10.5465/AMBPP.2022.322

# USING MACHINE LEARNING TO ADVANCE THEORY: PRIDE IN WORK AS A NOVEL ANTECEDENT OF JOB SATISFACTION

ABHISHEK SHEETAL
Central Queensland University and The Hong Kong Polytechnic University

YI HUANG
Nanyang Technological University

KRISHNA SAVANI
The Hong Kong Polytechnic University

## INTRODUCTION

Job satisfaction is one of the oldest and most widely researched topics in organizational behavior and industrial-organizational psychology (Judge & Kammeyer-Mueller, 2012). Given the importance of employees' job satisfaction, there has been ample research investigating the determinants of this construct. Although there is a large volume of research on job satisfaction, we asked whether researchers might have missed out on some important antecedents. Management researchers typically generate hypotheses from existing theories. However, if existing theories are limited, then researchers would be unable to discover important antecedents of a phenomenon that are not derived from the theory. This problem is not restricted to the management sciences; natural scientists have claimed that "human activities [are] a principal bottleneck in scientific progress" and make "scientific advancement more subject to error and harder to reproduce" (Gil et al., 2014, p. 171).

To address the human bottleneck in scientific progress, natural science researchers have suggested that artificial intelligence (AI) can serve as an aid in the process of hypothesis generation and discovery (Mjolsness & DeCoste, 2001). Relatedly, recently, psychologists have used a similar approach to generate a novel hypothesis from a large dataset (Sheetal et al., 2020; see also Carey, 2020). Despite the increasing trend toward using machine learning methods to aid the discovery process in other scientific disciplines (Gil et al., 2014; Sheetal et al., 2020), to our best knowledge, no research in the organizational sciences has used a similar approach.

In this research, we used machine learning analyses to identify top predictors of job satisfaction in the World Values Survey (WVS) dataset (Inglehart et al., 2020). We trained two different machine learning models to predict respondents' job satisfaction, and then asked the models to reveal the top predictors of job satisfaction; we then chose the topmost predictor from this list, and tested it in a field study. In this way, we sought to place the new construct in existing theories of job satisfaction and use machine learning for theory building.

## STUDY 1: MACHINE LEARNING ANALYSES

### Method

We followed the approach of Sheetal et al. (2020). The code used to develop the model, the final model, and related files are available in the online repository for this project: https://osf.io/s6vcy/?view_only=a3177d6c9d574ebbb939a866e0fd7712

10.5465/AMBPP.2022.322

We performed the machine learning analyses on the World Values Survey dataset (Inglehart et al., 2014). Of the 24,558 individuals in the second wave of WVS, 13,840 has answered the job satisfaction question – their responses were thus used in the analyses. Respondents indicated their level of job satisfaction on a 10-point scale ranging from "1 = dissatisfied" to "10 = satisfied". The distribution of participants' responses to the job satisfaction question in the WVS was highly skewed, with 10 and 8 being the most commonly selected responses. Thus, a median split combining participants who selected a response between 1 and 7 vs. 8 and 10 seemed the most sensible approach.

Before imputation, we first excluded variables that were created by WVS researchers, variables which included more than 20 categorical responses, variables with open-ended responses, variables that indicated responses such as "none" or "don't know", and blank variables. Second, we dummy-coded all categorical variables, including s003 (country). In addition, we added dummy-coded variables that represented the ISO sub-region code and region code of each country. We used the *holdout technique* (Harrison et al., 1997) to test our model. We randomly split the dataset into two chunks: the *seen* data, which included 90% of the observations, was used for building the machine learning model, and the *unseen* data, which included remaining 10%, was used for testing the model.

We used the *missRanger* package in *R* (Wright & Ziegler, 2015) for the imputation using these parameters: 100 tree, 15 iterations maximum, splitrule = "extratrees," and respect.unordered.factors = TRUE. After imputing missing values, we standardized all input variables in the *unseen* data to range from 0 and 1. After imputing missing values in the *seen* data, we appended the *unseen* data to the post-imputation *seen* data, and imputed missing values in the *unseen* data. We then standardized values in the *seen* data.

Then, we excluded variables that cannot reasonably serve as antecedents of job satisfaction (e.g., whether respondents would be comfortable with different types of neighbor). After this round of data cleaning, we were left with 316 predictors.

For lasso regression, we used the *glmnet* to predict participants' job satisfaction in the *unseen* data. For deep learning model building, we used a fully connected neural network with three hidden layers (Aizenberg et al., 1999). We used graphics cards rather than CPU to perform the model training so that we can reduce the model-building time (Shi et al., 2016; Steinkraus et al., 2005). We used the *Keras* package in *R* to implement a *feedforward mutli-layer perceptron* (Pal & Mitra, 1992), and used Intel's *PlaidML* libraries (Sotoudeh et al., 2019) to conduct matrix multiplications on graphics cards. In each node in each layer before the output layer, we used the *Relu* activation function (Nair & Hinton, 2010). As we wanted to classify participants into two categories (*less satisfied or more satisfied with their jobs*), in the output layer, we used the *Sigmoid* (Kwan, 1992) activation function, which provides a binary output.

We employed the *leave-p-out cross-validation* method to ensure that our model yields generalizable insights (Celisse & Robin, 2008). In each iteration of the model, we split the 90% *seen* data into two parts: 70% of the data was used for *training* the model, and 20% of the data was used for *validating* the model. In the first iteration, we assigned the initial weights for all links in the input layer and the two hidden layers to 1, and the last layer to 0. The model then predicted each respondent's job satisfaction in both the training data and the validation data, and used the *binary cross entropy* loss function (Shore & Gray, 1982) to compute the value of the loss function (i.e., the difference between the actual values of the dependent variable and the predicted values in the respective dataset). The model then back propagated the errors in the training data, adjusted the weights, and tested the new set of weights in both the training and the

10.5465/AMBPP.2022.322

validation data. We set a limit of maximum 200 iterations. Then model building part ended if the value of the loss in the validation data did not reduce appreciably over 10 successive iterations. To reduce the chances of over-fitting, we compelled the model to randomly forget a fraction of all weights in the model after each iteration (Hinton et al., 2012).

The *learning rate* of the model indicates how fast the weights of all connections change across successive iterations (Jacobs, 1988). We set the *learning rate patience* to half; i.e., the learning rate was reduced by half if the model did not improve after 10 iterations. We set the *early stopping patience* to 20; i.e., the model building phase automatically ended if the model did not improve after 20 iterations. We used the *adam* optimizing algorithm (Kingma & Ba, 2014) to adjust the weights across successive iterations of the model.

We initially varied a number of model parameters so that we can make educated guesses about the range of parameters within which the model has reasonably good performance. To systematize this procedure, we conducted a *hyperparameter search*, a process that varies a set of specified parameters within specified ranges across a large number of iterations. We chose to vary the total number of nodes in each layer of the neural network, the proportion of weights in each layer that the model forgets after each iteration, the batch size (i.e., how many participants' data the model processes at a time), and the rate of learning. We asked the program to try out 1000 different randomly selected combinations of these ten parameters. The program selected the model that had the smallest loss value. Once the model building process was completed, we asked the model to predict respondents' job satisfaction in the *unseen* data.

**Results**

The lasso regression's accuracy in classifying respondents' job satisfaction in the unseen data was 75.1%, 95 CI [72.8%, 77.4%]. The top 10 predictors according to the lasso regression are pride in work, freedom of decision-making in job, satisfaction with one's life, satisfaction with financial situation of household, satisfaction with home life, the importance of work in life, the importance of god in one's life, freedom of choice and control, degree of pride in one's nationality, and sharing of moral standards with partner as the top predictor.

The deep learning model's accuracy in classifying respondents' job satisfaction in the unseen data was 75.6% (95% CI [73.2%, 77.8%]). This accuracy was above chance, K = 51.1% (Ben-David, 2008). The model's *Area Under Receiver Operating Characteristics* (Creelman & Donaldson, 1968) (AUC ROC) was 83%, 95% CI [80.8%, 85.1%], indicating that if presented with two individuals, one with low and one with high job satisfaction, the model would rank the higher job satisfaction individual above the lower job satisfaction individual with this probability. The top 10 predictors according to the deep learning model are pride in work, freedom of decision-making in job, satisfaction with one's life, satisfaction with financial situation of household, satisfaction with home life, the importance of work in life, the importance of god in one's life, how business should be managed, opinion about scientific advances will help, equality above freedom as the top predictors.

## LITERATURE REVIEW AND THEORY DEVELOPMENT

Overall, most of the top predictors of job satisfaction identified by the machine learning models are aligned with the existing management literature (Judge et al., 2017). However, the topmost predictor identified by the two machine learning models—pride in work—appears to be

10.5465/AMBPP.2022.322

novel. Past research in management has predominantly examined the extent to which employees experienced *pride experienced at work* (e.g., Baer et al., 2015), but relatively silent on their *pride in work*. Thus, the construct here is work-directed pride, not state-level pride experienced at work.

Wollack et al.'s (1971) work values scale includes a subscale titled *pride in work* (a sample item is "a man should feel a sense of pride in his work"). However, this construct is different from the WVS item because it measures what individuals *should* feel toward their work, rather than what individuals *actually* feel about their work. Moreover, pride in work might appear similar to *work engagement*, which is defined as "a positive motivational state of vigor, dedication, and absorption" (Bakker et al., 2014: 389). Indeed, one item of the work engagement scale, "I am proud of the work that I do," captures this construct, but only as part of the *dedication* subscale. Similarly, pride in work might appear similar to *job involvement*, but job involvement is a cognitive construct, and is defined as "a cognitive belief state of psychological identification with one's job" (Brown & Leigh, 1996, p. 361). Finally, although the construct *affective commitment* encompasses an emotional element that is similar to pride in work (Allen & Meyer, 1990), this construct is directed to one's firm rather than one's work. Thus, pride in work has not been conceptualized as a construct in its own right but only as part of broader constructs that are simultaneously affective, cognitive, and experiential in nature.

The construct *pride in work* appears parallel to *meaning in work*, which has been extensively examined in the literature (Judge et al., 2017). Past research has found that meaning in work is an important antecedent of job satisfaction—people who find their work more meaningful tend to be satisfied with their job (Hackman & Oldham). Meaning in work and pride in work appear to be parallel constructs, with one cognitive and the other affective in nature: "Meaningfulness is a cognition which originates from the experiences in the task that s/he cares about" (Hackman & Oldham, 1976), whereas pride is an emotion that results from "the enhancement of our ego-identity by taking credit for a valued object or achievement" (Lazarus & Cohen-Charash, 2001, p. 55). How meaningful an employee finds their work is an evaluation of the work itself. In contrast, pride in work is related to both the employee's ego and the work as pride is an affirming appraisal about oneself. Thus, the two constructs are distinct.

Also, it is also important to distinguish *pride in work* from *pride at work*. The appraisal of pride involves an evaluation of one's self- or social esteem, along with an evaluation of the target to which esteem can be attributed (Conroy et al., 2017; Tracy & Robins, 2007). Pride in work is attributed to the work, whereas pride experienced at work can be attributed to the work (e.g., Magee, 2015), to the organization (e.g., Jones, 2010), to coworkers (e.g., Baer et al., 2015), and so on. Therefore, pride in work appears to be a unique construct in the management literature.

*Hypothesis 1: Pride in work will predict job satisfaction above and beyond meaningfulness of work and pride experienced at work.*

Given the parallel between meaning in work and pride in work, we consider whether pride in work might serve as a mechanism in the job characteristics model (Hackman & Oldham, 1976; Oldham & Hackman, 2010).

*Hypothesis 2: Skill variety, task identity, and task significance will predict pride in work.*

10.5465/AMBPP.2022.322

*Hypothesis 3: Pride in work will mediate the association between the three core job characteristics (i.e., skill variety, task identity, and task significance) and job satisfaction after controlling for the experienced meaningfulness of work and pride experienced at work.*

## STUDY 2: MULTI-WAVE SURVEY

### Method

We recruited full time employees from Amazon's Mechanical Turk. To reduce common method bias, we adopted a three-wave design. We measured core job characteristics, pride in work, pride experienced at work, meaningfulness of work, and job satisfaction.

### Results

Consistent with Hypothesis 1, pride in work predicted job satisfaction ($b$ = .21, $SE$ = .06, $p$ = .00) above and beyond meaningfulness of work and pride experienced at work. As expected, meaningfulness of work also predicted job satisfaction ($b$ = .44, $SE$ = .06, $p$ = .00). However, unexpectedly, pride experienced at work was negatively associated with job satisfaction ($b$ = -.26, $SE$ = .06, $p$ = .00). This finding shows that *pride in work* and *pride at work* are theoretically distinct constructs. Further, skill variety ($b$ = .27, $SE$ = .07, $p$ = .00, task identity ($b$ = .14, $SE$ = .07, $p$ = . 03), and task significance ($b$ = .12, $SE$ = .06, $p$ = .06) were all positively associated with pride in work, consistent with Hypothesis 2. The size of the coefficients associated with pride in work were similar to those associated with meaningfulness of work (skill variety: $b$ = .34, $SE$ = .07, $p$ = .00; task identity: $b$ = .14, $SE$ = .06, $p$ = .03; task significance: $b$ = .22, $SE$ = .06, $p$ = .00), and higher than those associated with pride at work (skill variety: $b$ = .01, $SE$ = .08, $p$ = .94; task identity: $b$ = -.20, $SE$ = .07, $p$ = .01; task significance: $b$ = -.06, $SE$ = .07, $p$ = .41). To test the mediation relationship stated in Hypothesis 3, we conducted bootstrapping regression-based path analyses using the SPSS PROCESS macro (Hayes, 2017; Preacher & Hayes, 2008). After controlling the influence of meaningfulness of work and pride experienced at work, there were significant indirect effects of skill variety (*95% CIs* [.02; .14]), task identity (*95% CIs* [.00; .08]), and task significance (*95% CIs* [.01; .11]) on job satisfaction through pride in work. Therefore, Hypothesis 3 was supported. All three indirect effects via meaningfulness of work were significant and were similar in size to those of pride in work (skill variety: *95% CIs* [.11; .31]; task identity: *95% CIs* [.02; .17]; task significance: *95% CIs* [.08; .22]). Only one indirect effect was significant via pride at work (task identity: *95% CIs* [.01; .10]).

## DISCUSSION

The current research used two machine learning models to predict whether people are satisfied with their job, and to generate novel hypotheses about the antecedents of job satisfaction. The machine learning models identified a number of predictors of job satisfaction. Some of those were already identified in the literature, such as life satisfaction (Judge & Watanabe, 1993). Others were relatively novel, such as pride in work. We integrated the findings of the machine learning models with the job satisfaction literature and argued that pride in work fits into the framework of job characteristics model as an affective parallel to meaningfulness of

10.5465/AMBPP.2022.322

work, which is cognitive construct. A three-wave survey found that pride in work mediated the association between job characteristics and job satisfaction over and beyond meaningfulness at work and pride experienced at work. Further, the size of the coefficients associated with pride in work was similar to those associated with meaning in work. Pride in work was distinct from pride experienced at work, which did not play a significant role in the model.

Our findings highlight that pride in work is an important but unexplored predictor of job satisfaction, and is one of the mechanisms through which skill variety, task identity, and task significance influence job satisfaction. Indeed, given the volume of research on meaning in work (Rosso et al., 2010), it is surprising that its affective counterpart—pride in work—is simply absent from the job attitudes literature. Without the aid of machine learning, it is unclear when, if ever, researchers would have uncovered this affective construct using traditional deductive methods of hypotheses generation.

In addition, we differentiated pride in work from pride experienced at work. Specifically, pride in work was positively associated with job-related outcomes, whereas pride experienced at work was negatively associated with job satisfaction. One possible explanation for the negative association is that employees feel proud about themselves think that they are better than or above their coworkers, their job, and their organization, (Conroy et al., 2017). If employees think that they are superior to others, they may perceive their job as less satisfying. Building on this finding, future research can pay greater attention to different forms of pride in the organizational context (Tracy & Robins, 2007).

This research presents a novel use of machine learning to generate a new hypothesis. Interpretable machine learning allows researchers to think beyond the parameters of existing theories (Sheetal et al., 2020). Although we highlighted the power of machine learning in generating novel insights, using machine learning is not an end goal—it is merely a step in the discovery process. This combination of machine learning plus a field study is the key feature that distinguishes our research from prior work using machine learning in management, which has typically not conducted follow-up studies to test the predictions of machine learning models.

Regarding practical implications, these findings suggest that organizations should not only attempt to demonstrate to employees that their work is meaningful, but also elevate employees' perception of pride in the work they do, which would exert positive influences on both employees and the organizations. In addition, since different forms of pride would lead to distinct consequences for employees' job satisfaction, organizations need to be cautious about making employees feel more pride in an unspecific manner.

## REFERENCES AVAILABLE FROM THE AUTHORS