

# Personalized Prediction of Behaviors and Experiences: An Idiographic Person–Situation Test



Emorie D. Beck<sup>1,2</sup> and Joshua J. Jackson<sup>3</sup>

<sup>1</sup>Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University;

<sup>2</sup>Department of Psychology, University of California, Davis; and <sup>3</sup>Department of Psychological and Brain Sciences, Washington University in St. Louis

Psychological Science

2022, Vol. 33(10) 1767–1782

© The Author(s) 2022

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09567976221093307

www.psychologicalscience.org/PS



## Abstract

A longstanding goal of psychology is to predict the things that people do and feel, but tools to accurately predict future behaviors and experiences remain elusive. In the present study, we used intensive longitudinal data ( $N = 104$  college-age adults at a midwestern university; total assessments = 5,971) and three machine-learning approaches to investigate the degree to which three future behaviors and experiences—loneliness, procrastination, and studying—could be predicted from past 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-specific perspective. Rather than pitting persons against situations, such an approach allows psychological phenomena, situations, and time to jointly predict future behaviors and experiences. We found (a) a striking degree of prediction accuracy across participants, (b) that a majority of participants' future behaviors are predicted by both person and situation features, and (c) that the most important features vary greatly across people.

## Keywords

idiographic, personality, prediction, machine learning, experience-sampling method (ESM), open data, open materials, preregistered

Received 6/19/21; Revision accepted 3/5/22

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 that people do and experience. Despite this persisting emphasis, accurately predicting future socioemotional behaviors and experiences remains elusive. Indeed, most of the existent research on prediction examines broad life outcomes (e.g., Beck & Jackson, 2022a; Joel et al., 2020). Whereas such broad life outcomes result from accumulating behaviors and experiences (e.g., Hampson et al., 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-specific, idiographic approach to the prediction of behavior and experiences, where the antecedents of everyday behavior and experiences are allowed to vary across people. We used three machine-learning approaches to

investigate the degree to which seven future behaviors and experiences, three of which we will focus on (future loneliness, procrastination, and studying), 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 measurement in psychology is that a measured construct is the same across people. A personality characteristic such as extraversion is

---

### Corresponding Author:

Emorie D. Beck, University of California, Davis

Email: edbeck@ucdavis.edu

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 et al., 2003; Fisher et al., 2018; Molenaar, 2004), then it is difficult to say whether extraversion predicts behavior and experiences 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, 2020b).

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, 2020a; Borkenau & Ostendorf, 1998; Molenaar, 2004). Across people, measures of the Big Five demonstrate five factors, whereas within people, 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 et al., 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 people 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 that individuals should respond to the “same” situations similarly. Even if they do, there is almost certainly heterogeneity in why individuals behave similarly. Consider, for example, small talk while at work. The same behavior in the same environment could be fueled by goals to get ahead and get along—such as avoiding an awkward interpersonal situation (get along) or making a long-term beneficial social connection (get ahead). Differences in person factors differ both across people and within them, making the interpretation of the situation quite subjective (i.e., idiographic).

### The Person–Situation Debate (Still)

Predicting behavior and experiences 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

### Statement of Relevance

Understanding and predicting behaviors and experiences are integral parts of everyday life. Situations clearly impact behavior, but individuals also have unique styles in their behaviors. Despite this, psychological studies of behavior and experiences typically examine person and situation factors separately or assumes that situations have similar behavioral or experiential consequences across people. But people exhibit different behaviors or have different experiences in the same contexts and similar behaviors in different contexts, indicating reciprocal person–situation relationships. In this study, we built personalized prediction models to predict future occurrences of procrastination, loneliness, and studying to test the extent to which people differed in their person- and situation-level antecedents. We found that these behaviors varied in how predictable they were for each person as well as that the person and situation features that best predicted each future behavior and experience varied widely across people. Such variability suggests that one-size-fits-all methods for predicting, explaining, and changing behavior and experiences are misguided at best and wholly wrong at worst.

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 association between predictor and behavior/experiences. If instead a behavior or experience is determined from many sources, theoretical estimates can go much higher than .3, although currently this possibility is only theoretical (Ahadi & Diener, 1989).

Another longstanding question in predicting behavior and experiences asks whether person or situation factors impact behavior and experiences more (Epstein & O'Brien, 1985). But most approaches hold either person or situation features constant to examine the association between the other and behavior/experiences (Kenrick & Funder, 1988). In other words, rather than viewing the triad of personality, situations, and behavior/experiences 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 as independent but not interactive influences (Sherman et al., 2015), leading people to continue studying these in isolation. Thus, it

remains unclear how situations coalesce with person factors to impact behavior and experiences.

## The Present Study

We argue that 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 and experiences (Renner et al., 2020). In the clinical-psychology domain, previous research has indicated (a) that future behaviors, such as smoking (Fisher & Soyster, 2019; Soyster et al., 2022) and food craving (Butter et al., 2020), can be predicted with high levels of accuracy using these methods and (b) 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 (a) the degree to which we can predict behavioral and experiential outcomes, (b) individual differences in how predictable such outcomes are, (c) whether certain domains (i.e., persons, situations, and time) out-predict others, (d) which features play the strongest role, and (e) whether and to what degree individuals differ in which features play the strongest role.

## Method

This study was preregistered on OSF (<https://osf.io/4nm5p/>); all data, analysis scripts, and results are also available on OSF (<https://osf.io/8ebyx/>). More details on the analyses and visual results depictions are available online at <https://github.com/emoriebeck/behavior-prediction> and in the R Shiny Web app at <https://emoriebeck.shinyapps.io/behavior-prediction/>. All data are completely deidentified. This study was approved by the institutional review board at Washington University in St. Louis (No. 201806124), and all data were collected in alignment with the American Psychological Association ethics code. Components of these data have been published elsewhere (Beck & Jackson, 2022b; Jackson & Beck, 2021).<sup>1</sup>

## Participants

Participants were 208 (71.96% female; age:  $M = 19.51$  years,  $SD = 1.27$ ) undergraduates at Washington University in St. Louis who enrolled in a study between October 2018 and December 2019. Sixty-nine participants identified as White, 67 as Asian, 34 as Black, and 30 as other race/ethnicity or mixed race/ethnicity (eight declined to answer). Nine participants were excluded for not completing any experience-sampling method

(ESM) surveys. The remaining participants completed a total of 8,403 surveys ( $M = 42.23$ ,  $SD = 24.01$ , range = 1–109). An additional 85 participants were excluded for having fewer than 40 ESM measurements, and 10 participants were excluded for having too little variance in one or both outcome measures.<sup>2</sup> The remaining 104 participants (72.82% female; age:  $M = 19.49$  years,  $SD = 1.31$ ) completed an average of 57.41 assessments ( $SD = 16.33$ , range = 40–109). Thirty-two participants identified as White, 33 as Asian, 14 as Black, and 16 as other (nine declined to answer).

## Measures

Participants responded to a battery of trait and ESM measures as part of the larger study (see the codebooks at <https://osf.io/8ebyx/>). The present study focused on a subset of preregistered ESM measures that were used to estimate idiographic prediction models.

**Personality and affect.** Personality was assessed using the full Big Five Inventory–2 (Soto & John, 2017) with a planned missing-data design (Revelle et al., 2016; <https://osf.io/pj9sy/>). Items capturing affect were a subset of the Positive and Negative Affect Schedule–Expanded Form (Watson & Clark, 1999), and items redundant with those on the Big Five Inventory–2 were removed. Each item was answered relative to what a participant was just doing on a 5-point Likert-type scale ranging from 1 (*disagree strongly*) to 5 (*agree strongly*).

**Situations.** Binary situation indicators were derived by asking research assistants to provide a list of the common social, academic, and personal situations in which they found themselves. From these, we derived a list of 20 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 past hour (1 = occurred, 0 = did not occur). Psychological features of situations were measured using the ultra-brief version of the “Situational Eight” DIAMONDS scale (Duty, Intellect, Adversity, Mating, pOsitivity, Negativity, Deception, Sociality; Rauthmann & Sherman, 2016) on a 3-point scale ranging from 1 (*not at all*) to 3 (*totally*).

**Time features.** Time features were created from the time stamps collected with each survey on the basis of approaches used in other studies of idiographic prediction (Butter et al., 2020; Fisher & Soyster, 2019). To do this, we created four dummy codes for time of day (morning, midday, evening, night) and seven for day of the week. 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-hr period (e.g., two-

$$\text{period sine} = \sin \frac{2\pi}{12} \times \text{Cumulative Time}_t \text{ and one-period} \\ \text{sine} = \sin \frac{2\pi}{24} \times \text{Cumulative Time}_t).$$

**Outcomes.** Procrastination, loneliness, and studying were assessed by asking participants to check a box if it had occurred in the past hour (1 = occurred, 0 = did not occur). Each was lagged such that time  $t$  features would predict time  $t + 1$  procrastination, loneliness, and studying.<sup>3</sup>

## Procedure

Participants responded to two types of surveys—trait and state ESM measures—for which they were paid separately. More information on the procedure of this study sample has been reported elsewhere (Beck & Jackson, 2022b; Jackson & Beck, 2021) and is available at <https://osf.io/8ebyx/>.

## Analytic plan

The present study used three machine-learning classification models: (a) elastic-net regression (ENR; Friedman et al., 2010), (b) the best-items scale that is cross-validated, correlation-weighted, informative, and transparent (BISCWIT; Elleman et al., 2020), and (c) random-forest models (Kim et al., 2019). More details on these methods and the procedure can be found at <https://osf.io/8ebyx/> but are summarized below.

Because we had 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 test which group of indicators is the most predictive of future procrastination, loneliness, and studying, we also examined these in several sets: (a) psychological indicators (personality + affect; 25), (b), situation indicators (binary + DIAMONDS; 25), and (c) full set (personality + affect + binary situations + DIAMONDS; 49). Each of these was also tested with and without the 18 timing indicators, for a total set of six combinations of the 68 features.

In each of these methods, we used cumulative rolling-origin forecast validation,<sup>4</sup> which comprised the first 75% of the time series (i.e., training data), and held out the remaining 25% of the data set for the test set. In the rolling-origin forecast validation, we used the first one third of the training-data time series as the initial set, five observations as the validation set, and set skip to

one (i.e., move two observations forward for each fold to reduce the number of folds to roughly equate tenfold cross-validation), which resulted in 10 to 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 hr previously). Tuning results are available for each participant, feature set, outcome, and model at <https://osf.io/8ebyx/> and on the Web app (“Model Tuning Figures”).

Out-of-sample prediction was tested on the basis of classification error and area under the receiver-operating-characteristic curve (AUC) in the test set (the last 25% of the time series). Classification error is a simple estimate of the percentage of the test sample that was correctly classified by the model. In addition, the AUC captures the trade-off between sensitivity and specificity. An AUC of .5 indicates binary classification at chance levels. AUCs are available at <https://osf.io/8ebyx/> and on the Web app (“ROC”).

ENR uses L1 (ridge) and L2 (least absolute shrinkage and selection operator [LASSO]) regularization to shrink coefficients on the basis of a penalty ( $\lambda$ ) that is tuned to minimize error using cross-validation. We tuned on the basis of penalty and mixture (set to 10 values each). ENR was performed using the *tidymodels* package (Kuhn & Wickham, 2020) in the R programming environment (Version 4.1.0) to estimate the ENR models by calling the `logistic_reg()`, setting the engine as “glmnet” (mode = “classification”; Friedman et al., 2010).

The BISCWIT is a correlation-based machine-learning technique. Using the `best.scales()` function in the *psych* package (Version 2.1.3; Revelle, 2020), we used rolling-origin validation to choose the best number of items to be retained. 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 models are a variant of decision-tree classification algorithms that use ensemble methods. Because random-forest models use 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” (mode = “classification”; Wright & Ziegler, 2017), with importance set to “permutation” in order to extract variable importance.



## Results

### ***Can we predict future procrastination, loneliness, and studying?***

First, we tested to what extent we could predict future incidences of procrastination, loneliness, and studying for each person by their previous assessments using ENR, BISCWIT, and random-forest models. Figure 1 presents histograms and descriptive statistics of accuracy and AUC across the full sample for each outcome and model. The same figures for all other outcomes are available at <https://osf.io/8ebyx/> and on the Web app (“Sample-Level Performance Distributions”). As is clear in the figure, predictive accuracy was high overall, with mean accuracy of .87 ( $Mdn = .91-.92$ ) for future loneliness, between .82 and .83 ( $Mdn = .88-.89$ ) for future procrastination, and between .77 and .79 ( $Mdn = .81-.83$ ) for future studying. Similarly, the AUC was also well above the .5 threshold; means ranged from .70 to .76 ( $Mdn = .75-.80$ ) for future loneliness, from .69 to .70 ( $Mdn = .70-.75$ ) for future procrastination, and from .63 to .68 ( $Mdn = .65-.76$ ) for future studying. Participant-level descriptive statistics across models, feature sets, and outcomes are available at <https://osf.io/8ebyx/> and on the Web app (“Model Performance Tables”).<sup>5</sup>

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

Figure 2 presents the median, 66%, and 95% range of classification accuracy for a random sample of 25 participants, ordered by the median accuracy. Other outcomes as well as the AUC for all outcomes are available at <https://osf.io/8ebyx/> and on the Web app (“Person-Level 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.

### ***Do psychological, situational, time, or full feature sets perform best?***

Table 1 presents the number of and percentage of participants whose best-performing 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 random-forest models indicated

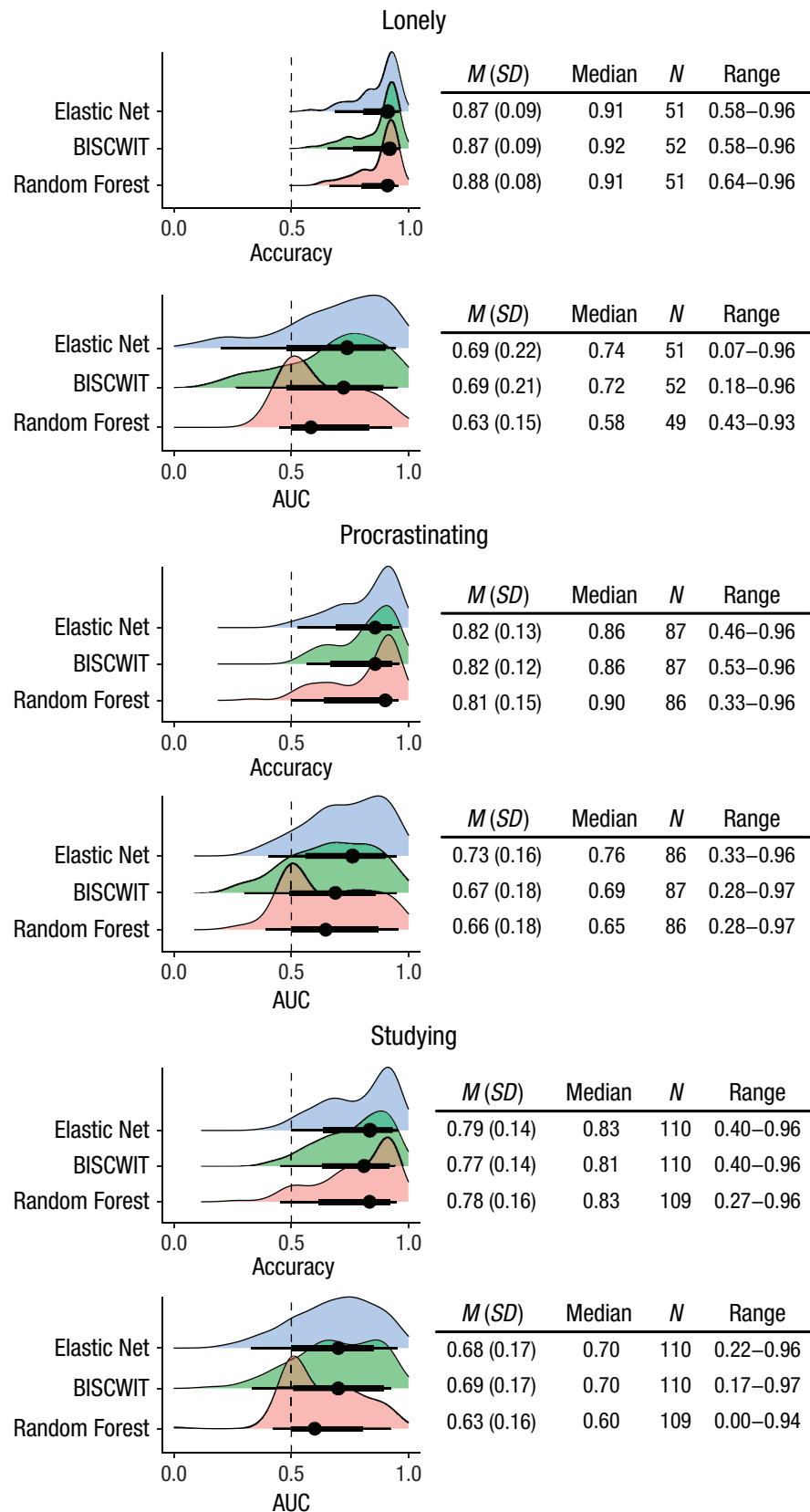
that situation feature models performed better than psychological feature models. We next examined the breakdown of selected features for each participant. As is clear in Figure 3, which shows proportions of features for all participants’ best-performing models for each method, there were individual differences in the proportion of psychological, situational, and time features. Some participants’ best-performing models included exclusively psychological or situational features, with most showing a varying mixture of both. Similar patterns were displayed across all outcomes, which are included at <https://osf.io/8ebyx/> and on the Web app (“Feature Proportions”).

### ***Which features are most associated with future procrastination, loneliness, and studying?***

To examine which features were most important, we extracted the top five features and calculated the proportion of the sample that had each feature in their top five. In Figure 4, larger, darker circles indicate that a higher proportion of individuals had this feature in their top five, whereas smaller, lighter circles indicate that a smaller proportion of individuals had this feature in their top five. Overall, most features were not shared by a majority of participants, and the maximum proportion of participants shared a single feature of 40% (energy level predicting arguing with a friend or family member). Across outcomes, the results were very similar, although which features were most frequent varied across outcomes, as can be seen at <https://osf.io/8ebyx/> and on the Web app (“Feature Frequency”).

Figure 4 also has several takeaways relative to categories of features.<sup>6</sup> First, across models, time features were less frequent, with the exception of linear, quadratic, and cubic trends (T12–T14) across the ESM period and a 24-hr sinusoidal diurnal cycle for ENR. Second, for ENR and BISCWIT, psychological features were slightly more frequent than situation features. Third, one consequence of the higher frequency of situation-feature random-forest models being selected than for the other two models was that situation features were both more frequent and more variable (more different sized circles) for the random-forest 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, and neither dominated the feature space.

Lastly, there are several specific features of note in Figure 4. For each outcome in Figure 4, there were some frequent features that made sense at face value. For example, the extraversion feature of energy level,



**Fig. 1.** Histograms showing classification accuracy and area under the receiver-operating-characteristic curve (AUC) for the best-performing models predicting loneliness, procrastinating, and studying. For each outcome, distributions from three models are shown: elastic net; best-items scale that is cross-validated, correlation-weighted, informative, and transparent (BISCWIT); and random forest. Shaded regions indicate the density of the data, dots represent medians, wide lines represent 66% confidence intervals, and thin lines represent 95% confidence intervals. AUC = area under the receiver-operating-characteristic curve.

the openness features of intellectual curiosity and aesthetic sensitivity, and Internet use were relatively frequent predictors of future procrastination across all models. Some were less intuitive; for example, happiness was infrequently predictive of future procrastination. For random-forest models, some situational features were also very frequent; for example, perceptions of situations from the DIAMONDS as inviting

sociality or being positive or negative predicted future procrastination. Similarly, for future loneliness, the extraversion feature of sociability and the neuroticism feature of emotional volatility were frequent. For random-forest models, the perception of a situation as positive or negative was the most frequent predictor, suggesting that the perceived valence of participants' current situations was predictive of future loneliness

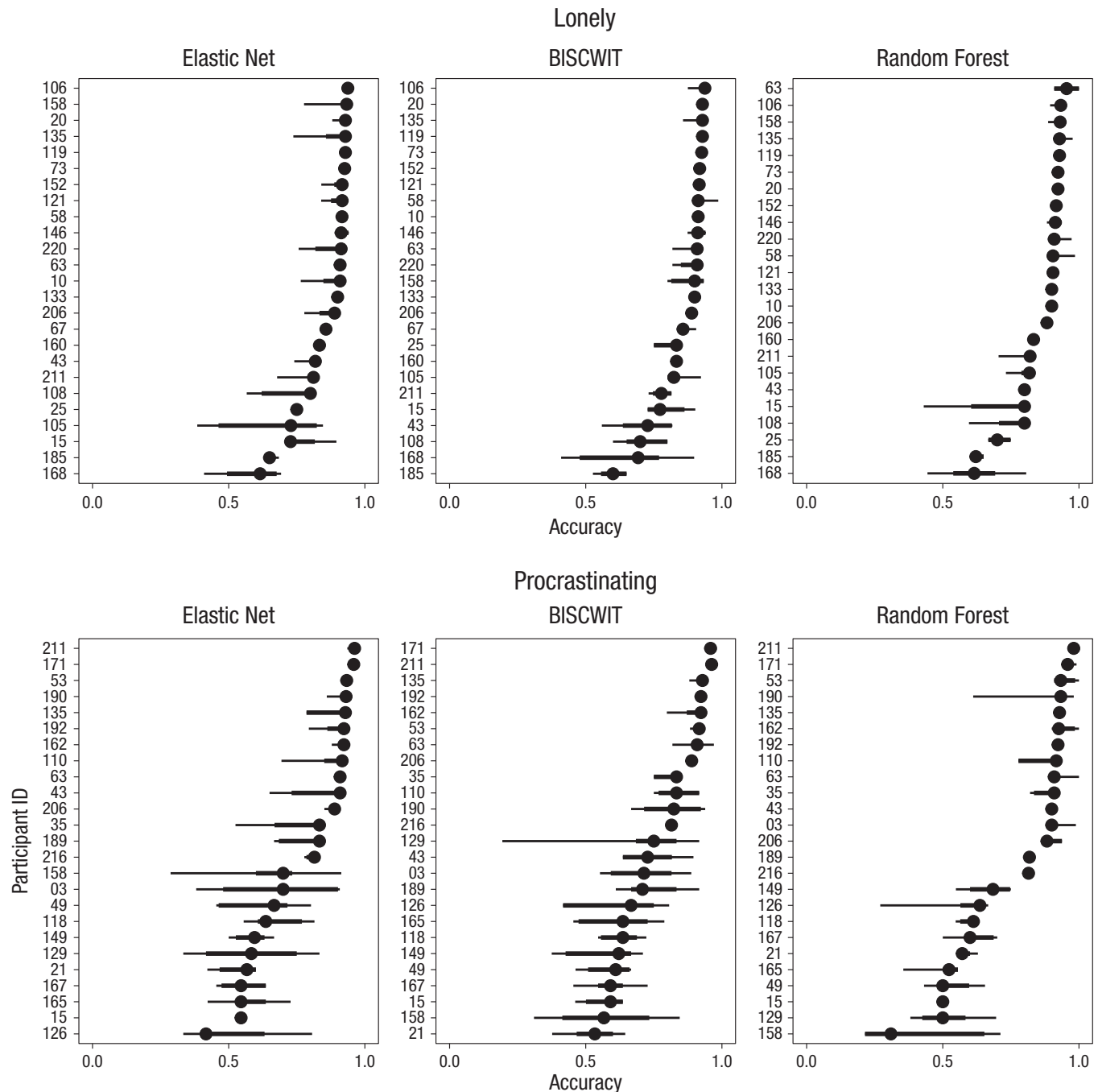
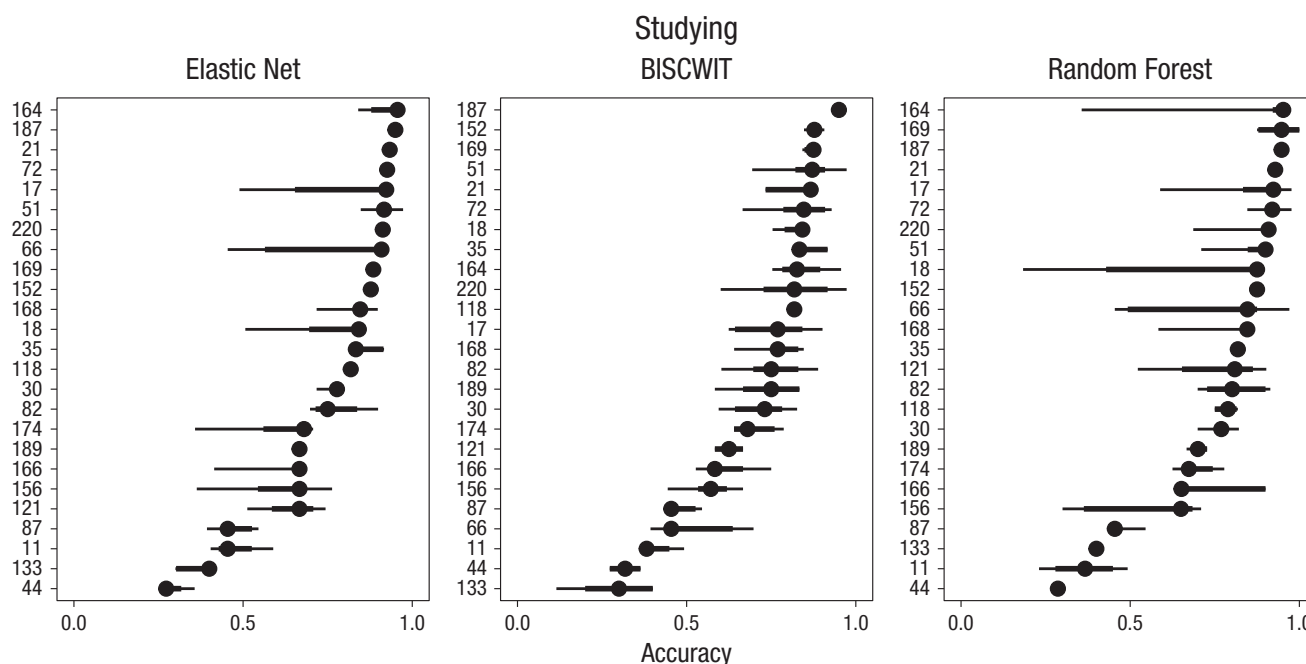


Fig. 2. (continued on next page)



**Fig. 2.** Person-level distributions of classification accuracy for 25 randomly selected sample participants, separately for classifiers predicting loneliness, procrastinating, and studying. Dots represent medians, wide lines represent 66% quantiles, and thin lines represent 95% quantiles. BISCWIT = best-items scale that is cross-validated, correlation-weighted, informative, and transparent.

approximately 4 hr later. Finally, for studying, the most frequent features were quite similar to those for procrastination; sociality and intellect from the DIAMONDS scale were frequently predictive of future studying in random-forest models as well as relatively more frequent time features than for loneliness.

### ***Do people vary in which features are most important?***

To demonstrate how people differ in which features were important, we have included participants' profiles of retained features along with their variable importance at <https://osf.io/8ebyx/> and on the Web app for each participant, outcome, and model ("Participant Coefficient Plots"). In addition, Figure 5 presents the profiles of all participants' coefficients (i.e., correlations) in their best-performing models for BISCWIT for future procrastination. All combinations of outcomes and models are available at <https://osf.io/8ebyx/> and on the Web app ("Participant Coefficient Profiles"). From the figure, it is clear that only a relatively small number of participants' best-performing models had time features. Moreover, even common features varied widely across people in presence, direction, and magnitude without exception. No two profiles are the same

even just in which features were included, let alone in direction and magnitude of the associations.

## **Discussion**

The current study investigated personalized, idiographic prediction models for seven socioemotional behaviors and experiences, three of which we focused on (feeling lonely, procrastinating, and studying in the future) and four of which are detailed at <https://osf.io/8ebyx/> (interacting, arguing, feeling tired, and feeling sick in the future). Rather than assuming that antecedents of different outcomes were shared, we used an idiographic approach to build  $N = 1$  personalized prediction models. Overall, three main conclusions emerged: First, psychological, situational, and time variables accurately predicted future everyday behaviors and experiences. Second, psychological and situational variables were both important, almost equally so, and neither was a predominant antecedent of behavior and experiences. Third, individual differences reigned supreme—people differed on how predictable outcomes were, which domains performed best, and which features were most important. Moreover, across the three behaviors and experiences, one experiential and two performative, the results were quite consistent. These findings



**Table 1.** Frequencies With Which the Full, Psychological, and Situation Feature Sets Were the Best Predictors of Loneliness, Procrastination, and Studying, With or Without the Inclusion of Time

Feature set and time	Accuracy						AUC					
	Elastic-net regression		BISCWIT		Random forest		Elastic-net regression		BISCWIT		Random forest	
	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%	<i>N</i>	%
Lonely												
Full												
Without	35	68.6%	33	63.5%	22	43.1%	13	25.5%	12	23.1%	14	28.6%
With	5	9.8%	5	9.6%	2	3.9%	7	13.7%	3	5.8%	4	8.2%
Psychological												
Without	9	17.6%	7	13.5%	6	11.8%	8	15.7%	19	36.5%	10	20.4%
With	0	0	3	5.8%	1	2.0%	9	17.6%	5	9.6%	2	4.1%
Situations												
Without	1	2.0%	4	7.7%	20	39.2%	8	15.7%	8	15.4%	13	26.5%
With	1	2.0%	0	0	0	0	6	11.8%	5	9.6%	6	12.2%
Procrastinating												
Full												
Without	48	55.2%	38	43.7%	34	39.5%	14	16.3%	14	16.1%	24	27.9%
With	12	13.8%	9	10.3%	4	4.7%	6	7.0%	6	6.9%	8	9.3%
Psychological												
Without	15	17.2%	17	19.5%	14	16.3%	19	22.1%	23	26.4%	18	20.9%
With	6	6.9%	7	8.0%	1	1.2%	12	14.0%	10	11.5%	12	14.0%
Situations												
Without	5	5.7%	12	13.8%	29	33.7%	19	22.1%	18	20.7%	17	19.8%
With	1	1.1%	4	4.6%	4	4.7%	16	18.6%	16	18.4%	7	8.1%
Studying												
Full												
Without	56	50.9%	45	40.9%	38	34.9%	22	20.0%	17	15.5%	34	31.2%
With	18	16.4%	6	5.5%	8	7.3%	9	8.2%	14	12.7%	9	8.3%
Psychological												
Without	17	15.5%	26	23.6%	14	12.8%	19	17.3%	27	24.5%	17	15.6%
With	9	8.2%	13	11.8%	4	3.7%	23	20.9%	19	17.3%	12	11.0%
Situations												
Without	6	5.5%	12	10.9%	36	33.0%	22	20.0%	18	16.4%	22	20.2%
With	4	3.6%	8	7.3%	9	8.3%	15	13.6%	15	13.6%	15	13.8%

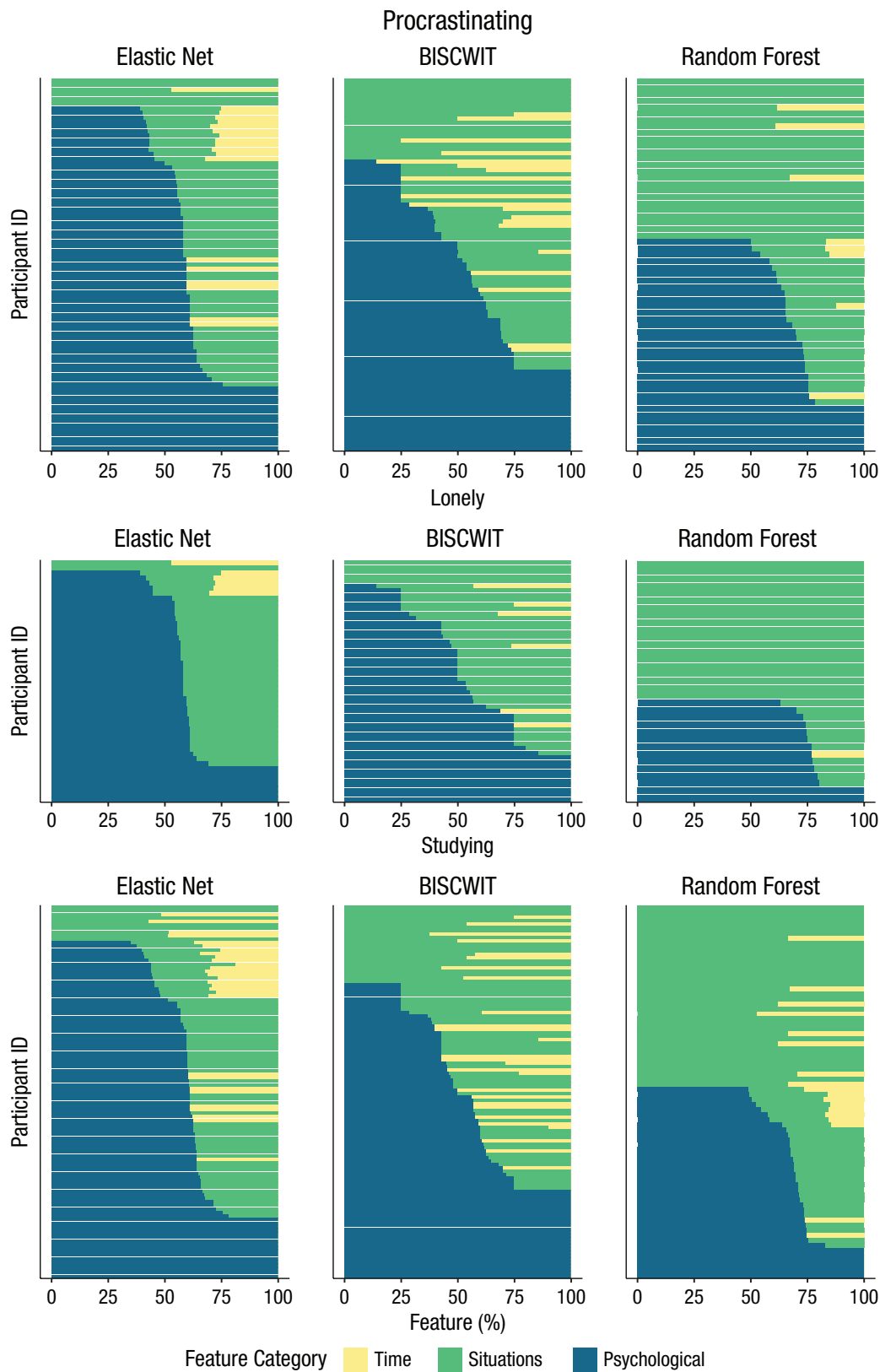
Note: AUC = area under the receiver-operating-characteristic curve; BISCWIT = best-items scale that is cross-validated, correlation-weighted, informative, and transparent.

indicate the utility of an idiographic approach to psychological assessment relative to standard between-person approaches that are routinely used.

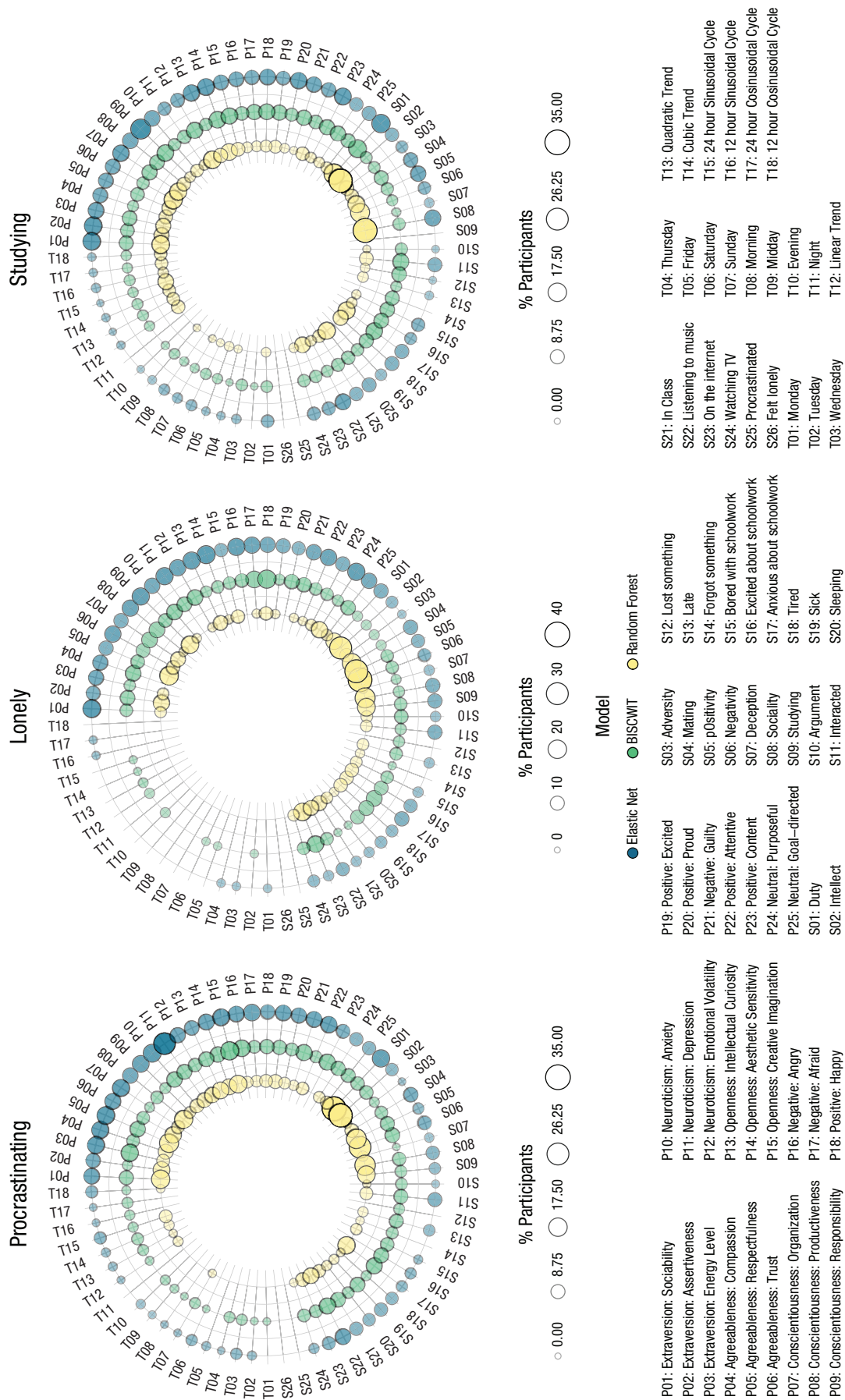
### ***On predicting more behaviors more of the time***

We found accurate out-of-sample prediction of future procrastination, studying, and feelings of loneliness when using a suite of psychological and situational factors. Predicting individual experiences and behaviors has long eluded psychologists. However, the results of

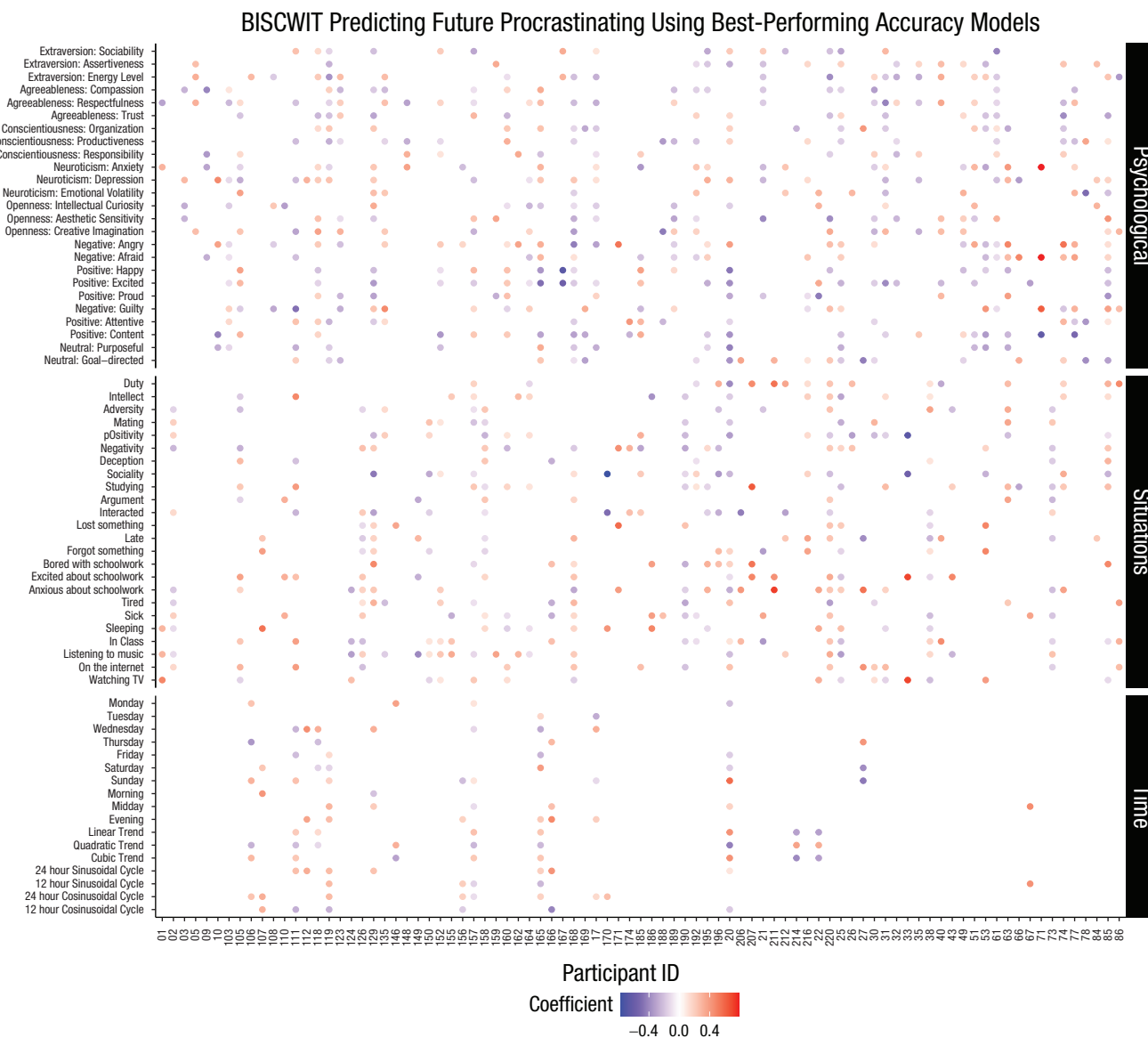
the present study suggest that this is because of two assumptions: (a) that the same psychological and situational antecedents should predict the same behaviors across people and (b) that psychological or situational features should out-predict the other. Neither of these assumptions holds. Whereas there are between-person individual differences in future loneliness, procrastination, and studying, there was also within-person variability in terms of how and when people demonstrated these behaviors and experiences. Typical prediction models within psychology have largely focused on which between-person features predict life outcomes



**Fig. 3.** Sequence plots showing the percentage of features from the psychological, situational, and time features sets for each participant for each outcome and model. BISCWIT = best-items scale that is cross-validated, correlation-weighted, informative, and transparent.



**Fig. 4.** Percentage of each feature appearing as a top-five variable-importance feature across the full sample for future procrastinating (left), loneliness (middle), and studying (right). Larger, darker circles indicate higher percentages, whereas smaller, lighter circles indicate lower percentages. Each feature's corresponding label is listed below the graphs. BISCWIT = best-items scale that is cross-validated, correlation-weighted, informative, and transparent.



**Fig. 5.** Coefficient profile (i.e., correlations) for all participants' best-performing BISCWIT (best-items scale that is cross-validated, correlation-weighted, informative, and transparent) models predicting future procrastination. Features are grouped by category: psychological, situation, and time.

or other aggregated behaviors (e.g., Beck & Jackson, 2022a; 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, study, or feel lonely at a future time point—not only whether they tend to procrastinate, study, or feel lonely *in general*.

Notably, predictions were made under the assumption that individuals have unique antecedents of each outcome. 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 that 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 procrastinate, study, 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 very accurately predict future behaviors, whereas for others, we could not. Similarly, people differed in which and to what degree 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 demonstration of the empirical validity of such thinking. Participants differ in the important situational and psychological features that predict future behavior and experiences.

Pragmatically, that people differed in (a) which features were predictive of future procrastination, studying, and loneliness and (b) the proportion of these that were situational and psychological features has implications for behavioral prediction in applied contexts, such as outpatient clinical work, worker performance and well-being, and more. The results of the present study suggest that using a very short, standard battery of psychological and situational indicators may not well capture the antecedents of these or other behaviors but that machine-learning approaches are useful as feature-selection tools when used in conjunction with larger batteries. However, what this study did not address is the use of true idiographic assessment in which participants respond to unique, tailored batteries of items rather than (or in addition to) standard batteries. It is possible that such approaches may be useful in reducing participant burden and improving overall prediction. However, we expect that such an approach would only broaden the realm of antecedents in behaviors, highlighting the broad range of individual differences within and across people in psychological and situational antecedents.

### ***The person–situation debate revisited***

Half a century ago, the seeming limits of behavioral prediction sparked the person–situation debate and led to research being formulated around the question of whether person or situation features matter more. Although most scholars agree that both matter, there

are few examples of demonstrating the joint importance of them for the same outcome (cf. Sherman et al., 2015). We found evidence that person and situation features were both important for most individuals, and only a minority demonstrated that person or situation features alone were most predictive of future procrastination, loneliness, and studying. In other words, the person–situation debate was always a false debate. The dynamic relations among person, situation, and behavior/experiences indicate that attempts to understand behavior and experiences 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 and experiences that are regimented by time of day and day of week (Larsen, 1985; Matthews, 1988), it would be natural to expect that behavior and experiences largely vary within and across people as a function of these cycles. For example, people work less on the weekends and at night, which is a change in their behavior and experiences. Similarly, time of day and day of week govern situations that people can enter. Although across the three focal outcomes (future procrastination, studying, and loneliness) as well as the five additional outcomes we tested as robustness-checks models without time were less likely to be selected as the best-performing feature set, there was some variability across outcomes. For example, time features were proportionally more prevalent for procrastination, interacting with friends/family members, and studying than they were for feeling lonely, tired, or sick. Although it may be expected that work-related behaviors, such as procrastination and studying, may be associated with time, that feeling tired was not strongly associated with day of the week, time of day, diurnal cycles, and more was more surprising. It is possible that time variables were less important because they were already captured by the more proximal person or situational features. In other words, time is likely important but works through person and situation variables rather than being a separate factor. However, the full pattern of results across the seven tested outcomes does not paint a clear picture of when this is true.

### ***Limitations and conclusion***

This study is not without its limitations. First, relatively low variance in each of the outcomes led us to drop a number of participants from analyses. Thus, the participants in the present study are representative only of participants who experienced somewhat frequent loneliness, procrastination, and studying, as well as the other outcomes. Second, we examined prediction over a 2-week interval for most participants, so long-term



prediction accuracy is unclear. Third, this study was the first in a line of planned research focusing on individual differences in the accuracy, antecedents, and timing of prediction models of future behaviors and experiences. Finally, the participants in the sample used in this study were predominantly female college students at a private university in the midwestern United States. Thus, we were unable to address age, language, cultural, or other sociodemographic impacts on the observed pattern of results. Future studies must address challenging questions about the long-term consistency of antecedents, their consequences, and more.

In the current study, we created personalized prediction models to help understand antecedents of future loneliness, procrastination, and studying. We found that 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 only longer assessments—but assessments that are tailored and important for the individual. Behavior and experiences appear to be highly predictable, so our next task is identifying personalized antecedents.

## Transparency

*Action Editor:* Mark Brandt

*Editor:* Patricia J. Bauer

### Author Contributions

E. D. Beck and J. J. Jackson conceptualized the idea, cowrote the Introduction and Discussion, and edited the manuscript. E. D. Beck collected the data, wrote the preregistration, ran the analyses, wrote the Method and Results, and prepared all materials on OSF (<https://osf.io/8ebyx/>) and Github (<https://github.com/emoriebeck/behavior-prediction>). Both of the authors approved the final manuscript for submission.

### Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

### Funding

The preparation of this article was supported by a grant from the National Science Foundation (BCS-1125553) to J. J. Jackson. Data collection was supported by a Society for Personality and Social Psychology Inside the Grant Panel grant awarded to E. D. Beck. E. D. Beck was supported by National Institute on Aging Grants T32 AG00030-3, R01-AG067622, and R01-AG018436.

### Open Practices

All data, materials, and analysis scripts have been made publicly available via OSF and can be accessed at <https://osf.io/8ebyx/> and <https://github.com/emoriebeck/behavior-prediction>. The design and analysis plans were preregistered at <https://osf.io/4nm5p/>. This article has received the badges for Open Data, Open Materials, and

Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



## ORCID iD

Emorie D. Beck  <https://orcid.org/0000-0002-8152-3510>

## Notes

1. The data used in the present study were collected as a part of a personalized intervention study targeting one social (procrastination) and one emotional (loneliness) behavior or experience among college students. In this longitudinal burst study, participants completed baseline surveys and a wave of ESM and then were later contacted with the opportunity to be part of the intervention component. Because of delays and the onset of the COVID-19 pandemic, a number of participants became ineligible for the intervention component of the study, leaving more than 200 participants with data designed to help understand antecedents and consequences of procrastination and loneliness in students' lives. Rather than simply starting over, we decided to use these data to refine and validate our personalized prediction procedure before resuming the study with a new cohort of participants. In the present study, we chose to focus on procrastination and loneliness in order to validate our predictions of these socioemotional behaviors and experiences. However, thanks to helpful comments during the review process, we included five additional outcomes: feeling sick, feeling tired, studying, arguing with a friend or family member, and interacting with a friend or family member. These results are available at <https://osf.io/8ebyx/> as well as in the R Shiny Web app.
2. Participants were excluded on an outcome-by-outcome basis—that is, if a participant had too little variance in Outcome 1, they were excluded from those analyses, but if they had enough variance in Outcome 2, they were included in those analyses. The 10 participants here are those who had too little variance across outcomes.
3. In addition to examining future procrastination, loneliness, and studying, we also examined four other future outcomes. These outcomes, as well as studying, were deviations from our preregistered analysis plan, to replicate our main conclusions on the basis of suggestions received during the review process. The full results of these models are available at <https://osf.io/8ebyx/>. On the whole, the results of these unregistered models were largely congruent with the results of the preregistered loneliness and procrastination models in the aggregate, with a few exceptions that we will note in later sections.
4. We preregistered the use of tenfold cross-validation on the training set. Because of how tenfold cross-validation separates and combines folds, this would have resulted in validation sets that temporally occurred before observations in the other nine folds. As an alternative, we elected to use rolling-origin validation. In rolling origin, the validation set always occurs after the earlier "rolls" (as opposed to folds).
5. Of the additional four outcomes we tested, the results largely converged. Three (argument, sick, and tired) had median

accuracy between .81 and .92. Interaction had a lower accuracy, with median accuracies between .69 and .72 ( $SDs = .14-.17$ ).

6. This pattern was largely replicated in the additional five behaviors and experiences. Across these, the prevalence of any feature was heavily right skewed, ranging from 0 (38.14% of all timing frequencies) to 17.72 ( $M = 2.40$ ,  $Mdn = 1.80$ ,  $SD = 2.87$ ). The non-zero frequencies were largely driven by two outcomes: interacting with family or friends and studying. However, the mean frequencies of time features relative to psychological and situational features across these outcomes remained low.

## References

- 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>
- 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](https://doi.org/10.1207/s15327043hup1804_3)
- Beck, E. D., & Jackson, J. J. (2020a). Consistency and change in idiographic personality: A longitudinal ESM network study. *Journal of Personality and Social Psychology*, 118(5), 1080–1100.
- Beck, E. D., & Jackson, J. J. (2020b). Idiographic traits: A return to Allportian approaches to personality. *Current Directions in Psychological Science*, 29(3), 301–308.
- Beck, E. D., & Jackson, J. J. (2022a). A mega-analysis of personality prediction: Robustness and boundary conditions. *Journal of Personality and Social Psychology*, 122(3), 523–553.
- Beck, E. D., & Jackson, J. J. (2022b). Idiographic personality coherence: A quasi experimental longitudinal ESM study. *European Journal of Personality*, 36(3), 391–412.
- Borkenau, P., & Ostendorf, F. (1998). The Big Five as states: How useful is the five-factor model to describe intra-individual variations over time? *Journal of Research in Personality*, 32(2), 202–221.
- 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>
- 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>
- 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>
- 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>
- 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>
- 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, USA*, 115(27), E6106–E6115.
- Fisher, A. J., & Soyster, P. (2019). Generating accurate personalized predictions of future behavior: A smoking exemplar. PsyArXiv. <https://doi.org/10.31234/osf.io/e24v6>
- Fodor, J. A. (1968). The appeal to tacit knowledge in psychological explanation. *The Journal of Philosophy*, 65(20), 627–640.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1). <https://doi.org/10.18637/jss.v033.i01>
- Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality*, 40(1), 21–34.
- 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.
- 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.
- Jackson, J. J., & Beck, E. D. (2021). Using idiographic models to distinguish personality and psychopathology. *Journal of Personality*, 89(5), 1026–1043.
- 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 multi-method, 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>
- Joel, S., Eastwick, P. W., Allison, C. J., Arriaga, X. B., Baker, Z. G., Bar-Kalifa, E., Bergeron, S., Birnbaum, G. E., Brock, R. L., Brumbaugh, C. C., Carmichael, C. L., Chen, S., Clarke, J., Cobb, R. J., Coolsen, M. K., Davis, J., de Jong, D. C., DeBrot, A., DeHaas, E. C., . . . 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, USA*, 117(32), 19061–19071.
- 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>
- 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), Article e14149. <https://doi.org/10.2196/14149>
- Kuhn, M., & Wickham, H. (2020). *Tidymodels: A collection of packages for modeling and machine learning using tidyverse principles*. <https://www.tidymodels.org>

- Larsen, R. J. (1985). Individual differences in circadian activity rhythm and personality. *Personality and Individual Differences*, 6(3), 305–311.
- Matthews, G. (1988). Morningness–eveningness as a dimension of personality: Trait, state, and psychophysiological correlates. *European Journal of Personality*, 2(4), 277–293.
- 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>
- 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>
- 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.
- 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.
- Parrigon, S., Woo, S. E., Tay, L., & Wang, T. (2017). CAPTIONing 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>
- 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, USA*, 117(28), 16273–16282.
- 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>
- Rauthmann, J. F., & Sherman, R. (2016). Ultra-brief measures for the Situational Eight DIAMONDS domains. *European Journal of Psychological Assessment*, 32(2), 165–174.
- Renner, K. H., Klee, S., & von Oertzen, T. (2020). Bringing back the person into behavioural personality science using Big Data. *European Journal of Personality*, 34(5), 670–686.
- Revelle, W. (2020). *psych: Procedures for personality and psychological research* (Version 2.1.3). <https://CRAN.R-project.org/package=psych>
- 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. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (2nd ed., pp. 578–595). SAGE.
- Roberts, B. W. (2009). Back to the future: Personality and assessment and personality development. *Journal of Research in Personality*, 43(2), 137–145.
- 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>
- Snyder, M., & Ickes, W. (1985). Personality and social behavior. *Handbook of Social Psychology*, 2(3), 883–947.
- 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>
- Soyster, P. D., Ashlock, L., & Fisher, A. J. (2022). Pooled and person-specific machine learning models for predicting future alcohol consumption, craving, and wanting to drink: A demonstration of parallel utility. *Psychology of Addictive Behaviors*, 36(3), 296–306. <https://doi.org/10.1037/adb0000666>
- Titchener, E. B. (1898). The postulates of a structural psychology. *The Philosophical Review*, 7(5), 449–465.
- Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the Positive and Negative Affect Schedule-Expanded Form*. The University of Iowa.
- Wright, M. N., & Ziegler, A. (2017). ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*, 77(1). <https://doi.org/10.18637/jss.v077.i01>