Research Article

Psychological Science
2020, Vol. 31(10) 1222–1235

DOI: 10.1177/0956797620959594
www.psychologicalscience.org/PS

Using Machine Learning to Generate Novel Hypotheses: Increasing Optimism About COVID-19 Makes People Less Willing to Justify Unethical Behaviors

Abhishek Sheetal[1], Zhiyu Feng[2], and Krishna Savani[1]
[1]Nanyang Business School, Nanyang Technological University, and [2]Department of Organization and Human Resources, Business School, Renmin University of China

Abstract
How can we nudge people to not engage in unethical behaviors, such as hoarding and violating social-distancing guidelines, during the COVID-19 pandemic? Because past research on antecedents of unethical behavior has not provided a clear answer, we turned to machine learning to generate novel hypotheses. We trained a deep-learning model to predict whether or not World Values Survey respondents perceived unethical behaviors as justifiable, on the basis of their responses to 708 other items. The model identified optimism about the future of humanity as one of the top predictors of unethicality. A preregistered correlational study (N = 218 U.S. residents) conceptually replicated this finding. A preregistered experiment (N = 294 U.S. residents) provided causal support: Participants who read a scenario conveying optimism about the COVID-19 pandemic were less willing to justify hoarding and violating social-distancing guidelines than participants who read a scenario conveying pessimism. The findings suggest that optimism can help reduce unethicality, and they document the utility of machine-learning methods for generating novel hypotheses.

Keywords
COVID-19, machine learning, optimism, neural network, unethical behavior, open data, open materials, preregistered

Received 4/8/20; Revision accepted 8/27/20

Unethical behaviors can have substantial consequences in times of crisis. For example, in the midst of the COVID-19 pandemic, many people hoarded face masks and hand sanitizers; this hoarding deprived those who needed protective supplies most (e.g., medical workers and the elderly) and, therefore, put them at risk. Despite escalating deaths, more than 50,000 people were caught violating quarantine orders in Italy, putting themselves and others at risk. Governments covered up the scale of the pandemic in that country, thereby allowing the infection to spread in an uncontrolled manner. Thus, understanding antecedents of unethical behavior and identifying nudges to reduce unethical behaviors are particularly important in times of crisis.

Unethical behavior is a major research topic within psychology and other behavioral sciences. Researchers have identified dozens of predictors of unethical behavior, including characteristics of the person and of the situation. For example, people are more ethical when honesty is the descriptive norm or the salient prescriptive norm, when they are reminded of God or religion, when they are feeling less anxious, when they

Corresponding Authors:
Zhiyu Feng, Department of Organization and Human Resources, Business School, Renmin University of China, 59 Zhongguancun St., Haidian District, Beijing 100872, China
E-mail: fengzhiyu@rmbs.ruc.edu.cn

Abhishek Sheetal, Nanyang Business School, Nanyang Technological University, 50 Nanyang Ave., Block S3, B1A-20, Singapore 639798
E-mail: asheetal@ntu.edu.sg

Krishna Savani, Nanyang Business School, Nanyang Technological University, 50 Nanyang Ave., Block S3, 01C-76, Singapore 639798
E-mail: ksavani@ntu.edu.sg

are not depleted, and so on (for reviews, see Ellemers, van der Toorn, Paunov, & van Leeuwen, 2019; Gerlach, Teodorescu, & Hertwig, 2019). However, these interventions to reduce unethical behaviors cannot be easily implemented in the field. For example, during a stay-at-home order, it would be advisable for people to *not* follow the social norm—if there are too many people outside, it is advisable to stay indoors; if there is no one outside, then it is safe to go out.

In this research, we sought to identify novel antecedents of unethical behavior by examining existing data sets that were designed for other purposes (Goldstone & Lupyan, 2016). Specifically, we used the World Values Survey (WVS; Inglehart et al., 2014), which contains measures of unethical behavior. Using this data set, past research has identified a number of predictors of unethical behavior, including Big Five personality traits (Simha & Parboteeah, 2019); happiness and belief in free will (Martin, Rigoni, & Vohs, 2017); filial piety and materialism (Cullen, Parboteeah, & Hoegl, 2004); political orientation, pride in nation, generalized trust, and satisfaction with household income (Sommer, Bloom, & Arikan, 2013); and religiosity, risk aversion, interest in politics, and trust in the political system (Dong & Torgler, 2009). Most of these are individual-difference variables that cannot easily be experimentally manipulated and, therefore, cannot be easily used by policymakers to help arrest the COVID-19 pandemic. However, given that the WVS asked respondents hundreds of questions, there are likely other predictors of unethical behavior in the data that researchers have not yet examined.

There are many ways to generate novel hypotheses from large data sets. Researchers could examine which variables in the WVS data set are most strongly correlated with the variables measuring unethical behavior. Researchers could run regressions with regularization methods (e.g., lasso, ridge, and elastic net) to select an optimum number of predictors (Hastie, Tibshirani, & Friedman, 2009). However, the large proportion of missing values in the WVS data set limits the use of these regression-based methods, as they can be run only on observations without any missing values. Further, linear regressions require that key assumptions, such as homoscedasticity and independently, identically, and normally distributed residuals, are met. Researchers could also use machine-learning methods, such as random forest, gradient boost, *k*-nearest neighbors, support-vector machine, and neural networks (Alpaydin, 2020). These methods do not make any auxiliary assumptions and can impute even large volumes of missing data (either in a separate stage prior to modeling or during the process of modeling).

Once a machine-learning model is trained, we can query it to identify the top predictors. Certain challenges emerge, however, when attempting to do so. In particular, identifying the top predictors in a large data set is a *nondeterministic-polynomial-time-complete* problem (Karp, 1975), and there is no known closed-form solution to this problem. Only approximate solutions are possible for all problems of this class, and any given approximate solution can neither be proven to be the best solution nor be proven to be an inferior solution. Various regression-based and machine-learning methods merely provide a possible solution; neither the similarity of solutions provided by different methods nor their difference is guaranteed (Reyzin, 2019). Thus, researchers can freely choose any method to identify the top predictors in a large data set as long as the data meet the assumptions of the method and researchers have sufficient computing power.

In the present research, we chose to use a deep neural network to generate novel hypotheses about antecedents of unethical behavior in the WVS. We chose deep learning because this method has been the source of recent groundbreaking discoveries in physics (e.g., novel particles; Baldi, Sadowski, & Whiteson, 2014), chemistry (e.g., novel materials; Jha et al., 2018), and biology (e.g., novel antibiotics; Stokes et al., 2020). Further, regression-based methods limit the range of possible predictor variables to those that have a mostly linear and direct relationship with the dependent variable; in contrast, deep-learning models can capture nonlinear effects and complex interactions.

Statement of Relevance

This research is likely to be of interest to all researchers in the social-behavioral sciences who work on hypothesis testing, because it demonstrates a general method to generate novel hypotheses using machine-learning techniques. This method can be applied in any field in which researchers have access to reasonably large data sets. The present research significantly expands the scope of machine learning in psychology, which has been nearly exclusively focused on prediction until now. The current research demonstrates that machine-learning methods can be used simultaneously for prediction and for theory development. The context in which we tested the hypothesis generated by the machine-learning method—unethical behaviors surrounding the COVID-19 pandemic—is immediately relevant to policymakers and the general public who wish people to act in a more ethical manner to arrest the pandemic. Our experimental materials provide messages that policymakers and public-interest organizations can immediately use.

Study 1: Machine Learning

The goal of Study 1 was to identify novel predictors of unethical behavior using a deep-learning model. We used the WVS because it contains items that could be used to measure people's willingness to engage in unethical behaviors, as well as items associated with variables that might predict their willingness to engage in unethical behaviors. Many predictors of unethical behavior uncovered by the deep-learning model might already have been examined in past research. However, it is also possible that some predictors might not match those that have been discussed in the literature, which would be an interesting and novel result.

Method

Figure 1 illustrates the procedure that we used for the machine-learning analysis. The code that was used to build the model and the final model is available in the OSF data repository for this project (https://doi.org/10.17605/OSF.IO/A6Y7R). We used a desktop computer running Linux with a 16-thread CPU, 128 GB RAM, and four GeForce GTX1070 8GB graphics cards (Nvidia, Santa Clara, CA) and running OpenCL drivers to support the graphics card's interface with R. We used Intel's *PlaidML* libraries to conduct the lower level matrix-multiplication programming on the graphics card.

Data set. The WVS contains 348,532 rows of data, representing individuals, and 975 variables with at least one nonmissing data point. The surveys were conducted in 98 countries across six waves: 1981–1984, 1989–1993, 1994–1998, 1999–2004, 2005–2009, and 2010–2014. Different items were administered in different waves and in different countries within each wave. The survey was administered during a face-to-face interview at respondents' homes. Participants responded to between 17 and 361 items. The survey was translated into all local languages that were spoken by at least 15% of the country's residents. The survey items were decided by a multidisciplinary group of social scientists and spanned multiple domains: perceptions of life, environment, work, family, politics and society, religion and morals, national identity, security, science, and sociodemographics. We chose this data set because it contained a measure of unethical behavior and a large number of other items that could potentially serve as predictors.

Outcome variable. Following past research (Martin et al., 2017), we used four items administered in all six waves of the WVS as a measure of justifiability of unethical behavior: "Justifiable: claiming government benefits to which you are not entitled" (variable f114 in the WVS data set), "Justifiable: avoiding a fare on public transport" (variable f115), "Justifiable: cheating on taxes if you have a chance" (variable f116), and "Justifiable: someone accepting a bribe in the course of their duties" (variable f117; α = .79). Participants responded to these items on a 10-point scale ranging from 1, *never justifiable*, to 10, *always justifiable*. The WVS included additional items about the justifiability of unethical behavior (e.g., variable f114_01, "Justifiable: stealing property"), but these items were not administered in all six waves, and thus we did not use them.

As the modal response for each of these four items was "never justifiable" (chosen by 58%, 57%, 63%, and 74% of the respondents, respectively), we converted participants' responses from a continuous variable to a binary variable, with 0 indicating "never justifiable" and 1 indicating "justifiable" (i.e., if participants selected a response greater than 1). Of the 348,532 individuals in the data set, 12,226 had a missing value for all four items and were thus dropped from the analyses. Of the remaining 336,306 individuals, we coded 139,412 individuals (41.45%) who selected "not justified" to all four items as *ethical* and the 196,894 remaining individuals (58.55%) as *unethical*.

Data cleaning before imputation. We first dropped columns in the data file that were created by WVS researchers (e.g., weights, administrative codes, and factor variables), items that contained a large number of categorical responses or open-ended responses, and items that represented "don't know" or "none" responses. Next, we dummy-coded all categorical response options, including the WVS wave (variable s002 in the WVS data file) and country (variable s003; "Bosnia" and "Srpska" Republic were merged with "Bosnia and Herzegovina"). Further, we added dummy variables indicating the International Organization for Standardization (ISO) region code and ISO subregion code for each country. The full list of variables deleted and dummy-coded is available in this project's OSF data repository (https://doi.org/10.17605/OSF.IO/A6Y7R).

Splitting the data. To test our model, we used the *holdout technique*. We split the above data set randomly into two parts. We used 90% of the observations for the training or model-building phase (called the *seen* data) and reserved the remaining 10% of the data to test the model (called the *unseen* data). To ensure that the unseen data did not influence the training data in any way, we performed this split before imputing missing values in the seen data. This way, the unseen data could not influence the deep-learning model in any way.

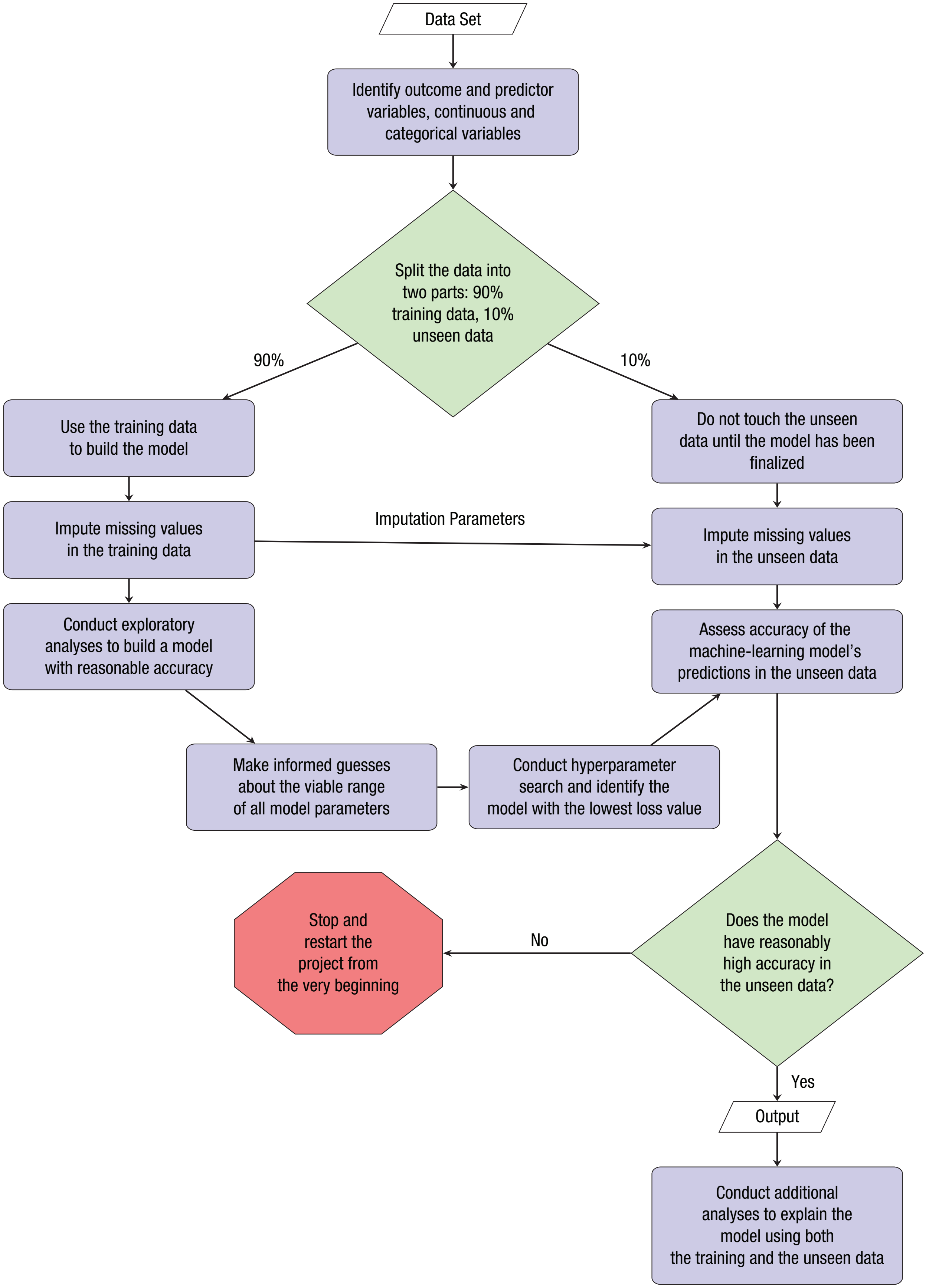

Fig. 1. Flowchart showing the machine-learning procedure used in Study 1.

Imputing missing data. Because the WVS administered different items in different waves and countries, at least 60% of the data were missing for every respondent. Thus, it was not possible to consider all items administered in the WVS as predictors without imputing missing data, as every participant would otherwise drop out from the data set. Overall, 75.55% of the values were missing in our truncated data set. Because the machine-learning package that we used would also drop all rows with any missing value, we imputed all missing values in the seen data using a machine-learning algorithm.

Although traditional approaches recommend multiple imputation only when the missing data are sparse and missing at random, newer machine-learning-based imputation methods can be used when data are missing systematically (e.g., many items were not administered in multiple waves and multiple countries in the WVS) and when a majority of the data are missing (Deng, Chang, Ido, & Long, 2016). To impute the missing values, we used a random-forest algorithm, the *missRanger* package in *R*, which is an improved version of the older *missForest* package (Stekhoven & Bühlmann, 2012). We ran 15 iterations of the imputation, which completed in 14 days, using these parameters: "num.trees = 100 trees, maxiter = 15, respect.unordered.factors = TRUE, splitrule = "extratrees."

Once we imputed missing information in the seen data, we appended the unseen data to the imputed seen data and imputed all missing values in the unseen data using the *missRanger* package. This way, any bias in the imputation would transfer over from the seen data to the unseen data, so if the imputation were unreliable or erroneous, the model would have low accuracy when predicting the nonimputed dependent variable in the unseen data. Note that the dependent variable is never imputed. If the model had reasonably high accuracy in predicting the dependent variable in the unseen data, that would mean that the imputation was sufficiently accurate. In any case, if the imputation process yielded spurious predictors of unethical behavior, then we would fail to find causal support for these spurious predictors.

Data cleaning (second round). After all missing values were imputed, we excluded variables f114, f115, f116, and f117, which were used to compute the dependent variable. We also excluded all variables between f114 and f144, as well as variable f199, because these variables were asked in the same format as f114 to f117 (i.e., starting with "Justifiable:") and thus might be correlated with f114 through f117 because of common method variance. We also deleted a number of items that would not help generate testable hypotheses about predictors of unethical behavior, such as demographic questions, membership to various groups, and confidence in various international organizations. The full list of items deleted and dummy-coded is available at https://doi.org/10.17605/OSF.IO/A6Y7R. We were left with 708 predictors after the second round of data cleaning.

Model building. We thus analyzed the training data using a fully connected deep-learning neural network. We experimented by varying the number of hidden layers in the mode from zero to seven. We observed sizable increases in accuracy up to three hidden layers but less than 1% increase in accuracy with additional layers; however, computing time increased substantially with more hidden layers. We thus selected a model with three hidden layers. We first standardized responses to all items to range from 0 to 1. The model training was performed on graphics cards instead of the CPU to reduce the model-building time. We used the *Keras* package in R (Chollet, 2015), which implemented a *multilayer perceptron*.

Figure 2 illustrates the model. Using an initial set of weights that we had seeded, the model predicted the

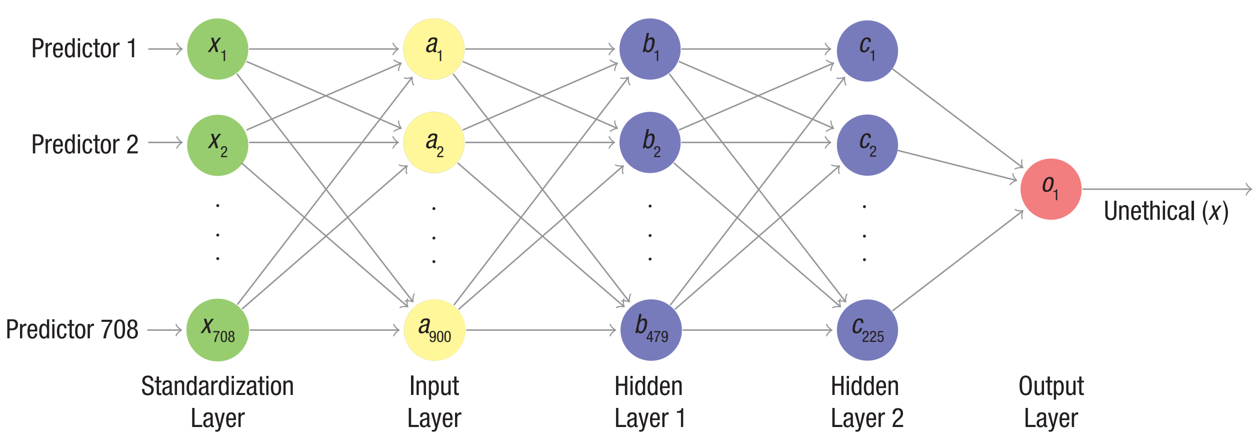

Fig. 2. Illustration of the deep-learning model used in Study 1. The green circles represent standardized versions of the 708 predictors. The yellow and purple circles represent perceptrons containing the relu activation function. The bottom row indicates the total number of perceptrons in each layer. The pink circle represents the perceptron in the final layer containing the sigmoid activation function, which predicted whether participants were ethical or unethical.

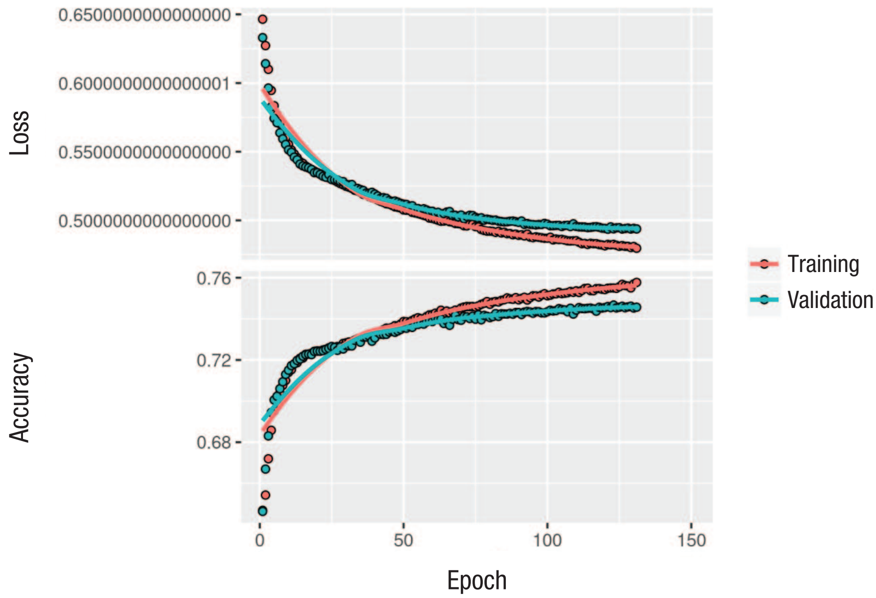

Fig. 3. Accuracy and binary-cross-entropy loss in the model-building (training) data and the validation data across successive iterations (Study 1).

outcome variable (ethical or unethical) for each respondent and then computed the loss (i.e., the gap between the actual values and the predicted values across the entire data set) using the *binary-cross-entropy* loss function. The model then adjusted the weights and recomputed the loss. This procedure continued until either a maximum of 200 iterations were run or the loss did not reduce in 10 consecutive iterations. Figure 3 depicts the loss across all the iterations of our model.

We used the *leave-p-out* (LPO) cross-validation technique, which allowed us to assess the generalizability of the model (Celisse, 2014). In each iteration of the model, the LPO technique randomly split the 90% training data into two components: 70% of the data were used for model-building and 20% for validation. In each iteration, the deep-learning algorithm built a model using just the model-building data and then tested the performance of this model on the validation data (in terms of accuracy and the binary-cross-entropy loss). The magnitude of the loss in the validation data indicates the model fit. For the model-building data, the accuracy and loss typically reach an asymptote toward 1 and 0, respectively. For the validation data, once the model begins converging, the loss typically reduces and the accuracy increases. But after some point, the model tries to overfit the model-building data (Hansel, Mato, & Meunier, 1992) and then the loss begins to increase while the accuracy begins to decrease in the validation data. To discourage the model from overfitting, in each iteration, we dropped a proportion of all neural connections (Hinton, Srivastava, Krizhevsky, Sutskever, & Salakhutdinov, 2012).

Hyperparameter search. To identify successively better models (i.e., models with a smaller loss), we used the hyperparameter search procedure, which adjusts a number of model parameters over several iterations. Specifically, we varied the number of perceptrons in each layer, the proportion of neural connections that are dropped after every iteration in each layer, the batch size (i.e., the number of rows that the model processes in one step of one iteration), and the learning rate (see Table 1). The hyperparameter search procedure experimented with 1,000 different random combinations of the 10 parameters. Of these 1,000 models, we chose the parameter combination that generated the smallest binary-cross-entropy loss. Table 2 presents the parameters used in our final model (see Fig. 2 for a depiction of the perceptrons in the final model).

Because this model was run on four graphics cards rather than on the CPU, and because each of our graphics cards contained 1,920 mini-CPUs, the computations within the individual perceptrons ran in parallel. This parallel processing introduced an indeterminacy because the order in which all the perceptrons of the model are computed is not guaranteed. Thus, if the same model is run repeatedly with identical parameters, the results of the model may vary on the scale of 0.1%. This minor variation is acceptable given that the time needed to build the model on graphics cards is approximately 200-fold faster than doing so on the CPU.

Model selection. Figure 3 depicts the accuracy and the binary-cross-entropy loss of the final model's classification in the model-building data and the validation data

Table 1. Parameters Values Used in the Hyperparameter Search (Study 1)

Parameter	Value
Number of neurons in first layer	100–1,000
Number of neurons in second layer	100–1,000
Number of neurons in third layer	100–500
Number of neurons in fourth layer	10–50
Dropped connections in each layer	0.1–0.8
Batch size	65, 128, 256, 512
Learning rate	$[200–1,500] \times 10^{-6}$

across successive iterations. As expected, in the training data, accuracy increased monotonically and loss decreased monotonically with successive iterations. However, with the validation data, accuracy increased and loss decreased up to a certain point and thereafter plateaued. *Keras* automatically saved the model that yielded the minimum loss in the validation data. The results reported are from this model.

Holdout testing. Once the model was finalized, we analyzed how accurately it classified individuals in the unseen data as ethical or unethical.

Alternate model. We also tried analyzing the data using a random-forest model, which ran on the CPU rather than on graphics cards and was much easier and simpler to run than deep-learning models. However, the random-forest model did not complete a single iteration on our computer within 10 days. We thus abandoned this approach.

Table 2. Parameters of the Final Model (Study 1)

Parameter	Value
Neurons in first layer	900
Neurons in second layer	479
Neurons in third layer	225
Neurons in fourth layer	46
Dropped-connection rate for first layer	0.2101
Dropped-connection rate for second layer	0.1660
Dropped-connection rate for third layer	0.6732
Dropped-connection rate for fourth layer	0.1455
Batch size	64
Learning rate	460×10^{-6}
Kernel initializer for first three layers	All 1s
Kernel initializer for output layer	All 0s
Activation function in first three layers	Relu
Activation function in output layer	Sigmoid
Optimizer	Adam
Learning-rate patience	.50
Early-stopping patience	20

Results

Reliability of the outcome measure. We had classified WVS respondents as either ethical or unethical on the basis of their responses to four items administered in all waves of the survey. To assess the reliability of this measure, we administered these four items to 204 U.S. residents recruited from Amazon Mechanical Turk (MTurk; age: M = 41.45 years, SD = 12.04; 87 women, 56 men, 61 missing values). Participants responded to the same four items used in the WVS (on the same response scale) twice, with an approximately 2-month gap in between the two measures.

Each time participants took the survey, we classified them as either ethical or unethical using the same method that we used to classify the WVS respondents. We found that 81.86% of the MTurk participants received the same classification on both occasions (63.49% for individuals classified as ethical at Time 1; 90.07% for individuals classified as unethical at Time 1). Thus, 81.86% (95% confidence interval, or CI = [75.88%, 86.90%]) represents the upper bound for the deep-learning model's accuracy.

Predictability of unethical behavior. Overall, in the unseen data (i.e., data that were not used in any way to build the model), the deep-learning model accurately classified respondents as ethical or unethical in 73.7% of the cases (95% CI = [73.2%, 74.2%]). Given that the dependent measure's test-retest reliability, 81.86%, imposed a theoretical upper bound on the deep-learning model's accuracy, our model's accuracy was 90% of the theoretical maximum accuracy. The overall accuracy of the model was above chance level, κ = 38.6%. The model's *specificity* (i.e., its accuracy in classifying ethical individuals) was 50.1%, whereas the model's *sensitivity* or *recall* (i.e., its accuracy in classifying unethical individuals) was 86.3%. This pattern is congruent with the results from the test-retest reliability study reported above, in which individuals classified as ethical at Time 1 were classified as ethical at Time 2 in only 63% of the cases, but individuals classified as unethical at Time 1 were classified as unethical at Time 2 in 90% of the cases. Table 3 presents the confusion matrix of the predictions of the deep-learning model using the unseen data.

The model's area under the receiver-operating-characteristic curve was 79.6%, indicating that if presented with a randomly selected ethical individual and a randomly selected unethical individual, the model would rank the unethical individual as unethical with this probability. The model's precision was 76.5%, indicating that if the model predicted that a person is unethical, its prediction was accurate in this proportion of the cases. The model's F_1 statistic, the harmonic mean

Table 3. Number of Actual Versus Predicted Cases in the Unseen Data (Study 1)

	Actual value	
Predicted value	Ethical participants	Unethical participants
Ethical participants	5,848	3,008
Unethical participants	5,817	18,901

Note: Accuracy was 73.7% (κ = 38.6%).

of precision and recall, which is often used to indicate the accuracy of a binary classifier with unbalanced data (i.e., data in which a binary outcome variable's responses are not distributed 50–50), was 81.1%.

Predictors of unethicality. Numerous methods exist to identify the top predictors in a deep-learning model, including *iml, LIME*, and *DALEX* (Molnar, 2020). The packages work by altering the values of one predictor at a time and assessing the extent to which this permutation influences the error term of the model; variables whose permutations cause a bigger change in the model's error term are assigned higher importance (Boehmke & Greenwell, 2019). Because all methods provide approximate solutions, the top predictors identified by one method might or might not match those identified by another method.

We used the *DALEX* package (Biecek, 2018) to identify the top predictors. This analysis was performed on the seen data because the model was built on this data set. The top 50 predictors of unethicality in the entire data set, in each wave of the WVS, in each subregion, and in each country are available at https://doi.org/10.17605/OSF.IO/A6Y7R. Specifically, the *DALEX* package estimated the increase in the model's binary-cross-entropy loss if each of the 708 variables in the data set was dropped from the model one at a time. The top 10 predictors in the entire data set are presented in Table 4.

For comparison, Table 5 presents the top 10 predictors with the highest point-biserial correlation with the dependent variable in the original WVS data set (before imputation). Quite strikingly, there was absolutely no overlap between the top 10 predictors identified by the deep-learning model and the top 10 predictors identified by a linear correlational analysis. Thus, the set of cause–effect hypotheses can be generated from a machine-learning analysis, and those that can be generated from a correlational analysis are nonoverlapping in the present case. This disjunction shows that predictor variables did not need to be highly correlated with the dependent variable to make a significant contribution to the deep-learning model's prediction accuracy.

We could not conduct a regression on the original WVS data because every respondent had many missing values. However, we ran a generalized linear model (GLM) with a logistic link function to predict unethicality in the imputed data based on all 708 variables. The model's accuracy was 72.4% (95% CI = [71.9%, 72.8%]), which was slightly lower than the lower deep-learning model's accuracy of 73.7%; however, the 95% CI of the GLM's accuracy was below the lower bound of the 95% CI of the deep-learning model's accuracy (95% CI = [73.2%, 74.2%]), indicating that the deep-learning model was significantly more accurate than the GLM. To identify the top predictors, we conducted a logistic regression with lasso regularization on the imputed data

Table 4. Top 10 Predictors of Unethicality Based on the Deep-Learning Model (Excluding Country, Wave, and Subregion Dummy Variables That Showed Up as Top Predictors; Study 1)

Item	WVS variable number	ΔDropout loss
How important: daily prayer	f194	0.4211
Religion is a cause of terrorism	f186	0.4180
Confidence: the presidency	e069_39	0.4175
Civil marriage is very important because it helps maintain the family	f193	0.4162
Aims of respondent: first choice (a stable economy)	e005 (1)	0.4161
Aims of respondent: first choice (the fight against crime)	e005 (4)	0.4153
Traits in a woman: woman wearing veil	d067	0.4152
Humanity has a bright or bleak future (bright future)	b017 (1)	0.4147
Meaning of religion: to follow religious norms and ceremonies vs. to do good to other people (both)	f200 (4)	0.4144
Aims of respondent: first choice (progress toward a less impersonal and more humane society)	e005 (2)	0.4142

Note: Values in parentheses after the World Values Survey (WVS) variable numbers refer to the response option of dummy-coded items. The "ΔDropout loss" column refers to the change in the model's binary-cross-entropy loss if the predictor mentioned in that row is permuted.

Table 5. Top 10 Nonexcluded Items With the Highest Bivariate Point-Biserial Correlation With the Dependent Variable in the Original WVS Data Set (Before Imputation; Study 1)

Item	WVS variable number	Point-biserial *r*
I see myself as someone who: tends to be lazy	a215	.1661
Democracy: criminals are severely punished	e231	.1597
Democracy: people choose their leaders in free elections	e226	.1593
Democracy: the economy is prospering	e230	.1575
Why people work: work most important in my life	c042b5	.1498
Millennial development goal: reduce child mortality	e244	.1473
Importance of democracy	e235	.1450
Democracy: people can change the laws in referendums	e232	.1403
Humiliating to receive money without having to work for it	c037	.1382
National goals: free speech	e010	.1376

Note: WVS = World Values Survey.

(Hastie et al., 2009). We first used a 10-fold cross-validation to determine the minimal lambda value using the cv.glmnet command in R, and we then used this value in the logistic lasso regression. The lasso regression's accuracy, 72.4% (95% CI = [71.9%, 72.8%]) was virtually identical to the regular logistic regression's. Table 6 presents the top 10 predictors from a lasso regression. Four predictors identified by the lasso regression were also identified by the deep-learning model, but the rest were different.

Importantly, the rank ordering of the top predictors of the deep-learning model is not guaranteed, because changes in the model parameters can lead to different rank orderings. This is particularly the case when different predictors have very similar dropout loss, as in the current results. Thus, it is up to researchers to select predictors that they find most interesting to pursue. Figure S1 in the Supplemental Material available online depicts the nature of the relationship between three of these predictors and the dependent variable.

Overall, the top predictor of unethicality identified by the machine-learning model (i.e., importance of daily prayer) is consistent with existing psychological theories claiming that more religious people are more prosocial and ethical (Norenzayan & Shariff, 2008). The next predictor, "religion is a cause of terrorism," possibly reflects an antireligion bias, but it is difficult to be confident about the underlying psychological construct without additional research. The third predictor, "confidence in the presidency," is probably difficult to experimentally manipulate in the present times, in which people's views about the presidency are highly polarized, particularly in the United States. The next predictor is a double-barreled item—it is unclear whether participants were responding to "civil marriage is important" or "civil marriage helps maintain the

Table 6. Top 10 Predictors From a Logistic Lasso Regression Conducted on the Imputed WVS Data (Study 1)

Item	WVS variable number	Coefficient
How important: daily prayer	f194	0.5672
Civil marriage is very important because it helps maintain the family	f193	0.4803
Religions limit democratic processes	f175	0.4524
Confidence: the presidency	e069_39	0.4404
Who should decide: international peacekeeping (nonprofit/nongovernmental organization)	e135 (5)	0.4264
Who should decide: human rights (commercial enterprise)	e139 (6)	0.4211
Good human relationships (other answer)	a169 (5)	0.3982
Religion is a cause of terrorism	f186	0.3625
Importation of goods (other answers)	e062 (3)	0.3148
Who should decide: refugees (nonprofit/nongovernmental organization)	e138 (5)	0.2584

Note: Values in parentheses after the World Values Survey (WVS) variable numbers refer to the response option of dummy-coded items. Coefficients are unstandardized.

family." The next two predictors (and also the 10th predictor) were from the same WVS item, in which participants were asked to choose the important option from "a stable economy," "progress toward a less impersonal and more humane society," "ideas count more than money," and "the fight against crime." These response options varied across multiple dimensions rather than a single dimension. Thus, it is difficult to identify the precise psychological construct underlying participants' responses. The next predictor tapped how important it is that women wear a veil; this item is culture-specific, because women in most cultures do not wear a veil, and thus we did not pursue this construct.

The next predictor was, "Humanity has a bright or bleak future" (variable b017; response options: "bright future," "bleak future," "both," "neither," and "other"; dummy-coded in the analysis). The deep-learning model identified the first option, "bright future," as the one most diagnostic of ethicality. We interpreted this variable to reflect optimism about the future of humanity. There is a large literature on unethical behavior and a reasonably large literature on optimism. Yet researchers have not connected the two. Because we had imputed a large volume of data, we first sought to verify whether there is a relationship between participants' responses to this item and their ethicality in the unimputed WVS data set. We found a significant direct relationship: 61.21% of the respondents who believed that humanity has a bright future justified at least one of the four unethical behaviors, but 65.00% of the participants who believed that humanity has a bleak future did so, $\chi^2(1) = 82.92$, $p < .001$.

Item b017 was probably not identified as a top predictor by bivariate correlations or the lasso regression because it was only weakly correlated with the dependent variable, $r = -.0383$. In contrast to regression-based methods, deep-learning models can capture nonlinear effects and complex interactions. In this case, though, because we had dummy-coded b017, the deep-learning model could map only a linear relationship between optimism about humanity and unethicality. Nevertheless, b017 probably contributed to the deep-learning model's predictions through interactions with other items included in the WVS. In any case, the deep-learning model did identify a direct relationship between optimism and unethicality (see Fig. S1). The correlation in the original WVS data further confirms that notwithstanding any interaction effects, a direct relationship does exist between optimism and ethicality. We thus decided to test the hypothesis that optimism reduces the justifiability of unethical behavior.

Although we focus on a hypothesis based on the worldwide top predictors from the deep-learning model, the model also generated region-specific and country-specific top predictors (see https://doi.org/10.17605/OSF.IO/A6Y7R). For example, a top predictor in Northern Africa was variable f175 ("Religions limit democratic processes"); variable e003, Response Option 3 ("Aims of respondent, first choice: fighting rising prices") in Eastern Asia; and variable e003, Response Option 2 ("Aims of respondent, first choice: give people more say") in Latin America. Future research can examine these region-specific hypotheses, thereby helping researchers expand their theorizing beyond ideas generated from Western, educated, industrialized, rich, and democratic (WEIRD) cultural contexts (Henrich, Heine, & Norenzayan, 2010). Future research can test such culture-specific hypotheses.

Study 2: Correlational Replication

The goal of Study 2 was to provide a conceptual replication of the key result found in Study 1 using different measures of the underlying constructs. Instead of the single-item measure of optimism included in the WVS, we used Scheier and Carver's (1985) optimism scale, which taps a general and global positive expectancy about the future. Further, instead of the WVS questions asking people whether unethical behaviors are justifiable, we administered Detert, Treviño, and Sweitzer's (2008) unethical-decision-making scale, because people who are higher on this scale are more likely to engage in actual unethical behaviors.

Method

In this study, we report all conditions, measures, and participants. The study was conducted in a single wave, and data were analyzed only after data collection was completed. We preregistered the methods and analyses of this study at https://osf.io/4v3sg/.

Participants. In a previous study using the same measures, we found an effect size in the predicted direction, $r = -.172$. A power analysis with this correlation, an α of .05 (one-tailed), and power equal to 80% indicated that we needed to recruit 205 participants. A survey seeking 205 U.S. residents was posted on MTurk. In response, 218 participants completed the survey (age: $M = 42.18$ years, $SD = 13.61$, 7 missing values; 93 women, 117 men, 8 missing values). All participants had unique Internet protocol (IP) addresses.

Procedure. We measured people's dispositional optimism using the eight-item scale developed by Scheier and Carver (1985). Participants were asked to respond to sample items such as "I always look on the bright side of things" on a 7-point scale ranging from *strongly disagree* to *strongly agree* ($\alpha = .89$). We measured people's willingness to engage in unethical behaviors using the eight-item unethical-decision-making scale developed by Detert et al. (2008, Appendix B). Participants were presented with eight ethically charged scenarios and were asked to rate how likely they would engage in the unethical behaviors described in these scenarios using a 7-point scale ranging from *not at all likely* to *extremely likely* (α = .88). A sample scenario is "You work as an office assistant for a department in a large company. You're alone in the office making copies and realize you're out of copy paper at home. You therefore slip a ream of paper into your backpack." Thereafter, we asked participants to respond to an open-ended statement: "Please summarize the main point of the statements that you just responded to in this survey."

Results

Following our preregistered analysis plan, we excluded 30 participants from our final analyses because they provided gibberish or irrelevant responses to the open-ended statement asking them to summarize the main point of the two measures that they responded to (see the Supplemental Material for the responses that were judged to be gibberish).

We found that dispositional optimism was negatively related to people's willingness to engage in unethical behaviors, $r(186) = -.213$, 95% CI = [−.351, −.070], p = .003. Therefore, this correlational study provides support for the hypothesis generated by the machine-learning analysis: More optimistic people are less willing to engage in unethical behaviors. Importantly, we replicated the key finding from Study 1 using measures that differed from those included in the WVS.

Study 3: Experiment

Study 3 tested the hypothesis that experimentally manipulating optimism would reduce people's tendency to perceive unethical behaviors as justifiable. We did so in the context of the COVID-19 epidemic. Two experiments reported in the Supplemental Material confirmed our prediction but with a slightly broader measure of unethical behavior in the context of COVID-19. In Experiment 3, we sought to replicate these findings with a more precisely defined measure of unethical behavior.

Method

As in Study 2, we report all conditions, measures, and participants. The study was conducted in a single wave, and data were analyzed only after data collection was completed. We preregistered the method and analysis plan for this study at https://osf.io/hwu9x/.

Participants. In the two prior experiments (reported in the Supplemental Material), we found an average effect size (Cohen's d = 0.36). A power analysis based on this effect size, an α of .05 (one-tailed), and power equal to 90%, indicated that we needed to recruit 266 participants. In the two previous studies, 8% of the participants, on average, were excluded because they provided gibberish or irrelevant responses to an open-ended statement. We thus posted a survey seeking 289 (i.e., 266/(1%–8%)) U.S. residents on MTurk. In response, 294 participants completed the survey (age: M = 34.87 years, SD = 11.93; 157 women, 135 men, 1 other, and 1 missing value). All responses came from unique IP addresses. Each participant was randomly assigned to either the *pessimism* condition or the *optimism* condition.

Procedure. Each participant was presented with a scenario stating that the future of the COVID-19 epidemic was either bright or bleak (see the Supplemental Material for detailed scenarios). For example, participants in the optimism condition were told that the virus should be contained within two months, that a vaccine should be ready within 6 months, and that the rate of new infections is guaranteed to go down. In contrast, participants in the pessimism condition were told that it will be very difficult to contain the virus, that a vaccine is unlikely to be ready in time, and that the rate of new infections is still pretty high. After they read the scenario, participants were asked to summarize the main idea it expressed and to respond to a manipulation-check question: "What are your expectations for the future of the coronavirus situation?" Participants responded on a 7-point scale ranging from 1, *it will be a very bleak future*, to 7, *it will be a very bright future*.

We finally measured the dependent variable by asking participants to rate the extent to which they found five unethical behaviors related to the COVID-19 situation justifiable: (a) "It is OK to go to a park that is closed for some fresh air and exercise, especially because there are likely to be very few people in a closed park"; (b) "It is justifiable to hoard face masks and hand sanitizers given how hard it is to buy them anywhere"; (c) "We must buy as many groceries as possible and stock up because who knows when supermarkets will run out of food"; (d) "Despite all the social distancing guidelines, it is OK to shake hands when meeting someone because otherwise one would appear extremely rude"; and (e) "Despite the social distancing guidelines, it is OK to get together with a few friends for a drink, as long as everyone doesn't show any respiratory symptoms" (see the Supplemental Material for the full items). Participants were asked to respond on a 11-point scale ranging from −5, *definitely express my disagreement*, to 5, *definitely express my agreement*.

Results

Following our preregistered plan, we excluded 52 participants who provided gibberish or irrelevant responses to the open-ended statement asking them to summarize the main point of the information provided in the scenario (see the Supplemental Material for the responses that were judged to be gibberish).

An independent-samples t test revealed that participants in the optimism condition thought that the future of the coronavirus situation was more bright (M = 4.88, 95% CI = [4.65, 5.13], SD = 1.35) than those in the pessimism condition (M = 3.21, 95% CI = [2.96, 3.47], SD = 1.42), $t(240)$ = 9.40, p < .001, Cohen's d = 1.21, 95% CI = [1.04, 1.39]. This finding indicates that our experimental manipulation was successful.

The five-item measure of justifiability of unethical behaviors had high reliability, α = .80. Another independent-samples t test found that participants in the optimism condition were less likely to find unethical behaviors justifiable (M = −3.15, 95% CI = [−3.41, −2.87], SD = 1.55) than participants in the pessimism condition (M = −2.79, 95% CI = [−3.07, −2.49], SD = 1.72), $t(240)$ = 1.67, p = .048 (one-tailed, given the preregistered directional hypothesis), Cohen's d = 0.215, 90% CI = [0.002, 0.427]. This experiment thus provided causal support for the hypotheses generated by the machine-learning analysis: Increasing participants' optimism about the COVID-19 epidemic reduced the extent to which they justified unethical behaviors related to the epidemic.

General Discussion

In the current research, we used a deep-learning model to predict whether people perceive unethical behaviors as justifiable and to generate novel hypotheses about antecedents of perceived justifiability of unethical behaviors. The deep neural network that we built was able to classify respondents of the WVS as ethical or unethical with high accuracy—the model's accuracy was 90% of the test-retest accuracy of the measure of unethicality.

Notably, the top 10 predictors of unethicality identified by the deep-learning model did not overlap at all with the top 10 predictors identified by a correlational analysis and partially overlapped with those identified by a lasso regression. This probably occurred because the deep-learning model could have modeled any number of interactions and nonlinear effects. Researchers have recently made advances in uncovering interactions from deep-learning models (Tsang, Cheng, & Liu, 2018). In the current research, we focused on the main effect of optimism on unethical behavior because we reasoned that if a top predictor does not directly cause the outcome variable, its interactions with other variables on the outcome effects would be of limited theoretical and practical utility. Nevertheless, given that optimism was not identified as a top predictor by the lasso regression and by bivariate correlations, the bulk of its contributions to predicting unethicality in the deep-learning model likely occurred through interactions. This means that in other data sets, if the unknown variables that the optimism variable has been interacting with are not included, then the predictive value of optimism might be lower.

We formulated a novel hypothesis—that optimism reduces unethicality—on the basis of the deep-learning model's finding that whether people think that the future of humanity is bleak or bright is a strong predictor of unethicality. This variable was not flagged as a top predictor either by the correlational analysis or by the lasso regression. Consistent with this idea, the results of a correlational study showed that people higher on dispositional optimism were less willing to engage in unethical behaviors. A following experiment found that increasing participants' optimism about the COVID-19 epidemic reduced the extent to which they justified unethical behaviors related to the epidemic. The behavioral studies were conducted with U.S. American participants; thus, the cultural generalizability of the present findings is unclear. Future research needs to test whether optimism reduces unethical behavior in other cultural contexts.

Although we could not locate research documenting the link between optimism and unethical behavior, we found research on mood and unethical behavior. Whereas Vincent, Emich, and Goncalo (2013) report that people in a more positive mood are more likely to engage in unethical behaviors, a meta-analysis by Kong and Drew (2016) suggests the opposite, namely, that people in a more positive mood are more likely to disengage from unethical behaviors. Optimism and mood are related in that more optimistic people tend to have more positive mood (Marshall, Wortman, Kusulas, Hervig, & Vickers, 1992; Segerstrom, Taylor, Kemeny, & Fahey, 1998), and inducing positive mood in people increases their optimism (Salovey & Birnbaum, 1989). Thus, past research on mood does not lead to clear predictions about the relationship between optimism and unethical behavior. Machine learning aided generation of the current directional hypothesis.

A key shortcoming of the machine-learning procedure to generate hypotheses is that machine-learning models give us ideas for a novel cause–effect relationship but do not provide a theory or a mechanism explaining this relationship. It is up to us researchers to theorize about and identify mechanisms underlying the cause–effect relationship. In the present context, we provide a conjecture for why optimistic people might be less likely to engage in unethical behaviors: People typically engage in unethical behaviors because they want to obtain some positive outcomes that they think cannot be obtained otherwise, and by definition, more optimistic people believe that they are more likely to obtain positive outcomes in the future. Thus, optimism might obviate the need to engage in unethical behaviors to obtain a positive outcome because people think that they already have a higher chance of obtaining the outcome. Future research can test this and other explanations as to why optimism reduces the justifiability of unethical behavior.

In the context of the COVID-19 pandemic, our findings suggest that if we want people to act in an ethical manner (e.g., to not hoard, to follow social-distancing guidelines), we should give people reasons to be optimistic about the future of the epidemic. For example, the media and governments can emphasize (a) that with sufficient measures, it is possible to contain the epidemic, as we know from China's experience; (b) reductions in the rate of new infections and death, even while total infections and death are increasing; and (c) that initial trials of numerous vaccines have been successful. One limitation of the current research is that we tested the hypothesis generated only with U.S. residents. Future research can test whether optimism reduces unethical behaviors in other parts of the world.

The current research presents significant advances in the use of machine-learning techniques in psychology. Past research at this intersection has primarily focused on prediction (Bleidorn & Hopwood, 2019), and researchers have conceptualized machine-learning models' ability to predict behavior as an alternative approach to traditional research's ability to explain behavior (Yarkoni & Westfall, 2017). However, in the current research, we used machine-learning tools simultaneously for predicting behavior (with 90% accuracy) and explaining behavior (identifying a novel cause–effect relationship). Thus, the current research demonstrates that using machine-learning algorithms, we can achieve relatively high predictability while also generating novel theoretical insights that are borne out by experimental tests.

Transparency

Action Editor: Marc J. Buehner
Editor: Patricia J. Bauer
Author Contributions
A. Sheetal conceived the idea for the study and designed and performed the machine-learning research. Z. Feng and K. Savani designed the behavioral research. Z. Feng collected and analyzed the behavioral data. All authors wrote the manuscript and approved the final version for submission. A. Sheetal and Z. Feng contributed equally to this study.
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
This research was supported by a Nanyang Assistant Professorship grant awarded by Nanyang Technological University to K. Savani.
Open Practices
Data for Studies 1 through 3, modeling code for Study 1, materials for Studies 2 and 3, and information about Studies S1 and S2 have been made publicly available via OSF and can be accessed at https://doi.org/10.17605/OSF.IO/A6Y7R. The design and analysis plan for Studies 2 and 3 were preregistered at https://osf.io/4v3sg/ and https://osf.io/hwu9x/, respectively. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797620959594. 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 iDs

Abhishek Sheetal https://orcid.org/0000-0002-4585-5358
Krishna Savani https://orcid.org/0000-0002-6934-1917

Acknowledgments

We thank Ee Hwee Lau, Dayana Bulchand, Sylvia Chin, and Andrea Low for invaluable research assistance.

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797620959594

References

Alpaydin, E. (2020). *Introduction to machine learning*. Cambridge, MA: MIT Press.

Baldi, P., Sadowski, P., & Whiteson, D. (2014). Searching for exotic particles in high-energy physics with deep learning. *Nature Communications*, *5*, 4308–4316.

Biecek, P. (2018). DALEX: Explainers for complex predictive models in R. *The Journal of Machine Learning Research*, *19*, 3245–3249.

Bleidorn, W., & Hopwood, C. J. (2019). Using machine learning to advance personality assessment and theory. *Personality and Social Psychology Review*, *23*, 190–203.

Boehmke, B., & Greenwell, B. M. (2019). *Hands-on machine learning with R*. Boca Raton, FL: CRC Press.

Celisse, A. (2014). Optimal cross-validation in density estimation with the L^2-loss. *The Annals of Statistics*, *42*, 1879–1910.

Chollet, F. (2015). Keras [Computer software]. Retrieved from https://github.com/keras-team

Cullen, J. B., Parboteeah, K. P., & Hoegl, M. (2004). Cross-national differences in managers' willingness to justify ethically suspect behaviors: A test of institutional anomie theory. *Academy of Management Journal*, *47*, 411–421.

Deng, Y., Chang, C., Ido, M. S., & Long, Q. (2016). Multiple imputation for general missing data patterns in the presence of high-dimensional data. *Scientific Reports*, *6*, Article 21689. doi:10.1038/srep21689

Detert, J. R., Treviño, L. K., & Sweitzer, V. L. (2008). Moral disengagement in ethical decision making: A study of antecedents and outcomes. *Journal of Applied Psychology*, *93*, 374–391.

Dong, B., & Torgler, B. (2009). Corruption and political interest: Empirical evidence at the micro level. *Journal of Interdisciplinary Economics*, *21*, 295–325.

Ellemers, N., van der Toorn, J., Paunov, Y., & van Leeuwen, T. (2019). The psychology of morality: A review and analysis of empirical studies published from 1940 through 2017. *Personality and Social Psychology Review*, *23*, 332–336.

Gerlach, P., Teodorescu, K., & Hertwig, R. (2019). The truth about lies: A meta-analysis on dishonest behavior. *Psychological Bulletin*, *145*, 1–44.

Goldstone, R. L., & Lupyan, G. (2016). Discovering psychological principles by mining naturally occurring data sets. *Topics in Cognitive Science*, *8*, 548–568.

Hansel, D., Mato, G., & Meunier, C. (1992). Memorization without generalization in a multilayered neural network. *EPL (Europhysics Letters)*, *20*, 471–476.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). New York, NY: Springer Science & Business Media.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv*. Retrieved from http://arxiv.org/abs/1207.0580

Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., . . . Puranen, B. (Eds.). (2014). World Values Survey: All rounds - country-pooled datafile. Retrieved from https://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp

Jha, D., Ward, L., Paul, A., Liao, W.-K., Choudhary, A., Wolverton, C., & Agrawal, A. (2018). *ElemNet*: Deep learning the chemistry of materials from only elemental composition. *Scientific Reports*, *8*, Article 17593. doi:10.1038/s41598-018-35934-y

Karp, R. M. (1975). On the computational complexity of combinatorial problems. *Networks*, *5*, 45–68.

Kong, D. T., & Drew, S. (2016). Meta-analyzing the differential effects of emotions on disengagement from unethical behavior: An asymmetric self-regulation model. In D. T. Kong & D. R. Forsyth (Eds.), *Leading through conflict: Into the fray* (pp. 23–44). New York, NY: Palgrave Macmillan.

Marshall, G. N., Wortman, C. B., Kusulas, J. W., Hervig, L. K., & Vickers, R. R., Jr. (1992). Distinguishing optimism from pessimism: Relations to fundamental dimensions of mood and personality. *Journal of Personality and Social Psychology*, *62*, 1067–1074.

Martin, N. D., Rigoni, D., & Vohs, K. D. (2017). Free will beliefs predict attitudes toward unethical behavior and criminal punishment. *Proceedings of the National Academy of Sciences, USA*, *114*, 7325–7330.

Molnar, C. (2020). *Interpretable machine learning: A guide for making black box models explainable*. Retrieved from https://christophm.github.io/interpretable-ml-book/

Norenzayan, A., & Shariff, A. F. (2008). The origin and evolution of religious prosociality. *Science*, *322*, 58–62.

Reyzin, L. (2019). Unprovability comes to machine learning. *Nature*, *565*, 166–167.

Salovey, P., & Birnbaum, D. (1989). Influence of mood on health-relevant cognitions. *Journal of Personality and Social Psychology*, *57*, 539–551.

Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology*, *4*, 219–247.

Segerstrom, S. C., Taylor, S. E., Kemeny, M. E., & Fahey, J. L. (1998). Optimism is associated with mood, coping, and immune change in response to stress. *Journal of Personality and Social Psychology*, *74*, 1646–1655.

Simha, A., & Parboteeah, K. P. (2019). The Big 5 personality traits and willingness to justify unethical behavior—A cross-national examination. *Journal of Business Ethics*. Advance online publication. doi:10.1007/s10551-019-04142-7

Sommer, U., Bloom, P. B. N., & Arikan, G. (2013). Does faith limit immorality? The politics of religion and corruption. *Democratization*, *20*, 287–309.

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—nonparametric missing value imputation for mixed-type data. *Bioinformatics*, *28*, 112–118.

Stokes, J. M., Yang, K., Swanson, K., Jin, W., Cubillos-Ruiz, A., Donghia, N. M., & Tran, V. M. (2020). A deep learning approach to antibiotic discovery. *Cell*, *180*, 688–702.

Tsang, M., Cheng, D., & Liu, Y. (2018). Detecting statistical interactions from neural network weights. *arXiv*. Retrieved from https://arxiv.org/abs/1705.04977

Vincent, L. C., Emich, K. J., & Goncalo, J. A. (2013). Stretching the moral gray zone: Positive affect, moral disengagement, and dishonesty. *Psychological Science*, *24*, 595–599.

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*, 1100–1122.