Personalized Behavior Prediction: An Idiographic Person-Situation Test

Emorie D. Beck

Northwestern University Feinberg School of Medicine

Joshua J. Jackson

Washington University in St. Louis

Authors Note. Correspondence concerning this article should be addressed to Emorie D Beck, 633 N St. Clair St., Chicago, IL 60611. Email: emorie_beck@northwestern.edu. The preparation of this manuscript was supported by a grant from the National Science Foundation to Joshua Jackson (BCS-1125553). Data collection was supported by a Society for Personality and Social Psychology Inside the Grant Panel grant awarded to Emorie Beck. Emorie Beck was supported by National Institute on Aging Grants T32 AG00030-3, R01-AG067622, and R01-AG018436.

Introduction, Discussion, Footnotes, Acknowledgments, and Appendices: no more than 2,000 words combined.
In almost all cases, an adequate account of method and results can be achieved in 2,500 or fewer words for Research Articles. Methodological minutiae and fine-grained details on the Results—the sorts of information that only “insiders” would relish and require for purposes of replication—should be placed in Supplemental Online Materials-Reviewed, not in the main text.

Abstract

A longstanding goal of psychology is to predict the things people do, but tools to predict accurately future behaviors remain elusive. In the present study, we used intensive longitudinal data (N = 104; total assessments = 5,971) and three machine learning approaches to investigate the degree to which two behaviors – loneliness and procrastination – could be predicted from psychological (i.e. personality and affective states), situational (i.e. objective situations and psychological situation cues), and time (i.e. trends, diurnal cycles, time of day, and day of the week) phenomena from an idiographic, person-centered perspective. Rather than pitting persons against situations, such an approach allows psychological phenomena, situations, and time to jointly inform prediction. We find (1) a striking degree of accuracy across participants, (2) that a majority of participants models are best informed by both person and situation features, and (3) that the most important features vary greatly across people.

Keywords: idiographic, personality, prediction, machine learning, ESM

"Let me repeat this question, for it is the one that more than any other has haunted me over the years. What makes the system cohere in any one person?"

- Allport, 1960, p. 308

A longstanding goal of psychology is to describe (e.g., Titchener, 1898), predict (e.g., Meehl, 1954), and explain (e.g., Fodor, 1968) the things people do. Although much research has developed tools to describe and explain such behaviors, both in terms of person and situation characteristics and using both observational and experimental techniques, tools for accurately predicting future behaviors have remained elusive. Indeed, most of the existent research on prediction examines broad life outcomes (e.g., Beck & Jackson, 2021; Joel et al., 2020). While such broad life outcomes result from accumulating behaviors (e.g., Hampson, Goldberg, Vogt, & Dubanoski, 2007), how predictable those behaviors are is unknown.

We argue that the elusiveness of accurate predictions of future behaviors stems from an almost exclusive focus of a between-person perspective. In the present study, we offer an alternative person-centered, idiographic approach to behavior prediction, where the antecedents of everyday behavior are allowed to vary across people. We use three machine learning approaches to investigate the degree to which two future behaviors – loneliness and procrastination – can be predicted from psychological phenomena (i.e. personality and affective states), situations (i.e. objective situations and psychological situation cues), and time (i.e. trends, diurnal cycles, time of day, and day of the week).

An individualized, idiographic approach to assessment

A major assumption of psychometric assessment is that a measured construct is the same across people. A personality characteristic like Extraversion is Extraversion for everyone, and how it is related to Neuroticism is the same for everyone. If this assumption is violated, and it almost always is (Borsboom, Mellenbergh, & van Heerden, 2003; Fisher, Medaglia, & Jeronimus, 2018; Molenaar, 2004), then it hard to say whether or not Extraversion predicts behavior at the level of a person. Indeed, doing so would be to make an incorrect within-person interpretation of a between-person model (Borsboom et al., 2003). Alternatively, idiographic approaches sidestep such assumptions by focusing on a single individual, attempting to identify variables that are meaningful for them, which may not be meaningful for another person (Beck & Jackson, 2020a).

To what extent are idiographic approaches necessary? A growing body of work demonstrates that the within-person structure of emotion and personality differs across people (Beck & Jackson, 2020b; Borkenau & Ostendorf, 1998; Molenaar, 2004). Across people, measures of the Big Five demonstrate five factors, while within-person, they range from two to seven and have different content within them. Thus, common taxonomies used to describe populations may not describe an individual.

Similarly, taxonomic work on situations is only beginning (Parrigon, Woo, Tay, & Wang, 2017; Rauthmann et al., 2014). In part, this is because of the great range in behaviors exhibited within similar situations, which makes identifying coherent patterns that are found between-persons difficult. It is likely that situations impact a person largely idiosyncratically (i.e. idiographically) for each person (Mischel, 2004). Simply – there is little reason to believe individuals should respond to the "same" (objective) situations similarly. Even if individuals do respond to the same situations similarly, there is almost certainly causal heterogeneity in why individuals respond similarly to the same situation. Consider, for example, a common behavior like small talk at work. In the simplest case, it could be fueled by common goals – to get ahead and get along – like avoiding an awkward interpersonal situation (get alone) or making a long-term beneficial social connection (get ahead). Differences in goals likely differ both across people and within them, and differences in these goals drive not just the behavior but also the content of the conversation.

Persons *and* Situations versus Persons *or* Situations

Predicting behavior using person and situation factors has mostly been ignored as a lingering consequence of the Person-Situation Debate (e.g., Barrick & Mount, 2005), which implied that the threshold on predicting single behaviors was low. The "personality coefficient" of .3 was seen as an upper bound of what is possible in behavioral prediction and is typically resolved by focusing on aggregating behaviors to increase predictive validities (Epstein, 1979), discussing the relevance of a .3 correlation (Funder, 2009; Roberts, 2009), or focusing on strong situations (Snyder & Ickes, 1985). However, these discussions typically concern a one-to-one predictor-behavior association. If instead a behavior is multi-determined from many sources, theoretical estimates can go much higher than .3 (Ahadi & Diener, 1987).

Another longstanding question in the prediction of behavior asks whether person or situational factors impact behavior more (Epstein & O'Brien, 1985). But most approaches hold either person or situation features constant to examine the association between the other and behavior (Kenrick & Funder, 1988). In other words, rather than viewing the triad of personality, situations, and behavior simultaneously, most studies examine these in isolation (Funder, 2006). Of those that do examine person and situational features simultaneously, findings indicate the importance of both persons and situations as *independent* but not interactive influences (Sherman, Rauthmann, Brown, Serfass, & Jones, 2015), leading people to continue studying these in isolation. Thus, it remains unclear how situations coalesce with person factors to impact behavior.

The Present Study: A Machine Learning Approach to Behavioral Prediction

We argue that by adopting an idiographic, machine learning-based prediction approach that incorporates information about persons, situations, and time relative only to a single person's experience will allow us to accurately predict future behavior. In the clinical psychology domain, previous research has indicated that future behaviors, like smoking (Fisher & Soyster, 2019) and food craving (Butter et al., 2020) (1) can be predicted with high levels of accuracy using these methods and (2) that the degree of predictability and the important features across people vary considerably. Thus, in the domain of psychology more broadly, we believe that machine learning methods can be used to understand (1) the degree to which we can predict behavioral outcomes, (2) individual differences in how predictable such behavioral outcomes are, (3) whether certain domains (i.e. persons, situations, and time) out-predict others, (4) which features play the strongest, and (5) whether and to what degree individuals differ in which features play the strongest role.

Method

This study was preregistered on the Open Science Framework (OSF; https://osf.io/4nm5p) and all data, analysis scripts, and results are available on both the OSF (https://osf.io/8ebyx/) and GitHub (https://github.com/emoriebeck/behavior-prediction). More details on the analyses and visual results depictions are available online at https://emoriebeck.github.io/behavior-prediction/ and in the R Shiny webapp at https://emoriebeck.shinyapps.io/behavior-prediction/. All data are completely de-identified. This study was approved by the Institutional Review Board at Washington University in St. Louis (#201806124), and all data were collected in alignment with the APA ethics code. Components of these data have been published elsewhere (Beck & Jackson, 2021b; Jackson & Beck, 2021).

Participants

Participants were 208 (71.96% female; M_{age} = 19.51, SD_{age} = 1.27) undergraduates at Washington University in St. Louis who enrolled in a study between October 2018 and December 2019. 69 identified as white, 67 as Asian, 34 as Black, and 30 other race/ethnicity or mixed race/ethnicity (8 declined to answer). Nine participants were excluded for not completing any ESM surveys. The remaining participants completed a total of 8,403 surveys (*M* = 42.23; *SD* = 24.01; range 1-109). See Table S1 in the online materials for additional information. An additional 85 participants were excluded for having less than 40 ESM measurements, and 10 were excluded for having too little variance in one or both outcome measures. The remaining 104 participants (72.82% female; M_{age} = 19.49, SD_{age} = 1.31) completed an average of 57.41 assessments (*SD* = 16.33; range 40-109). 32 identified as white, 33 as Asian, 14 as Black, and 16 as other (9 declined to answer).

Measures

Participants responded to a battery of trait and ESM measures as part of the larger study (see the supplementary codebooks in the online materials). The present study focuses on a subset of preregistered ESM measures that were used to estimate idiographic personality prediction models.

ESM Measures

Personality and Affect. Personality was assessed using the full BFI-2 (Soto & John, 2017) using a planned missing data design (Revelle et al., 2016; https://osf.io/pj9sy/). Items capturing affect were a subset of the PANAS-X (Watson & Clark, 1999) with items redundant with the BFI-2 removed. Each item was answered relative to what a participant was just doing on a 5-point Likert-like scale from 1 “disagree strongly” to 5 “agree strongly.”

Situations. Binary situation indicators were derived by asking research assistants to provide list of the common social, academic, and personal situations in which they found themselves. From these, we derived a list of 19 unique situations. Separate items for arguing with or interacting with friends or relatives were composited in overall argument and interaction items. Participants checked a box for each event that occurred in the last hour (1 = occurred, 0 = did not occur). Psychological features of situations were measured using the ultra-brief version of the “Situational Eight” DIAMONDS scale (S8-I; Rauthmann & Sherman, 2015) on a 3-point scale from 1 “not at all” to 3 “totally.”

Timing Features. Time features were created from the time stamps collected with each survey based on approaches used in other studies of idiographic prediction (Fisher & Soyster, 2019; Butter et al., 2020). To create these, we created time of day (4; morning, midday, evening, night) and day of the week dummy codes. Next, we created a cumulative time variable (in hours) from first beep to create linear, quadratic, and cubic time trends as well as one and two period sine and cosine functions across each 24 period (e.g., 2 period sine = $\sin\frac{2\pi}{12} * \ cumulative\ time_t$ and 1 period sine = $\sin\frac{2\pi}{24} * \ cumulative\ time_t$).

Outcomes. Procrastination and loneliness were assessed by asking participants to check a box if it had occurred in the last hour (1 = occurred, 0 = did not occur). Each will be lagged such that time t features will predict time t+1 procrastination or loneliness.

Procedure

Participants responded to two types of surveys: trait and state (Experience Sampling Method; ESM) measures, for which they were paid separately. More information on the procedure of this study sample have been reported elsewhere (Beck & Jackson, 2021b; Jackson & Beck, 2021) and are available in the online materials.

Analytic Plan

The present study used three machine learning classification models: (1) Elastic Net Regression (Friedman, Hastie, & Tibshirani, 2010), (2) The Best Items Scale that is Cross-validated, Correlation-weighted, Informative and Transparent (BISCWIT; Elleman, McDougald, Condon, & Revelle, 2020), and (3) Random Forest Models (Kim et al., 2019). More details on these methods and the procedure can be found in the online materials but are summarized below.

Because we have a large number of features to test, we chose methods with variable selection procedures and methods for reducing overfitting. To both reduce the number of indicators used in each model and to test which group of indicators are the most predictive of future procrastination and loneliness, we will also examine these in several sets: (1) Psychological indicators (personality + affect) (25), (2), Situation indicators (binary + DIAMONDS) (24), and (3) Full set (personality + affect + binary situations + DIAMONDS) (49). Each of these will also be tested with and without the 18 timing indicators, for a total set of six combinations of the 67 features.[1]

In each of these methods, we used cumulative rolling origin forecast validation,[2] which was comprised of the first 75% of the time series, and held out the remaining 25% of the data set

[1] We preregistered testing each of the psychological and situation indicators as separate feature sets. However, for parsimony, these results are only reported in the online materials.
[2] We preregistered the use of 10-fold cross-validation on the training set, rather than rolling origin forecast validation. We ultimately elected to use rolling origin validation because 10-fold cross validation would have resulted in using future observations of the time series to predict past observations, while rolling origin forecast validation slides a cumulative window across the time series, always predicting future observations.

for the test set. In the rolling origin forecast validation, we used the first one-third of the time series as the initial set, five observations as the validation set, and set skip to one (to reduce the number of folds to roughly equate 10-fold cross-validation), which resulted in 10-15 rolling origin "folds." For all training and test sets, the outcomes were lagged such that each outcome was predicted by previous time point features (roughly 4 hours previously). Tuning results are available for each participant, feature set, outcome, and model in the online materials and webapp ("Model Tuning Figures").

Out of sample prediction was tested based on classification error and area under the ROC (receive operating characteristic) curve (AUC). Classification error is a simple estimate of the percentage of the test sample that was correctly classified by the model. In addition, the AUC will capture the trade-off between sensitivity and specificity across a threshold. In the present study, we used an AUC threshold of .5, which indicates binary classification at chance levels. AUC curves are available in the online materials and webapp ("ROC").

Elastic Net Regression (ENR) uses *L1* (Ridge) and *L2* (LASSO) regularization to shrink coefficients based on a penality (λ) that is tuned to minimize error using cross-validation. We tuned based on penalty and mixture (set to 10 values each). ENR was performed using the `tidymodels` package in `R` to estimate the ENR models by calling the `logistic_reg()`, setting the engine as "`glmnet`", and the mode as "`classification`".

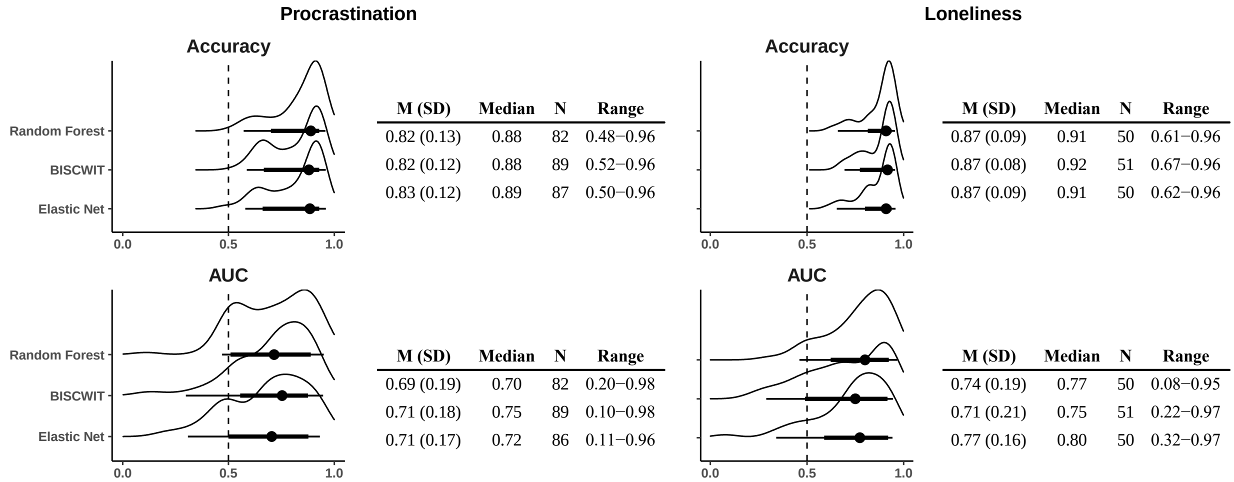
The Best Items Scale that is Cross-validated, Correlation-weighted, Informative and Transparent (BISCWIT) is a correlation-based machine learning technique. Using the `best.scales()` function in the `psych` package, we used rolling origin validation to choose the best number of items to be retained using validation set accuracy. Weighted scores were calculated by extracting the correlations from the best scales object and using it in the `scoreWtd()` function to create the correlation weighted scores.

Random forest (RF) models are a variant of decision tree classification algorithms that utilize ensemble methods) methods. Because random forest using bagging (i.e. bootstrapping with aggregation), we performed a series of steps that make bootstrapping appropriate with time series data. Models were tuned using `mtry` (i.e. the number of predictors that will be randomly sampled at each split when creating tree models) and `min_n` (i.e. the minimum number of data points in a node that is required for the node to be split further), which were each set to 10 values. We used the `tidymodels` package in `R` to estimate the random forest models by calling the `rand_forest()`, setting the engine as `"ranger"`, with `importance = "permutation"` in order to extract variable importance, and the mode as `"classification"`.

Results [1395 words]

Can we predict future procrastination and loneliness?

First, we tested to what extent we could predict future incidences of procrastination and loneliness for each person by their previous assessments using ENR, BISCWIT, and RF. Figure 1 presents histograms and descriptive statistics of accuracy and AUC across the full sample for each outcome and model. As is clear in the figure, predictive accuracy was high overall, with mean accuracy of .87 (Median .91 to .92) for loneliness and between .82 and .83 (Median .88 to .89) for procrastination. Similarly, AUC was also well above the .5 threshold with means ranging from .70 to .76 (Median .75 to .80) for loneliness and .69 to .70 (Median .70 to .75) for procrastination. Participant level descriptives across models, feature sets, and outcomes are available in the online materials and webapp ("All Model Performance").



Procrastination — Accuracy

M (SD)	Median	N	Range
0.82 (0.13)	0.88	82	0.48−0.96
0.82 (0.12)	0.88	89	0.52−0.96
0.83 (0.12)	0.89	87	0.50−0.96

Loneliness — Accuracy

M (SD)	Median	N	Range
0.87 (0.09)	0.91	50	0.61−0.96
0.87 (0.08)	0.92	51	0.67−0.96
0.87 (0.09)	0.91	50	0.62−0.96

Procrastination — AUC

M (SD)	Median	N	Range
0.69 (0.19)	0.70	82	0.20−0.98
0.71 (0.18)	0.75	89	0.10−0.98
0.71 (0.17)	0.72	86	0.11−0.96

Loneliness — AUC

M (SD)	Median	N	Range
0.74 (0.19)	0.77	50	0.08−0.95
0.71 (0.21)	0.75	51	0.22−0.97
0.77 (0.16)	0.80	50	0.32−0.97

Figure 1. Histograms of classification accuracy and Area Under the Receiver Operator Curve (AUC) for participants' best models.

Are there individual differences in the idiographic range of prediction across people?

Given that future behavior predictions were high in classification accuracy and AUC, we next tested whether there were individual differences across feature sets in how predictable each behavior was for each person. Figure 2 presents the median, 66%, and 95% range of classification accuracy for a random sample of 25 participants, ordered by the median accuracy (AUC is available in the online materials and webapp ["Model Performance Distributions"]). As is clear in the figures, accuracy varies both across people and within them. In other words, although there are between-person differences in the degree of accuracy, there are also within-person differences, depending on which features are used.

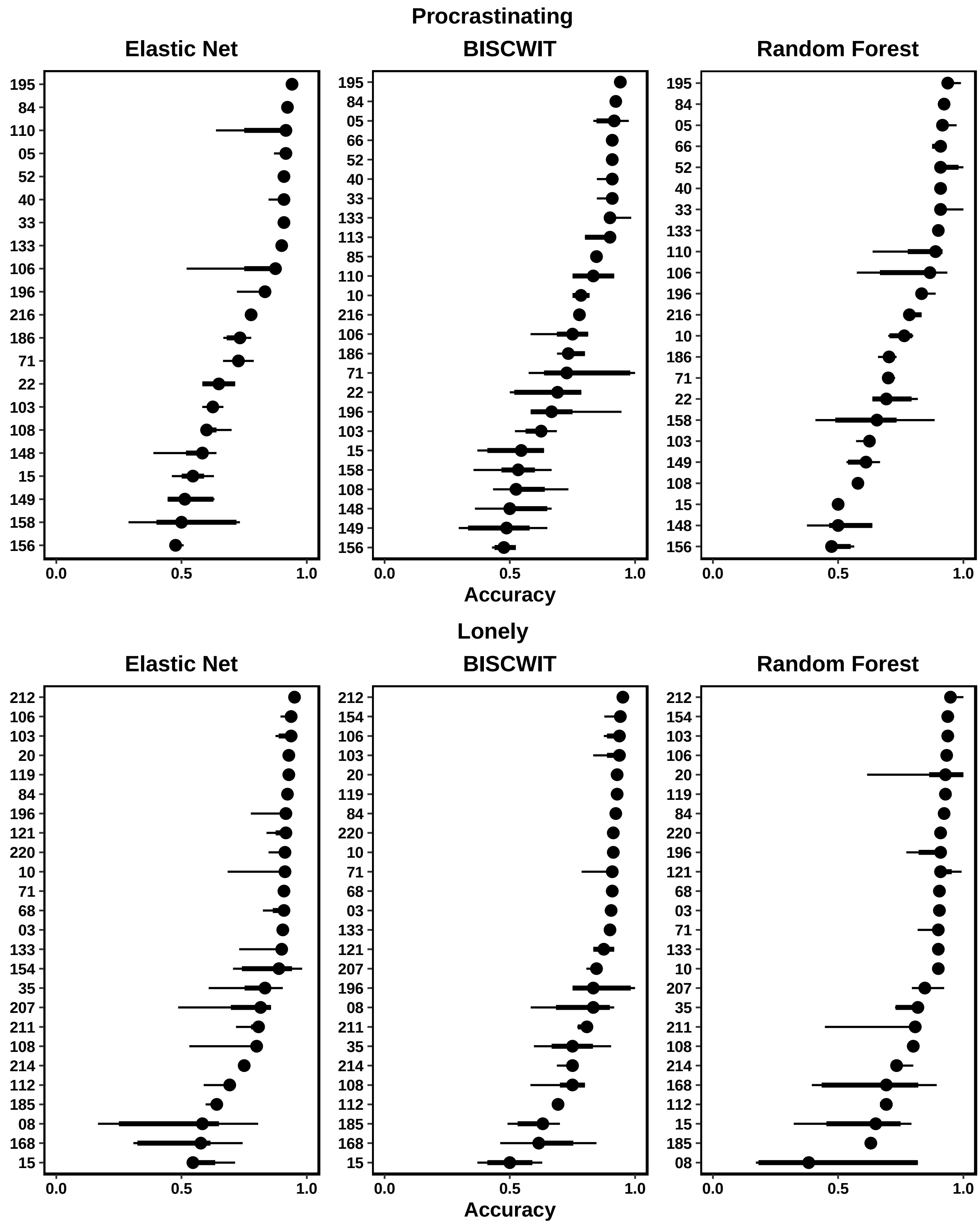


Figure 2. Person-level distributions (dot = median, wide line = 66% interval, thin line = 95% interval) of classification accuracy for 25 sample participants predicting Procrastination and Loneliness.

Do Psychological, Situational, Time, or Full Feature Sets Perform Best?

Given the range of classification accuracy within-person, we next examined which set of features produced the best model in terms of classification accuracy and AUC for each person. Table 1 presents the number of and percentage of participants whose best model was for each feature set. As is clear, feature sets without time performed better than those with time. Second, relative to AUC, using accuracy as the selection metric was more likely to indicate that the full feature set performed best. Third, with some slight differences, relative proportions were similar across the three methods. Finally, for accuracy but not AUC, only RF indicated that situation feature models performed better than psychological feature models. We next examined the breakdown of selected features for each participant. As is clear in Figure 3, which shows proportions of features for all participants' best models for each method, there were individual differences in the proportion of psychological, situational, and time features. Some participants' best models included exclusively psychological or situational features, with most showing a varying mixture of both.

Table 1

Frequencies of the Full, Psychological, and Situation Feature Sets with or Without Time Being the Best Model for a Participant

		Loneliness						Procrastination					
		Elastic Net		BISCWIT		Random Forest		Elastic Net		BISCWIT		Random Forest	
Set	**Time**	**#**	**%**	**#**	**%**	**#**	**%**	**#**	**%**	**#**	**%**	**#**	**%**
Accuracy													
Full	No	36	70.6%	32	62.7%	21	41.2%	57	64.8%	46	51.7%	30	33.7%
	Yes	3	5.9%	3	5.9%			5	5.7%	8	9.0%	7	7.9%
Psychological	No	6	11.8%	10	19.6%	4	7.8%	15	17.0%	15	16.9%	12	13.5%
	Yes	3	5.9%	3	5.9%	2	3.9%	6	6.8%	6	6.7%	3	3.4%
Situations	No	3	5.9%	3	5.9%	23	45.1%	3	3.4%	8	9.0%	31	34.8%
	Yes					1	2.0%	2	2.3%	6	6.7%	6	6.7%
AUC													

Table 1

Frequencies of the Full, Psychological, and Situation Feature Sets with or Without Time Being the Best Model for a Participant

		Loneliness						Procrastination					
		Elastic Net		BISCWIT		Random Forest		Elastic Net		BISCWIT		Random Forest	
Set	**Time**	**#**	**%**	**#**	**%**	**#**	**%**	**#**	**%**	**#**	**%**	**#**	**%**
Full	No	14	27.5%	7	13.7%	12	23.5%	14	15.9%	14	15.7%	21	23.9%
	Yes	5	9.8%	7	13.7%	10	19.6%	13	14.8%	10	11.2%	6	6.8%
Psychological	No	10	19.6%	12	23.5%	10	19.6%	24	27.3%	16	18.0%	17	19.3%
	Yes	6	11.8%	2	3.9%	6	11.8%	8	9.1%	23	25.8%	9	10.2%
Situations	No	10	19.6%	14	27.5%	10	19.6%	18	20.5%	18	20.2%	24	27.3%
	Yes	6	11.8%	9	17.6%	3	5.9%	11	12.5%	8	9.0%	11	12.5%

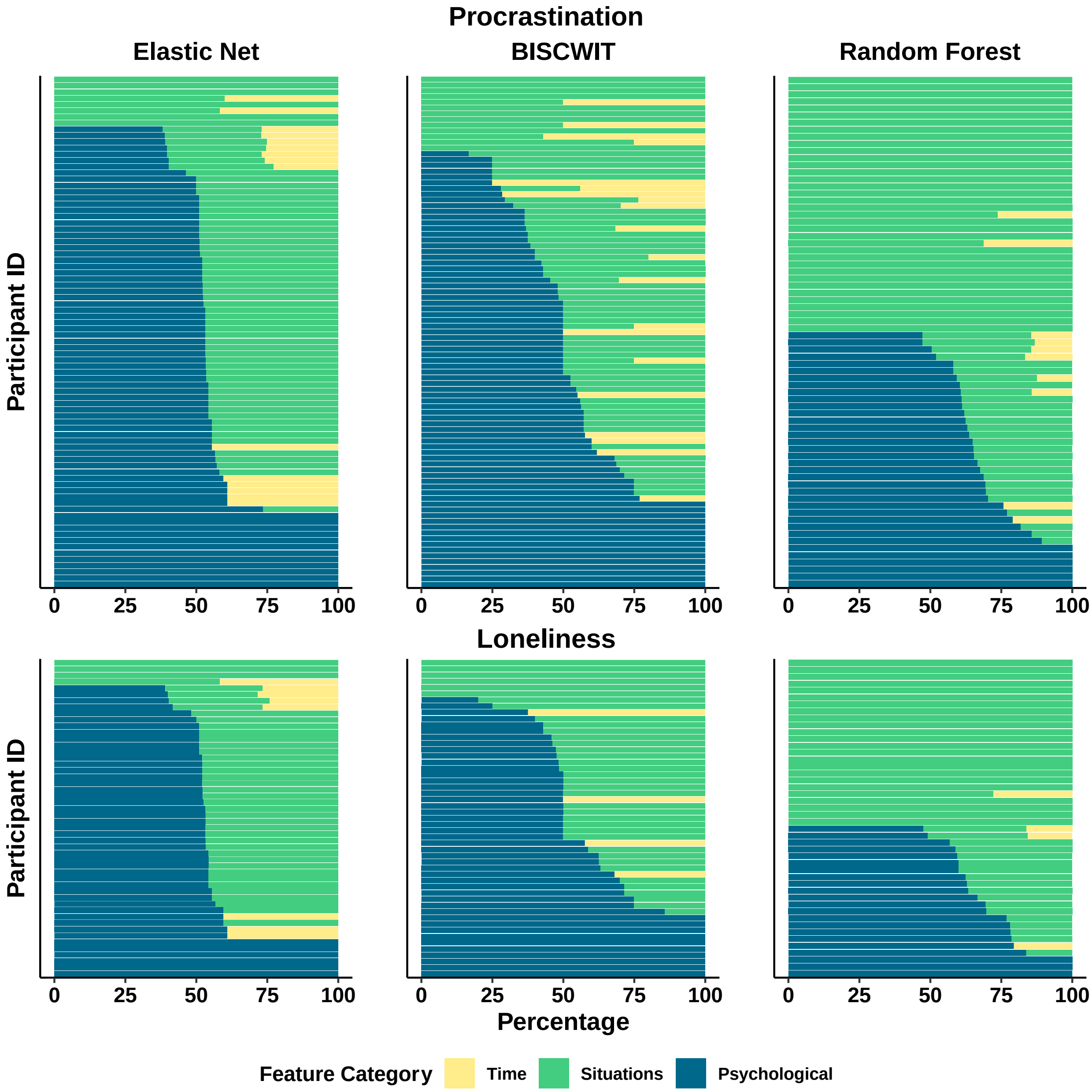


Figure 3. Sequence plots of the percentage of features from the Psychological, Situational, and Time Features Sets for each participant for each outcome and model.

Which features are most associated with future Procrastination and Loneliness?

To better understand which features were driving differences in which feature set produced the best model for each person, we next examined the variable importance metrics for each participant's best models. To do so, we extracted the top five features and calculated the proportion of the sample that had each feature in their top five. The resulting Figure 4 has several takeaways. First, across models, timing features were less frequent, with the exception linear, quadratic, and cubic trends (T12-T14) across the ESM period. Second, for ENR and BISCWIT, psychological features were slightly more frequent than situation features. Third, one consequence of the higher frequency of situation feature RF models being selected than for the other two models, top five situation features were both more frequent as well as more variable (more different sized circles) for the RF models than for ENR or BISCWIT (more similarly sized circles). Finally, and perhaps most crucially, this figure makes clear that person and situation characteristics were both key in predicting each outcome, with neither “dominating” the feature space.

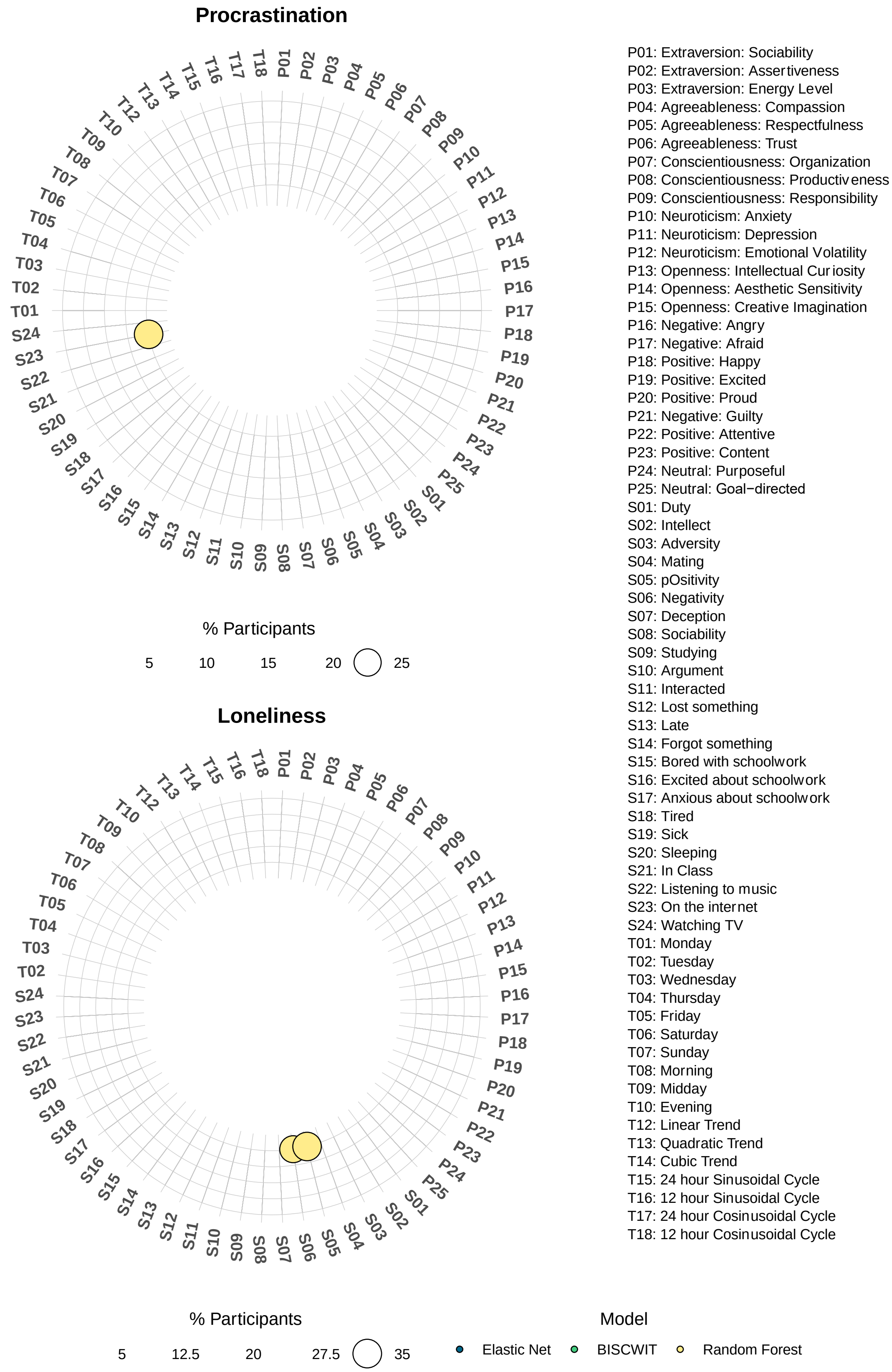

Procrastination
% Participants
5 10 15 20 25
Loneliness
% Participants
5 12.5 20 27.5 35
Model
Elastic Net
BISCWIT
Random Forest
P01: Extraversion: Sociability
P02: Extraversion: Assertiveness
P03: Extraversion: Energy Level
P04: Agreeableness: Compassion
P05: Agreeableness: Respectfulness
P06: Agreeableness: Trust
P07: Conscientiousness: Organization
P08: Conscientiousness: Productiveness
P09: Conscientiousness: Responsibility
P10: Neuroticism: Anxiety
P11: Neuroticism: Depression
P12: Neuroticism: Emotional Volatility
P13: Openness: Intellectual Curiosity
P14: Openness: Aesthetic Sensitivity
P15: Openness: Creative Imagination
P16: Negative: Angry
P17: Negative: Afraid
P18: Positive: Happy
P19: Positive: Excited
P20: Positive: Proud
P21: Negative: Guilty
P22: Positive: Attentive
P23: Positive: Content
P24: Neutral: Purposeful
P25: Neutral: Goal−directed
S01: Duty
S02: Intellect
S03: Adversity
S04: Mating
S05: pOsitivity
S06: Negativity
S07: Deception
S08: Sociability
S09: Studying
S10: Argument
S11: Interacted
S12: Lost something
S13: Late
S14: Forgot something
S15: Bored with schoolwork
S16: Excited about schoolwork
S17: Anxious about schoolwork
S18: Tired
S19: Sick
S20: Sleeping
S21: In Class
S22: Listening to music
S23: On the internet
S24: Watching TV
T01: Monday
T02: Tuesday
T03: Wednesday
T04: Thursday
T05: Friday
T06: Saturday
T07: Sunday
T08: Morning
T09: Midday
T10: Evening
T12: Linear Trend
T13: Quadratic Trend
T14: Cubic Trend
T15: 24 hour Sinusoidal Cycle
T16: 12 hour Sinusoidal Cycle
T17: 24 hour Cosinusoidal Cycle
T18: 12 hour Cosinusoidal Cycle

Figure 4. Percentage of each feature appearing as a top five variable importance feature across the full sample for Procrastination (top) and Loneliness (bottom). Larger, darker circles indicate higher percentages, while smaller, lighter circles indicate lower percentages. Each features' corresponding label is listed in the right side of the figure.

Do people vary in the which features are most important?

To demonstrate how people differ in which features were important, we will consider three participants whose best models were captured by the psychological, situational, or full feature sets. Similar plots for the remaining participants are available in the online materials and webapp ("Participant Profiles"). In addition, plots displaying profiles of participants coefficients in their best models as well as all combinations of outcomes, models, and feature sets are available in the online materials and webapp ("Participant Profiles").

First, participant 169's best model for procrastination used the psychological feature set without time for each of the three methods (accuracy = 0.94; AUC = 0.80). Variable importance (log odds ratios) for the features in their ENR model are shown in the bar graph in Figure 5. Across all three methods, there were some differences selected features, but consensus in the direction and general magnitude of them. Across all three, the top feature was the Openness to Experience facet Creative Imagination, perhaps indicating that this participant tended to procrastinate when they were feeling more creative or imaginative previously. As in clear in Figure 5, they also tended to procrastinate less when they were Intellectually Curious (O) and more when they felt afraid. Thus, it seems like this participant's procrastination may partially hinge upon competition between intellectual and creative pursuits, as well general fears.

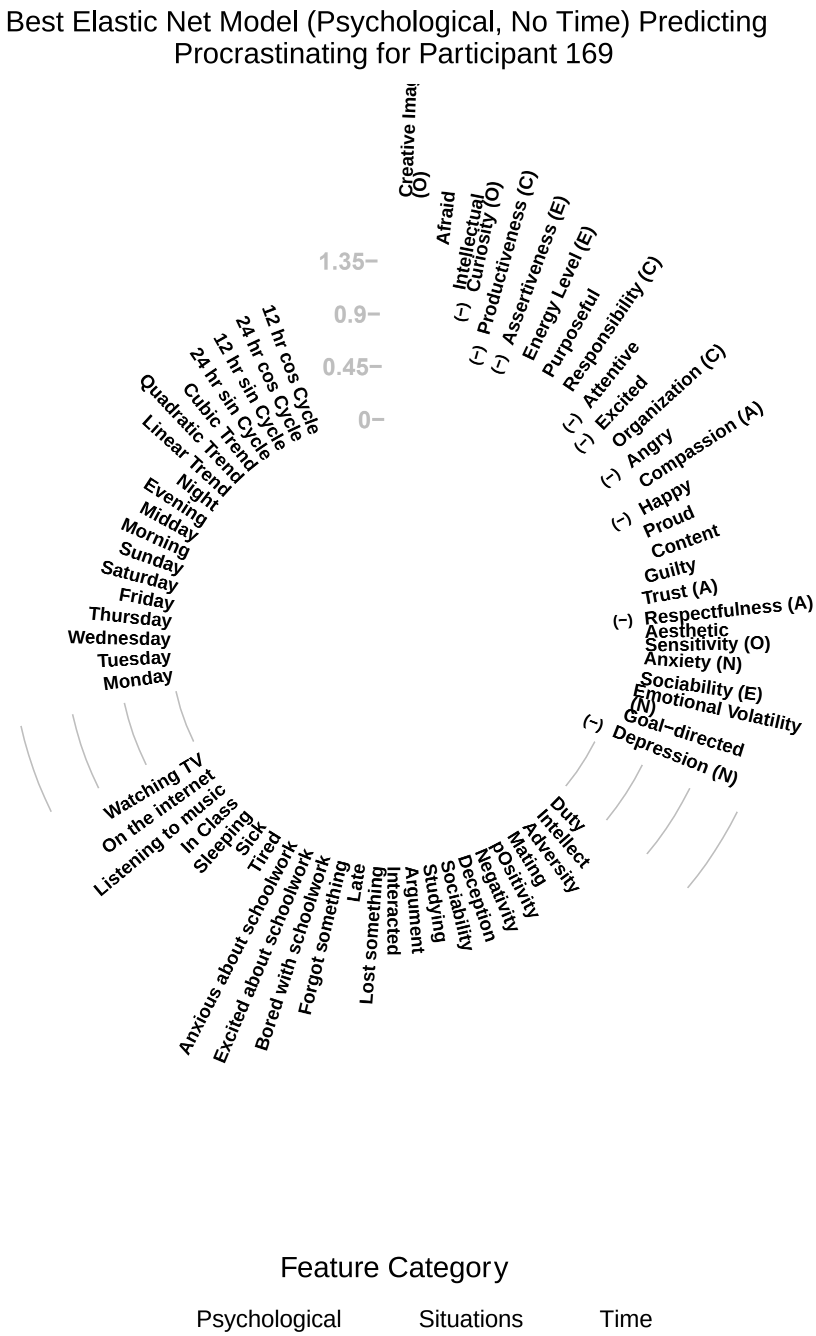


Figure 5. Variable Importance (absolute value of log odds) for Participant 169's best model predicting Procrastination. Each bar is a different feature. Color indicates feature category (psychological, situational, or time). Height of bars indicates the magnitude of the effect. (-)indicate negative effects (i.e. lower odds).

Second, participant 43's best model for loneliness used the situation feature set (without time; accuracy = 0.91; AUC = 0.83). As in seen in the bar graph of their ENR variable importance in Figure 6, the situation characteristics and features seem to indicate that obligations (e.g., duty, in class), physical health (e.g., sick, sleeping), and social interactions (Sociability, argument) were predictive of future feelings of loneliness. For example, both the situation feature Duty and being in class predicted less loneliness, while feeling sick and getting in an argument predicted more.

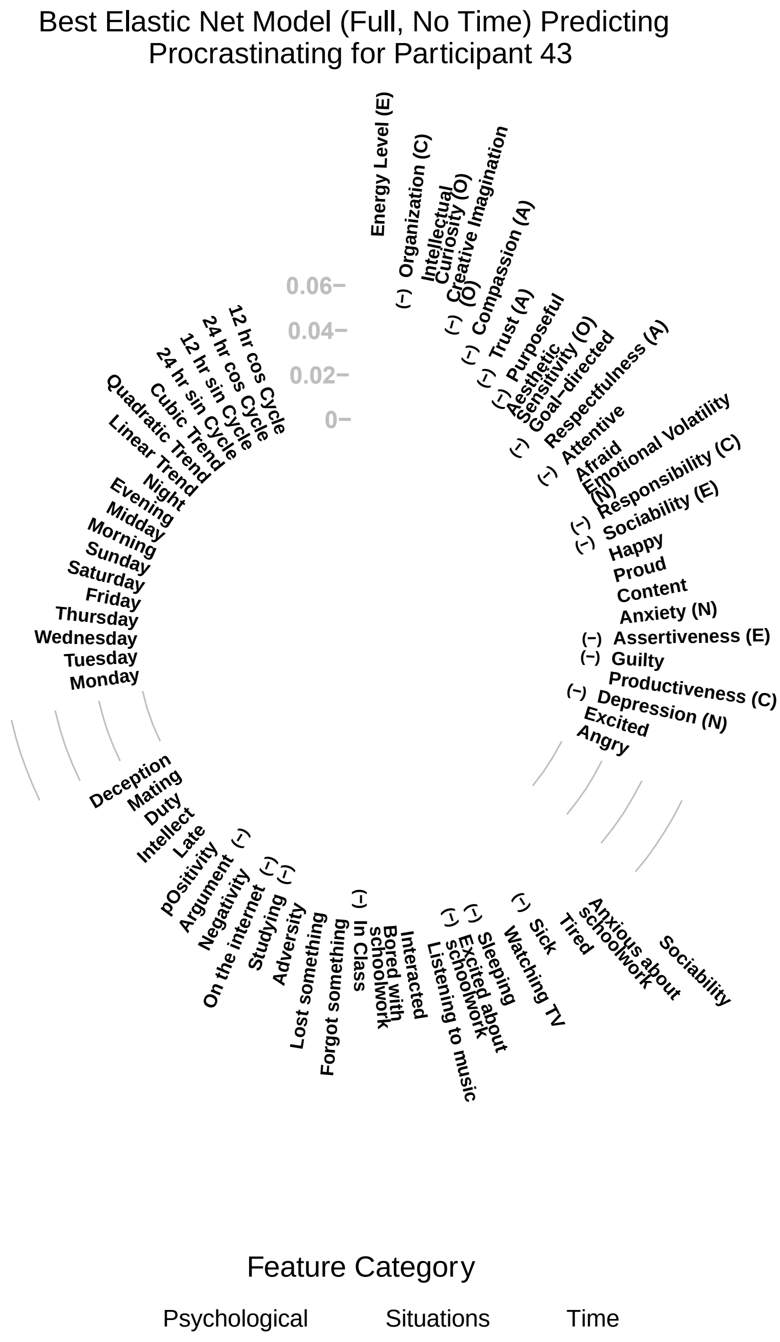


Figure 6. Variable Importance (absolute value of log odds) for Participant 43's best model predicting Loneliness. Each bar is a different feature. Color indicates feature category (psychological, situational, or time). Height of bars indicates the magnitude of the effect. (-) indicates negative effects.

Finally, participant 160’s best models for procrastination utilized the full feature set (i.e. psychological and situational features) without time (see Figure 7; accuracy = 0.89, AUC = 0.94). ENR and BISCWIT agreed on the top three features: Sociability (DIAMONDS; negative), Sleeping (positive), and Depression (Neuroticism; negative) were each associated with future procrastination. Moreover, other related features, like attentiveness and Assertiveness (Extraversion), were also predictive of both outcomes.

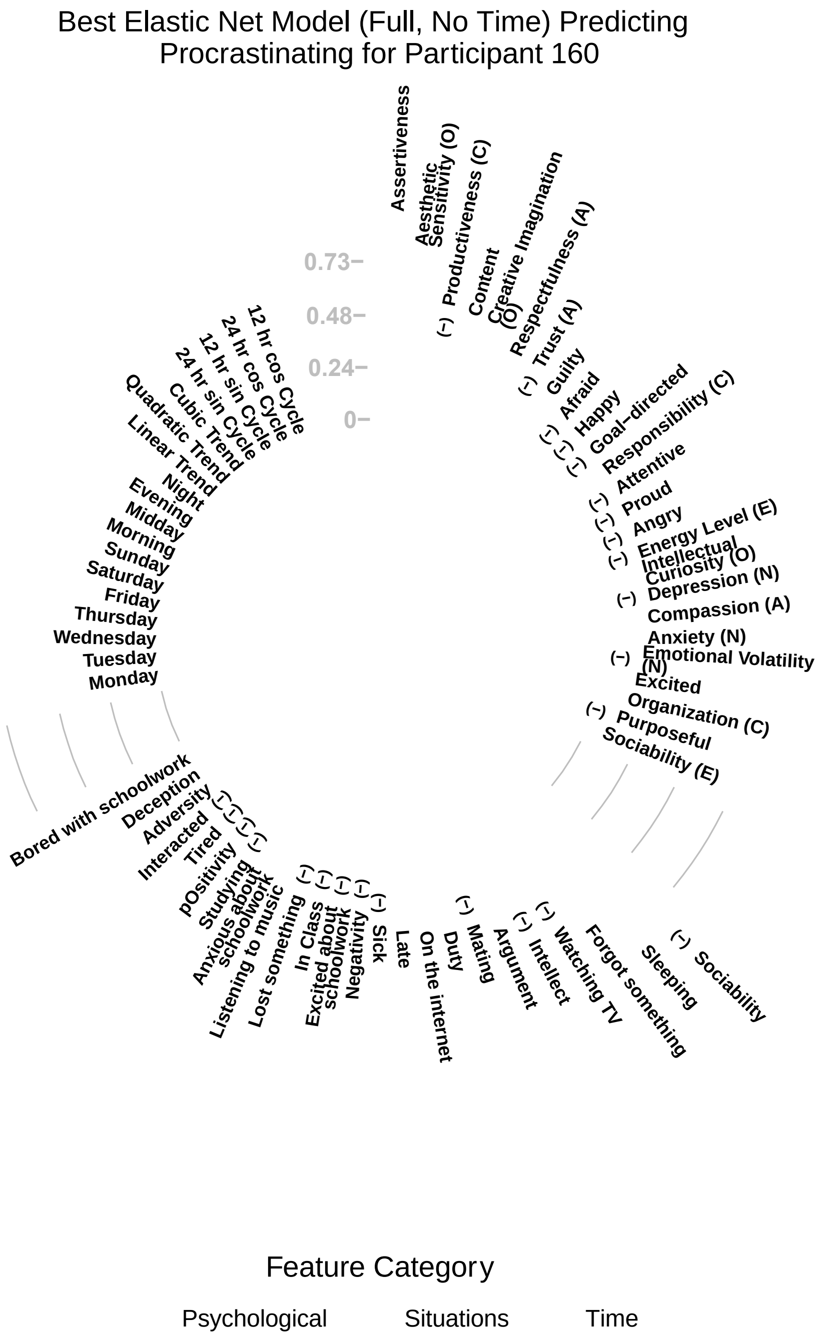


Figure 7. Variable Importance (absolute value of log odds) for Participant 160's best model predicting Procrastination. Each bar is a different feature. Color indicates feature category (psychological, situational, or time). Height of bars indicates the magnitude of the effect. (-) indicates negative effects (i.e. lower odds).

Discussion [1005 words]

The current study investigated personalized, idiographic prediction models for two behaviors, feeling lonely and procrastinating. Rather than assuming that antecedents of different outcomes were shared, our idiographic approach built *N*=1, personalized prediction models. Overall, three main conclusions emerged: First, psychological, situational, and time variables accurately predicted future everyday behaviors. Second, psychological and situational variables were both important, almost equally so, with neither being a predominant antecedent of behavior. Third, individual differences reigned supreme –people differed on how predictable outcomes were, which domains performed best, and which features were most important. These findings indicate the utility of an idiographic approach to psychological assessment relative to standard between-person approaches that are routinely used.

On predicting more behaviors more of the time

We found accurate out-of-sample prediction of procrastination and feelings of loneliness when using a suite of personality and situational factors. While there are between-person individual differences in both loneliness and procrastination, there was also within-person variability in terms of how and when people experienced these behaviors. Typical prediction models within psychology have largely focused on which between-person features predict life outcomes or other aggregated behaviors (e.g., Beck & Jackson, 2021a; Joel et al., 2020; Puterman et al., 2020). Here, in alignment with a growing emphasis on precision medicine approaches to improving physical health, well-being, and productivity, we demonstrate that within-person features are also predictable by psychological and situation features. These dynamic features tend to be less studied, which has resulted in little knowledge about why people vary within-person in these behaviors. Our findings suggest that from a fairly prescribed set of personality, situational, and time features, we can identify *when* someone is going to procrastinate or feel lonely at a future timepoint – not just if they tend to procrastinate or feel lonely *in general*.

Notably, predictions were made assuming individuals have unique antecedents of each behavior. Although this equifinality is often described in theoretical models, it is rarely implemented in statistical models. Instead, statistical models use a circumscribed set of predictors that are assumed to impact people similarly, depending on their rank-order on the predictor (e.g., Borsboom et al., 2003). For example, procrastination is associated with Conscientiousness (Jackson et al., 2009). Typically, this suggests if people are feeling low in Conscientiousness markers (responsibility, organization) they would be more likely to procrastinate. However, we found that markers of Conscientiousness were not important antecedents of procrastinating for everyone, nor were they the most important in general (with 10-15% of the sample having Conscientiousness features as important predictors). People both procrastinate and feel lonely for many different reasons. As a result, prediction models that assume similar associations between predictors and outcomes for everyone may underestimate potential predictive validity.

In general, we found individual differences in every aspect of the models – in accuracy, in feature sets, and in the importance of specific features. For some people, we could highly accurately predict future behaviors, while for others, we could not. Similarly, people differed in which and the degree to which the domains were important. Together these findings paint a picture of a psychological system that is highly unique to an individual. Although there is a longstanding consensus that behavior is the output of such highly unique dynamic psychological systems that are impacted by situational features (Mischel & Shoda, 1995), these have remained elusive and often ignored in practice. Thus, the present study is an initial demonstrate of the empirical validity of such thinking. These participants demonstrated unique important situational and psychological features predicted future behavior.

The person situation debate revisited

Half a century ago, the seeming limits of behavioral prediction that sparked the Person-Situation Debate and led to research being formulated around the question of whether person or situation features matter more. While most agree that both matter, there are few examples of demonstrating the joint importance of them for the same outcome (c.f., Sherman et al., 2015). We found evidence that person and situation features were both important for most individuals, with only a minority demonstrating that person or situation features alone were most predictive of future procrastination or loneliness. In other words, the Person-Situation Debate was always a false debate. The dynamic relations among person, situation, and behavior and indicate that attempts to understand behavior must incorporate both (Funder, 2006) – at least for most people.

Not only are person and situation variables important, but they were also more important than time variables. Given that people have natural cycles of behavior that are regimented by time of day and day of week (Mathews, 1988; Larson, 1985), it would be natural to expect that behavior largely varies within and across people as a function of these cycles. For example, people work (behavior) less on the weekends and at night, which is a change their behavior. Similarly, time of day and day of week govern situations people can enter. Why were time variables not that important? It is likely that these time indices were already captured by the more proximal person or situational features. Time is likely important, but works through person and situation variables rather than being a separate factor.

Limitations and Conclusion

This study is not without its limitations. First, relatively low variance in procrastination and loneliness led us to drop a number of participants from analyses. Thus, the participants in the present study are only representative of participants who experienced somewhat frequent loneliness and procrastination. Second, we examined prediction over a two week interval for most participants, so long-term prediction accuracy is unclear. Finally, we demonstrated high accuracy and AUC on average when predicting behavior four hours in the future, making it unclear how such models perform at different time intervals. However, given that processes unfold at different speeds both within- and between-person, model performance likely varies as a function of interval.

The current study created personalized prediction models to help understand antecedents of future loneliness and procrastination. We found psychological and situational predictors did well in predicting within-person variations in these behaviors. However, in contrast to many years of methodological orthodoxy, the antecedents of these behaviors differed greatly across people. Thus, there is a need for more personalized assessments – not just longer assessments – but assessments that are tailored and important for the individual. Behavior appears to be highly predictable, so our next task is identifying personalized antecedents.

References

1. Ahadi, S., & Diener, E. (1989). Multiple determinants and effect size. *Journal of Personality and Social Psychology*, *56*(3), 398–406. https://doi.org/10.1037/0022-3514.56.3.398
2. Barrick, M. R., & Mount, M. K. (2005). Yes, personality matters: Moving on to more important matters. *Human performance*, *18*(4), 359-372. https://doi.org/10.1207/s15327043hup1804_3
3. Beck, E. D., & Jackson, J. J. (2020a). Idiographic traits: A return to Allportian approaches to personality. *Current Directions in Psychological Science*, *29*(3), 301-308.
4. Beck, E. D., & Jackson, J. J. (2020b). Consistency and change in idiographic personality: A longitudinal ESM network study. *Journal of Personality and Social Psychology*, *118*(5), 1080-1100.
5. Beck, E. D. & Jackson, J. J. (2021a). A mega-analysis of personality prediction: Robustness and boundary conditions. *Journal of Personality and Social Psychology*.
6. Beck, E. D., & Jackson, J. J. (2021b). Idiographic Personality Coherence: A Quasi Experimental Longitudinal ESM Study. *European Journal of Personality*.
7. Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203–219. https://doi.org/10.1037/0033-295X.110.2.203
8. Borkenau, P., & Ostendorf, F. (1998). The Big Five as states: How useful is the five-factor model to describe intraindividual variations over time?. *Journal of Research in Personality*, *32*(2), 202-221.

9. Butter, B., Kaiser, T., Arzt, S., Pannicke, B., Ginzinger, S., & Blechert, J. (2020). Time-Lagged Prediction of Food Craving with Elastic Net Regression and BISCWIT: Prediction Performance of Qualitative Distinct Predictor Types. https://doi.org/10.31234/osf.io/5hj3x
10. Cattell, R. B. (1945). The description of personality: Principles and findings in a factor analysis. *The American journal of psychology*, *58*(1), 69-90.
11. Elleman, L. G., McDougald, S. K., Condon, D. M., & Revelle, W. (2020). That takes the BISCUIT: Predictive accuracy and parsimony of four statistical learning techniques in personality data, with data missingness conditions. *European Journal of Psychological Assessment, 36*(6), 948–958. https://doi.org/10.1027/1015-5759/a000590
12. Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology, 37*(7), 1097–1126. https://doi.org/10.1037/0022-3514.37.7.1097
13. Epstein, S., & O'Brien, E. J. (1985). The person–situation debate in historical and current perspective. *Psychological Bulletin, 98*(3), 513–537. https://doi.org/10.1037/0033-2909.98.3.513
14. Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences*, *115*(27), E6106-E6115.
15. Fisher, A. J., & Soyster, P. (2019). Generating accurate personalized predictions of future behavior: A smoking exemplar.
16. Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, *33*(1), 1.

17. Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality*, *40*(1), 21-34.

18. Funder, D. C. (2009). Persons, behaviors and situations: An agenda for personality psychology in the postwar era. *Journal of Research in Personality*, *43*(2), 120-126.

19. Hampson, S. E., Goldberg, L. R., Vogt, T. M., & Dubanoski, J. P. (2007). Mechanisms by which childhood personality traits influence adult health status: educational attainment and healthy behaviors. *Health Psychology*, *26*(1), 121-125.

20. Jackson, J. J., & Beck, E. D. (2021). Using idiographic models to distinguish personality and psychopathology. *Journal of Personality*.

21. Jackson, J. J., Bogg, T., Walton, K. E., Wood, D., Harms, P. D., Lodi-Smith, J., Edmonds, G. W., & Roberts, B. W. (2009). Not all conscientiousness scales change alike: A multimethod, multisample study of age differences in the facets of conscientiousness. *Journal of Personality and Social Psychology, 96*(2), 446–459. https://doi.org/10.1037/a0014156

22. Joel, S., Eastwick, P. W., Allison, C. J., Arriaga, X. B., Baker, Z. G., Bar-Kalifa, E., ... & Wolf, S. (2020). Machine learning uncovers the most robust self-report predictors of relationship quality across 43 longitudinal couples studies. *Proceedings of the National Academy of Sciences*, *117*(32), 19061-19071.

23. Kenrick, D. T., & Funder, D. C. (1988). Profiting from controversy: Lessons from the person-situation debate. *American Psychologist, 43*(1), 23–34. https://doi.org/10.1037/0003-066X.43.1.23

24. Kim, H., Lee, S., Lee, S., Hong, S., Kang, H., & Kim, N. (2019). Depression prediction by using ecological momentary assessment, actiwatch data, and machine learning: Observational study on older adults living alone. *JMIR mHealth and uHealth*, *7*(10), e14149.

25. Larsen, R. J. (1985). Individual differences in circadian activity rhythm and personality. *Personality and Individual Differences*, *6*(3), 305-311.

26. Matthews, G. (1988). Morningness–eveningness as a dimension of personality: Trait, state, and psychophysiological correlates. *European Journal of Personality*, *2*(4), 277-293.

27. Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence.* University of Minnesota Press. https://doi.org/10.1037/11281-000

28. 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

29. 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.

30. Molenaar, P. C. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement*, *2*(4), 201-218.

31. Parrigon, S., Woo, S. E., Tay, L., & Wang, T. (2017). CAPTION-ing the situation: A lexically-derived taxonomy of psychological situation characteristics. *Journal of Personality and Social Psychology, 112*(4), 642–681. https://doi.org/10.1037/pspp0000111

32. Puterman, E., Weiss, J., Hives, B. A., Gemmill, A., Karasek, D., Mendes, W. B., & Rehkopf, D. H. (2020). Predicting mortality from 57 economic, behavioral, social, and psychological factors. *Proceedings of the National Academy of Sciences, 117*(28), 16273-16282.

33. Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A., Ziegler, M., Jones, A. B., & Funder, D. C. (2014). The Situational Eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, *107*(4), 677–718. https://doi.org/10.1037/a0037250

34. Rauthmann, J., & Sherman, R. (2016). Ultra-Brief Measures for the Situational Eight DIAMONDS Domains. *European Journal of Psychological Assessment*, *32*(2), 165-174.

35. Revelle, W., Condon, D. M., Wilt, J., French, J. A., Brown, A., & Elleman, L. G. (2016). Web and phone based data collection using planned missing designs. *Sage handbook of online research methods (2nd ed., p. 578-595). Sage Publications, Inc*.

36. Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality*, *43*(2), 137-145.

37. Sherman, R. A., Rauthmann, J. F., Brown, N. A., Serfass, D. G., & Jones, A. B. (2015). The independent effects of personality and situations on real-time expressions of behavior and emotion. *Journal of Personality and Social Psychology, 109*(5), 872–888. https://doi.org/10.1037/pspp0000036

38. Snyder, M., & Ickes, W. (1985). Personality and social behavior. *Handbook of social psychology*, *2*(3), 883-947.

39. Soto, C. J., & John, O. P. (2017). The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of Personality and Social Psychology, 113*(1), 117–143. https://doi.org/10.1037/pspp0000096

40. Titchener, E. B. (1898). The postulates of a structural psychology. *The Philosophical Review*, *7*(5), 449-465.

41. Watson, D., & Clark, L. A. (1999). The PANAS-X: Manual for the positive and negative affect schedule-expanded form.