurige inschatting van parameters en het waarborgen van valide conclusies.

Na een inleiding over de ESM-methodologie en haar uitdagingen, is het proefschrift verdeeld in drie hoofddelen. **Deel I** beschrijft de systematische bottom-up ontwikkeling van Eerlijkheid-Nederigheid metingen op trait-, motivatie- en toestandsniveaus. **Deel II** verkent de interpersoonlijke emotiedynamiek in romantische relaties met behulp van geavanceerde dyadische ESM-benaderingen. **Deel III** behandelt fundamentele methodologische kwesties in ESM-data-analyse, waaronder het probleem van gecensureerde verdelingen en begrensde psychologische constructen.

Systematische Meetontwikkeling

Hoofdstukken 2, 3 en **4** vormen een geïntegreerde reeks gericht op het ontwikkelen en valideren van metingen van Eerlijkheid-Nederigheid die geschikt zijn voor zowel cross-sectionele trait-beoordeling als intensief longitudinaal onderzoek. Deze hoofdstukken

demonstreren een systematische, bottom-up benadering die onderzoekersbias minimaliseert, waarbij ecologische validiteit en psychometrische nauwkeurigheid gewaarborgd blijven.

Hoofdstuk 2 presenteert de ontwikkeling van de Adjective Checklist of Honesty (ACH), een uitgebreid 22-item adjectief-gebaseerd meetinstrument die de vier hoofdfacetten van Eerlijkheid-Nederigheid vastlegt: Oprechtheid, Eerlijkheid, Bescheidenheid en Hebzucht-Vermijding. Door gebruik te maken van een uitgebreid corpus van Italiaanse adjectieven verfijnd door onafhankelijke en deskundige beoordelaars, leverde deze systematische benadering belangrijke theoretische inzichten op. Een belangrijke uiktomst was dat waarheidsgetrouwheid en oprechtheid conceptueel niet te onderscheiden zijn in natuurlijk taalgebruik. De schaal toonde goede factoriële validiteit, vertoonde betekenisvolle associaties met theoretisch relevante constructen, en onthulde interessante overlappingen met consciëntieusheid in werkplaak-ethiek domeinen.

Hoofdstuk 3 breidt deze benadering uit met de ontwikkeling van de Goals for Honesty and Goals for Dishonesty (GH/GD) vragenlijst, om zo de motivationele kern van Eerlijkheid-Nederigheid bloot te leggen. Met gebruik van deelnemer-gegenereerde doelelicitatie uit meer dan 9.000 tekstuele reacties, geclassificeerd door experts met behulp van natural language processing algoritmen, identificeerden de onderzoekers 48 doelklassen die mogelijk gerelateerd zijn aan Eerlijkheid-Nederigheid. De uiteindelijke 78-item vragenlijst onthulde dat Goals for Honesty (GH) en Goals for Dishonesty (GD) twee individuele, maar slechts matig negatief gecorreleerde, motivationele oriëntaties vertegenwoordigen. Cruciaal is dat GD consistent feitelijk oneerlijk gedrag voorspelde in een geïncentiveerde valsspeel-taak. Dit was zelfs een betere voorspellende factor dan traditionele trait-metingen. GH, daarentegen, was gerelateerd was aan authentiek leven en zelfconcept-integriteit.

Hoofdstuk 4 voltooit deze meet-reeks door de GH/GD-vragenlijst aan te passen voor intensief longitudinaal onderzoek (GH/GD-EMA). In een 15-daagse ESM/EMA-studie met

198 deelnemers die elke dag vijf prompts ontvingen, bevestigden multilevel confirmatorische factoranalyses de twee-factorenstructuur op zowel binnen- als tussen-persoonsniveaus. De belangrijkste uitkomst is dat kortstondige belangrijkheid van doelen robuust de toestand van van Eerlijkheid-Nederigheid voorspelde, zowel gelijktijdig als tussen opeenvolgende metingen (temporeel). Met behulp van latent growth curve modeling, toonde de studie aan dat veranderingen in doelen tijdens de ESM-periode trait-niveau, ook verandering van baseline naar follow-up voorspelden door hun effecten op toestandsveranderingen. Dit levert bewijs voor bottom-up persoonlijkheidsontwikkelingsprocessen.

Interpersoonlijke Emotiedynamiek

Hoofdstuk 5 verschuift de focus naar interpersoonlijke processen, waarbij emotionele onderlinge afhankelijkheid in romantische relaties wordt onderzocht via een uitgebreid Longitudinal Actor-Partner Interdependence Model (L-APIM) raamwerk. Met behulp van intensieve longitudinale gegevens van 76 koppels die 30 dagen lang vijf keer per dag beoordelingen voltooiden, onderzocht de studie zowel gelijktijdige als temporele patronen van emotionele onderlinge afhankelijkheid tussen partners. Deze effecten werden gecorrigeerd voor autoregressieve effecten en cross-valence actor-effecten.

De resultaten onthulden een significante emotionele onderlinge afhankelijkheid voor positief affect, zowel gelijktijdig als temporeel. Dit geeft aan dat de positieve emoties van partners een bescheiden, maar consistente, koppeling vertonen in het dagelijks leven. Er waren echter geen significante partner-effecten voor negatief affect, wat suggereert dat emotionele transmissiepatronen kunnen verschillen per valentie. Contact tussen partners vertoonde een positief effect op het positieve affect van mannen, maar deze variabele was geen significante moderator van de sterkte van emotionele onderlinge afhankelijkheid. Er werd wel substantiële tussen-koppel variabiliteit in emotionele basislijnen waargenomen, waarbij partners een grotere gelijkenis vertoonden in negatieve affect basislijnen vergeleken met de positieve affect

basislijnen. Deze bevindingen benadrukken dat emotionele onderlinge afhankelijkheid complexer en heterogener is dan theoretische modellen vaak veronderstellen, en dat dit meer genuanceerde, koppel-specifieke benaderingen vereist.

Methodologische Vooruitgang

Hoofdstuk 6 behandelt een fundamentele uitdaging in ESM-onderzoek: het analyseren van scheve gegevens die vloer- of plafondeffecten vertonen. Veel variabelen die van belang zijn in ESM-studies, en met name negatief affect en psychopathologische symptomen, vertonen zeer scheve verdelingen. Dergelijke verdelingen schenden de onderliggende aannames van veelgebruikte statistische modellen, zoals lineaire mixed-effects modellen, wat kan leiden tot een vertekende inschatting van relevante parameters.

De studie stelt voor om scheve observaties te conceptualiseren als gecensureerde manifestaties van onderliggende latente variabelen. De resultaten demonstreren dat mixed-effects tobit-modellen superieur parameterherstel bieden in vergelijking met standaard lineaire mixed-effects modellen. Door kunstmatige censurering van aanvankelijk normaal positief affect-gegevens, toonden de resultaten aan dat tobit-modellen stabiele inschattingen van parameters behielden onder progressieve censureringsomstandigheden. Hiertot in tegenstelling, demonstreerde de standaard modellen een verslechterende inschatting van de helling, die nul benaderden. Wanneer deze toegepast worden op natuurlijk scheve negatieve affect-gegevens, schatten tobit-modellen consistent grotere effectgroottes, grotere tussen-persoon variabiliteit en meer plausibele intercept-helling correlaties over alle deelnemersgroepen. Deze bevindingen suggereren dat de standaard benaderingen systematische vertekening kunnen introduceren in de relatiesterktes en individuele verschillen in begrensde psychologische constructen.

Nieuwe Horizonten

Hoofdstuk 7 bespreekt de belangrijkste bevindingen en hun implicaties voor de

toekomst van ESM-onderzoek. Het proefschrift demonstreert dat het beoordelen van psychologische variabiliteit door intensieve longitudinale methoden cruciaal is voor het begrijpen van de dynamische processen die zich in het dagelijks leven voordoen. De systematische meetontwikkelingsprocedures, geavanceerde multilevel modelleertechnieken en innovatieve analytische oplossingen bieden een basis voor nauwkeuriger, genuanceerder en repliceerbaarder ESM-onderzoek. Toekomstige ontwikkelingen zullen zich waarschijnlijk richten op multimodale data-integratie, gepersonaliseerde beoordelingsbenaderingen en voortdurende verfijning van analysetechnieken voor steeds complexere psychologische fenomenen.

Sintesi

L'Experience Sampling Method e i suoi progressi metodologici

Il lavoro presentato in questa tesi mira a esplorare, applicare ed espandere le fondamenta metodologiche dell'Experience Sampling Method (ESM) attraverso lo sviluppo sistematico di strumenti di misurazione, l'applicazione di tecniche analitiche avanzate e l'elaborazione di soluzioni innovative alle comuni sfide legate alle distribuzioni dei dati. L'ESM, noto anche come Ecological Momentary Assessment (EMA), è un approccio metodologico che ha acquisito crescente importanza nella ricerca psicologica in quanto consente la raccolta di dati in tempo reale sulle variabili di interesse nel momento stesso in cui si manifestano nella vita quotidiana dei partecipanti. Tale metodologia presenta sfide analitiche proprie, tra cui la produzione di dati caratterizzati da una struttura annidata e complessa, che richiedono approcci specializzati per garantire stime parametriche accurate e la possibilità di trarre conclusioni valide.

Dopo un'introduzione alla metodologia ESM e alle sue principali sfide, la tesi si suddivide in tre sezioni principali. Nella **Parte I** viene descritto lo sviluppo sistematico, "dal basso verso l'alto", di misure di Onestà-Umiltà a livello di tratto, motivazione e stato. Nella **Parte II** si esplorano le dinamiche emotive interpersonali nelle relazioni romantiche utilizzando approcci diadici ESM avanzati. Infine, nella **Parte III** si affrontano le questioni metodologiche fondamentali nell'analisi dei dati ESM, con particolare riguardo alla gestione delle distribuzioni censurate e ai costrutti psicologici bounded.

Sviluppo sistematico delle misure

I **Capitoli 2, 3** e **4** costituiscono una sequenza integrata incentrata sullo sviluppo e sulla validazione di misure dell'Onestà-Umiltà adatte sia alla valutazione trasversale dei tratti, sia alla ricerca longitudinale intensiva. Questi capitoli dimostrano un approccio sistematico "dal basso verso l'alto" che minimizza il bias del ricercatore, garantendo al contempo validità ecologica e rigore psicometrico.

Nello specifico, il **Capitolo 2** presenta lo sviluppo dell'Adjective Checklist of Honesty (ACH), una misura di 22 aggettivi che cattura le quattro principali sfaccettature del tratto Onestà-Umiltà: Sincerità, Veridicità, Modestia e Greed-Avoidance (evitamento dell'avidità). Utilizzando un ampio corpus di aggettivi italiani vagliato da esperti indipendenti, questo approccio sistematico ha restituito intuizioni teoriche importanti, mostrando in particolare che "veridicità" e "sincerità" risultano concettualmente indistinguibili nel linguaggio naturale. La scala ha mostrato una buona validità fattoriale e associazioni significative con costrutti teoricamente rilevanti, rivelando inoltre interessanti sovrapposizioni con la coscienziosità nei contesti etici lavorativi.

Nel **Capitolo 3** tale approccio viene esteso all'indagine del nucleo motivazionale dell'Onestà-Umiltà, sviluppando il questionario Goals for Honesty and Goals for Dishonesty (GH/GD). Attraverso l'elicitazione di obiettivi generati dai partecipanti (oltre 9.000 risposte testuali classificate da esperti con l'ausilio di algoritmi di elaborazione del linguaggio naturale), i ricercatori hanno identificato 48 classi di obiettivi potenzialmente legate al costrutto. Il questionario finale, composto da 78 item, ha rivelato che i Goals for Honesty (GH) e i Goals for Dishonesty (GD) rappresentano due orientamenti motivazionali distinti e solo moderatamente negativamente correlati. Coerentemente alle ipotesi, la dimensione GD è risultata essere predittore, oltre delle tradizionali misure di tratto, anche dei comportamenti disonesti effettivi in un compito incentivato di cheating, mentre la dimensione GH è risultata maggiormente associata all'esistenza autentica e all'integrità del concetto di sé.

Il **Capitolo 4** si pone a completamento della sequenza di sviluppo di strumenti di misurazione, presentando l'adattamento del questionario GH/GD alla ricerca longitudinale intensiva (GH/GD-EMA). La struttura bi-fattoriale della scala è stata verificata sia a livello intraindividuale sia a livello interindividuale tramite analisi fattoriali multilivello confermative condotte su dati ottenuti in uno studio ESM/EMA di 15 giorni che ha coinvolto 198 partecipanti

con cinque rilevazioni giornaliere. In particolare, l'importanza momentanea degli obiettivi ha predetto in modo robusto la componente di stato di Onestà-Umiltà a livello contemporaneo e temporale. Utilizzando modelli di *latent growth curve*, lo studio ha dimostrato che i cambiamenti negli obiettivi registrati durante la raccolta dati ESM predicono il cambiamento della componente di tratto di Onestà-Umiltà, dalla baseline al follow-up, attraverso gli effetti sui cambiamenti di stato, fornendo evidenza empirica che lo sviluppo della personalità segua un processo *bottom-up*.

Dinamiche emotive interpersonali

Nel **Capitolo 5** l'attenzione viene spostata sui processi interpersonali, esaminando l'interdipendenza emotiva nelle relazioni romantiche attraverso l'applicazione di un'estensione del Longitudinal Actor-Partner Interdependence Model (L-APIM). Utilizzando dati longitudinali intensivi provenienti da 76 coppie che hanno completato valutazioni cinque volte al giorno per 30 giorni, lo studio ha indagato i pattern di interdipendenza emotiva tra i partner, sia contemporanei sia temporali, controllando per gli effetti autoregressivi e per quelli dell'esperienza emotiva con valenza opposta riportata dalla stessa persona (actor). I risultati hanno rivelato una significativa interdipendenza emotiva per lo stato affettivo positivo sia a livello contemporaneo sia temporale, indicando che le emozioni positive dei partner mostrano un legame modesto ma costante nella vita quotidiana. Tuttavia, non è emersa una significativa interdipendenza tra i partner per lo stato affettivo negativo, suggerendo che i pattern di trasmissione emotiva possono differire a seconda della valenza. Il contatto tra partner ha mostrato un effetto positivo sullo stato affettivo positivo maschile, ma non ha moderato in modo significativo la forza dell'interdipendenza emotiva. È stata inoltre osservata una notevole variabilità tra le coppie nei livelli di base di affettività, con una maggiore somiglianza nei livelli di attivazione affettiva negativa rispetto a quella positiva. Questi risultati evidenziano che l'interdipendenza emotiva è più complessa ed eterogenea di quanto spesso ipotizzato dai

modelli teorici, rendendo necessari approcci più sfumati e specifici a livello diadico.

Progressi metodologici

Nel **Capitolo 6** si affronta una sfida metodologica fondamentale nella ricerca ESM: l'analisi di dati asimmetrici che presentano effetti di pavimento o di soffitto. Molte variabili di interesse negli studi ESM, in particolare lo stato affettivo negativo e i sintomi psicopatologici, mostrano distribuzioni fortemente asimmetriche che possono violare le assunzioni dei modelli statistici comunemente usati, come i modelli lineari misti, portando a stime distorte dei parametri.

Lo studio propone di concettualizzare le osservazioni asimmetriche come manifestazioni censurate di variabili latenti sottostanti e dimostra che i modelli *tobit* a effetti misti offrono una capacità di stima dei parametri superiore rispetto ai modelli lineari standard. Applicando un processo di censura artificiale ai dati sullo stato affettivo positivo con distribuzione normale, i risultati hanno mostrato che i modelli *tobit* mantengono stime parametriche stabili anche sotto condizioni di censura progressiva, mentre i modelli standard producono stime di pendenza via via più deboli fino ad annullarsi. I modelli *tobit*, applicati ai dati sullo stato affettivo negativo con distribuzione asimmetrica, hanno mostrato di riuscire a stimare sistematicamente un *effect size* più grande, maggiore variabilità interindividuale e correlazioni *intercept-slope* più plausibili in tutti i gruppi di partecipanti, rispetto ai modelli standard. Questi risultati suggeriscono che gli approcci standard possono introdurre *bias* sistematici nella stima delle relazioni e delle differenze individuali nei costrutti psicologici *bounded*.

Nuovi orizzonti

Infine, nel **Capitolo 7** vengono discussi i principali risultati e le loro implicazioni per il futuro della ricerca ESM. La tesi dimostra che la valutazione della variabilità psicologica attraverso metodi longitudinali intensivi è fondamentale per comprendere i processi dinamici

che avvengono nella vita quotidiana. Procedure sistematiche di sviluppo delle misure, tecniche avanzate di modellizzazione multilivello e soluzioni analitiche innovative forniscono una base solida per una ricerca ESM più accurata, sfaccettata e replicabile. Si ritiene perciò probabile che gli sviluppi futuri si concentreranno sull'integrazione multimodale dei dati, su approcci di valutazione personalizzati e sul continuo perfezionamento delle tecniche analitiche per affrontare fenomeni psicologici sempre più complessi.

Knowledge Valorization

Societal Relevance

Psychological problems and relationship difficulties are prevalent in society, with significant associated costs for individuals, families, and society as a whole. Traditional approaches often focus on observable consequences (such as pathological symptoms, divorce, or misdeeds) rather than understanding the dynamic processes underlying intra- and interpersonal psychological functioning. From this perspective, the experience sampling method (ESM) of research offers a valuable tool for studying individual processes as they unfold in everyday life. However, with ESM being a relatively new and developing study methodology, there is still a lack of consistency, methodological rigor, and understanding of best practices in this type of research. The methodological and practical advances of ESM presented in this thesis address critical gaps in our ability to accurately assess and understand psychological processes as they occur in real time.

The systematic measurement development approaches demonstrated have broad applicability for creating ESM measures across various psychological domains. This is particularly relevant for mental health applications, where accurate measurement of symptoms, emotions, and therapeutic processes in natural environments is crucial for effective intervention and treatment monitoring. By providing more precise tools for capturing psychological dynamics, these methods enable the development of personalized interventions based on individual characteristics and integrate nomothetic knowledge with ideographic one (Olthof et al., 2023). Even more broadly, good-quality ESM studies create a fundamental basis for developing just-in-time smartphone adaptive interventions, which have recently been proposed as a promising tool for changing people's behavior by providing necessary support exactly when it is needed (Nahum-Shani, 2017).

Our findings in personality psychology not only contribute to the growing body of

research highlighting the need to disentangle dynamic personality states from more stable personality traits (Fleeson & Gallagher, 2009; Heller et al., 2007) but also reveal the importance of motivational aspects in personality-related processes and behaviors. The role that goals and motives play in state and trait changes makes them a promising target for behavior-changing interventions.

The solutions to analytical challenges of ESM research, particularly for skewed data common in clinical populations, have direct implications for understanding mental health conditions, treatment effectiveness, and individual differences in psychological functioning. More accurate parameter estimation enables the detection of subtle but clinically meaningful relationships that may be obscured by traditional analytical approaches, leading to better-informed clinical decisions and more effective interventions.

For interpersonal relationships, the advanced understanding of emotional dynamics between partners can inform more effective couples therapy and relationship counseling. By revealing the complexity and heterogeneity of emotional interdependence, this work supports individualized approaches rather than one-size-fits-all interventions. The findings also contribute to prevention efforts by helping identify patterns that promote relationship satisfaction and stability.

Innovation

This thesis introduces or develops several key methodological innovations that advance ESM research capabilities:

The systematic bottom-up measurement development framework combines natural language processing, expert judgment, and rigorous psychometric evaluation to create ecologically valid instruments. This multi-step approach, demonstrated through the development of the ESM questionnaire assessing momentary importance of honesty- and

dishonesty-related goals, provides a scalable methodology for creating ESM measures that capture constructs as they naturally manifest in participants' experiences rather than imposing researcher assumptions. Moreover, conducting a multilevel psychometric evaluation of the new measurement instrument, which provides separate reliability estimates for within-person and between-person components, addresses the lack of a comprehensive psychometric approach to ESM measurement validation. If adopted as a common standard, this practice would ensure that instruments used are appropriate for intensive longitudinal studies.

Advanced multilevel modeling techniques for complex data structures, particularly the extended APIM framework for dyadic ESM data, enable examination of interpersonal processes while appropriately accounting for temporal dependencies, autoregressive effects, and multiple levels of nesting. These techniques provide templates for analyzing sophisticated research questions previously difficult to address with available methods.

The mixed-effects tobit modeling approach for bounded, skewed psychological constructs represents a novel approach in ESM data analysis. Rather than treating floor and ceiling effects as purely statistical nuisances, this approach recognizes them as meaningful features of psychological measurement requiring appropriate analytical frameworks. This innovation has broad applicability beyond the specific constructs examined, extending to various ESM measures prone to similar distributional challenges.

Target Audience

The methodological advances described in this thesis are relevant for diverse stakeholders across research, applied, and clinical settings.

ESM researchers and methodologists across psychology, mental health, and related fields will benefit from the systematic measurement development procedures, advanced analytical techniques, and solutions to common methodological challenges. The frameworks

provide practical guidance for conducting more rigorous and accurate intensive longitudinal research.

Clinical psychologists and mental health professionals can apply these methods for routine outcome monitoring, personalized assessment, and the development of targeted interventions. The ability to accurately capture psychological processes in daily life enables more precise clinical decision-making and the development of individualized treatment approaches (see e.g., Burger et al., 2022; Harnas et al., 2021; Smith & Juarascio, 2019).

Digital health developers and technology researchers working on mobile assessment platforms can implement these methodological frameworks to improve the accuracy and validity of intensive longitudinal assessments in various applications, from mental health monitoring to behavioral intervention tools.

Policymakers and healthcare administrators can use evidence from more accurate ESM research to develop effective, cost-efficient prevention strategies and personalized care models, supporting the transition toward positive health approaches that emphasize resilience and autonomy (e.g., Meredith et al., 2015).

Implementation

To disseminate our findings among ESM researchers, we participate in both national (such as the Belgian-Dutch ESM Network) and international (such as the Society of Ambulatory Assessment) conferences. Within the Belgian-Dutch ESM Network, we actively collaborate with applied clinical researchers who can benefit from improved ESM research, also within their clinical practice.

For the publication of our results, we prioritize open science practices and efforts to ensure reproducibility. We openly share our study materials, data, and analytical code for the use of the research community. The systematic approaches demonstrated in our studies can

serve as templates for developing ESM measures in other domains, fostering methodological rigor and consistency across studies. For the future development of the tobit model approach for analyzing skewed ESM data, we plan to conduct a simulation study systematically comparing it with alternative approaches, as well as to provide applied researchers with a comprehensive tutorial for the model implementation.

The goal for future implementation efforts would be to establish the methodological advances proposed in this thesis as standard practices in ESM research, leading to more accurate scientific understanding of psychological processes and more effective interventions for improving mental health and interpersonal functioning in daily life contexts.

References

Burger, J., Epskamp, S., Van Der Veen, D. C., Dablander, F., Schoevers, R. A., Fried, E. I., & Riese, H. (2022). A clinical PREMISE for personalized models: Toward a formal integration of case formulations and statistical networks. *Journal of Psychopathology and Clinical Science*, *131*(8), 906–916. https://doi.org/10.1037/abn0000779

Fleeson, W., & Gallagher, P. (2009). The implications of Big Five standing for the distribution of trait manifestation in behavior: Fifteen experience-sampling studies and a meta-analysis. *Journal of Personality and Social Psychology*, *97*(6), 1097–1114. https://doi.org/10.1037/a0016786

Harnas, S. J., Knoop, H., Booij, S. H., & Braamse, A. M. J. (2021). Personalizing cognitive behavioral therapy for cancer-related fatigue using ecological momentary assessments followed by automated individual time series analyses: A case report series. *Internet Interventions*, *25*, 100430. https://doi.org/10.1016/j.invent.2021.100430

Heller, D., Komar, J., & Lee, W. B. (2007). The Dynamics of Personality States, Goals, and Well-Being. *Personality and Social Psychology Bulletin*, *33*(6), 898–910. https://doi.org/10.1177/0146167207301010

Meredith, S., Alessi, S., & Petry, N. (2015). Smartphone applications to reduce alcohol consumption and help patients with alcohol use disorder: A state-of-the-art review. *Advanced Health Care Technologies*, 47. https://doi.org/10.2147/AHCT.S65791

Nahum-Shani, I., Smith, S. N., Spring, B. J., Collins, L. M., Witkiewitz, K., Tewari, A., & Murphy, S. A. (2017). Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. *Annals of Behavioral Medicine*, *52*(6), 446–462. https://doi.org/10.1007/s12160-016-9830-8

Olthof, M., Hasselman, F., Aas, B., Lamoth, D., Scholz, S., Daniels-Wredenhagen, N., Goldbeck, F., Weinans, E., Strunk, G., Schiepek, G., Bosman, A. M. T., & Lichtwarck-Aschoff, A. (2023). The Best of Both Worlds? General Principles of Psychopathology in Personalized Assessment. *Journal of Psychopathology and Clinical Science*, *132*(7), 808–819. https://doi.org/10.1037/abn0000858

Smith, K. E., & Juarascio, A. (2019). From Ecological Momentary Assessment (EMA) to Ecological Momentary Intervention (EMI): Past and Future Directions for Ambulatory Assessment and Interventions in Eating Disorders. *Current Psychiatry Reports*, *21*(7), 53. https://doi.org/10.1007/s11920-019-1046-8