

January 27, 2021

NOTE TO THE
REVIEWER OF:

OMB CLEARANCE 1220-0141
“Cognitive and Psychological Research”

FROM:

Robin Kaplan
Research Statistician
Office of Survey Methods Research

SUBJECT:

Submission of Materials for the *Survey of
Occupational Injuries and Illnesses (SOII)*
Suggested Coding study

Please accept the enclosed materials for approval under the OMB clearance package 1220-0141 “Cognitive and Psychological Research.” In accordance with our agreement with OMB, we are submitting a brief description of the study.

The total estimated respondent burden hours for this study is 633 hours.

If there are any questions regarding this project, please contact Robin Kaplan at 202-691-7383.

1. Introduction

The Office of Safety, Health and Working Conditions (OSHS) is considering providing human coders with computer-automated suggested codes in the Survey of Occupational Injuries and Illnesses (SOII). The computer generated codes have been shown to be just as accurate as trained human coders, and sometimes even more so (Measure, 2014). However, the question remains as to whether providing human coders with these auto-suggested codes would improve accuracy, reduce coder burden, and/or reduce costs. OSHS approached the Office of Survey Methods Research (OSMR) to investigate the effects of auto-suggested codes on human coder accuracy and productivity.

The Survey of Occupational Injuries and Illnesses

The Survey of Occupational Injuries and Illnesses (SOII) is a mandatory establishment survey that generates annual estimates of the number and rate of nonfatal workplace injuries and illnesses. Businesses report on injuries and illnesses that occurred at their establishment by writing short written narratives in response to prompts that ask respondents to describe the injury or illness and what caused it.

SOII coders in BLS state and regional offices then manually code these narratives according to The Occupational Injury and Illness Classification System (OIICS), a classification system for use in coding the case characteristics of injuries and illnesses in SOII. OIICS includes four major case characteristics: Nature of the injury (e.g., fracture); Part of body affected (e.g., arm); Event which caused the injury (e.g., vehicular collision); and Source of the injury (e.g., the vehicle the person was in). In addition to the OIICS codes, coders must also use the narratives to determine the best occupation for the injured employee according to the Standard Occupational Classification (SOC) system used by SOII and other agencies to classify employees into occupational categories.

Coding SOII data is a complex task with thousands of available codes to select from. Coding rules and best practices can be difficult to understand and implement accurately. Human coding accuracy is low, ranging from 54% to 82% (Measure, 2014). Furthermore, SOII respondents'

narratives often lack sufficient information. Coders often have to call respondents back to obtain additional information, adding time and burden for both coders and respondents.

Research Questions. The overarching goal of this study is to test the effects of computer-generated code suggestions on coding data quality and coder productivity. In this study, participants will read fictional SOII narratives. Based on the information provided in the narratives, participants will attempt to select (or code) the best possible job titles for employees based a set of simplified rules designed to mimic the process of SOC coding. Some participants will see computer-suggested codes during the task and others will not see any (a control group). This design will allow us to identify ways in which the presentation of computer-suggested codes may affect code selections. In addition, we have identified six factors (framing of the codes, probability information, timing of code presentation, whether the codes are congruent with expectations, the amount of detail in the narratives, and the ambiguity of the possible options) that may influence the effect of the computer-generated code suggestions. These are described in detail below.

1.) How does the ‘framing’ of the suggested codes affect coder behavior?

As a necessary part of providing the suggested codes to the coders, OSHS must decide how to visually display the codes and how to instruct the coders to use the information. This ‘framing’ may influence how coders perceive the suggestions and integrate the computer-suggested codes in their decision making process; even providing no explanation at all would convey some information to the coders. For example, framing the suggestions as coming from a computer algorithm proven to be highly accurate may create a different impression compared to framing the suggestions as coming from a computer program that is in the middle of development (both of these may be argued to be fair characterizations of the OSHS algorithm). We’ve identified the following specific research aims to assess the framing of the suggested codes:

- When people are told the computer suggestions are highly accurate, are they more likely to uncritically accept the suggested codes?
- When people are told the computer is prone to error, are they more likely to actually use the suggested codes during the coding process to reduce burden, check their work, and assess the computer suggestion more critically?

2.) How does information about the quality of the suggested codes affect coder behavior?

The OSHS algorithm generates suggested codes along with probabilities for each suggested code being the correct code. This information is meaningful but may be difficult for coders to interpret appropriately and use effectively. This information can be presented in at least three ways: probabilities, ranks, or not at all. Probability information indicates the probability of the suggested code being accurate (e.g., 60% probability of being correct for one code versus a 39% probability of being correct for another code). Rank information indicates the computer's ranking in order from most to least likely to be correct. Use of probabilities, ranks, or providing no information about the probability that the code is correct may have different implications for how people process, judge, and ultimately select a code. For example:

- How do people assess auto-suggested codes that the computer indicates have a high probability of being correct? Are people more or less likely to select codes that have a high probability of being correct?
- Do people satisfice, or quickly select codes with a high probability of being correct, without displaying the cognitive effort to consider the others?
- How do people react to auto-suggested codes that have a low probability of being correct? Are these codes less likely to be selected? Are people more likely to dismiss these codes without critically assessing them?
- How do people use probability information when all probabilities are close in value? Are people more likely to guess in these instances, and display lower confidence in their selections?
- Are codes that are ranked #1 (as opposed to 2 or 3) more likely to be selected?
- If people disagree with the computer's suggested probability or rank, how do they reconcile that discrepancy? Do computer-suggested codes that appear inaccurate or illogical contaminate confidence in the accuracy of other suggested codes? (e.g., Dietvorst, Simmons, & Massey, 2014).

3.) Does the timing of computer-suggested codes affect data quality?

Suggested codes can be presented to coders at different times during the code-selection process. They can be presented after the coder has selected a code based on their own judgment, with the

opportunity to then decide whether to change their original selection based on the suggestions. Alternatively, codes can be presented before the coder selects a code, and thereby providing coders the opportunity to use the computer's suggestion to help inform their initial selection. The timing of these suggestions may seriously impact the coding decisions that coders make. For example:

- Are people more inclined to go with their original selection, despite what the computer suggests?
- Does presenting the suggested code before making a selection lead people to uncritically accept the computer suggested code quickly?
- Does presenting the code after making a selection cause people to doubt the accuracy of their original code selection, and instead accept the computer suggested code?

4.) What is the impact of the congruency of the computer suggested probabilities or ranks on coding decisions?

SOII coders read SOII narratives with a set of pre-existing knowledge, schemas, and assumptions about what workers in particular industries or occupations do in their work.

The computer-suggestions can be either congruent or incongruent with people's pre-existing knowledge. For instance, a narrative for an employee that works in the Fire Services industry and spends most of the time extinguishing fires would appear to have the job title 'Firefighter.' However, the algorithm might indicate that the probability that the job is 'Firefighter' is only 50%, or it could be ranked as the least likely job title. In both of these instances, the computer algorithm would be incongruent with expectations. In contrast, the computer algorithm might indicate that there is a 99% probability that the job title is 'Firefighter', or it could be the top-ranked selection (congruent with expectations). Whether the computer's suggested probabilities and ranks are congruent or incongruent with human expectations about a job title could seriously impact how much confidence coders hold in the algorithm's ability to identify the correct job title and ultimately which code they select. As such, we will explore these possibilities:

- When computer-suggested codes are congruent with expectations, will people make their selections more quickly compared to not having any information about the suggestions?
- When computer-suggested codes are incongruent with expectations, will people select the suggested code as frequently as they do when the code is congruent?

5.) Does the effect of suggested codes vary depending on the difficulty of the coding task?

Cases with less information from respondents are often more difficult to code. SOII narratives vary widely in the amount of detail they provide. Some provide enough information to make coding decisions without additional follow-ups, whereas others lack sufficient detail to make coding decisions. The amount of detail contained in the SOII narratives may seriously impact coding decisions and the use of suggested codes. For example:

- When narratives contain a lot of detail, are people less reliant on suggested codes and more likely to use their own judgment and knowledge?
- When narratives lack detail, are people more reliant on computer-suggested codes to make coding selections?

6.) Does the degree of similarity or dissimilarity of the set of code suggestions affect coding selections?

Cases for which there are many similar possible job titles are also more difficult to code. For instance, it's possible for each code in set of suggested codes to be highly similar to one another and equally plausible. Taking the example of the Firefighter job, other suggested codes might be Emergency Medical Technician (a closely related job, often paired with firefighter), a Fire Inspector, or an Emergency Responder. Each of these job titles are highly related to the job title 'Firefighter' and introduce some ambiguity with regard to which suggestion is most plausible. In contrast, sometimes the set of computer suggested codes are not highly similar to one another, and one stands out as an obvious choice. In the case of the Firefighter example, other suggested codes might include Police Officer, Lifeguard, or Medical Assistant. While somewhat related because these jobs all entail safety or emergency response, they aren't so similar to the job of Firefighter that they would be likely to create serious ambiguity in the task. Whether the set of computer-suggested codes create ambiguity with multiple plausible correct codes might affect the coding selections people make. For example:

- When the set of suggested codes are ambiguous, are people more likely to rely on the probability or rank information to form decisions?
- When the set of suggested codes are not ambiguous, are people more likely to rely on their own judgments and intuitions to make coding selections, and less on the computer?

- In circumstances where all codes look plausible, will people be more likely to select a code if it is labeled as #1 compared to when the probability information for that code is only slightly higher than for the other suggested codes?

2. Research Design

This research is part of a larger, multi-phased project aimed to assess the impact of auto-suggested codes on coding quality, accuracy, and productivity. The design of the proposed research was based in part on a previous study where OSMR conducted cognitive interviews with four SOII state coders to gain a better understanding of how they select codes in practice and use the coding interface. OSMR also received extensive input from various OSHS staff members who developed the computer algorithm to generate automated suggestions, staff who helped develop the original SOC coding system, staff that provide SOC coding training to SOII state coders, and other members of management.

Based on findings from the cognitive interviews and collaboration described above with OSHS, OSMR will conduct an online study using Amazon.com's Mechanical Turk. Ideally, we would explore this question with a large sample of SOII coders in a production coding environment to ensure findings are reliable and valid. However, this is not possible given the small population of SOII coders available to participate in a research study and the risk that any experience they have in a research study now may affect future implementations of computer-suggested codes (for example, a coder may dislike the version of the codes they see now and just never give them another chance (e.g., Dietvorst et al., 2014).

Given these limitations, we will use an alternative approach using Mechanical Turk, which is comprised of an online panel of "workers" who are typically given small incentives for completing survey tasks. In addition, Mechanical Turk has a group of what are known as "master" workers. These participants have demonstrated a high degree of accuracy in categorization tasks – similar to the work of SOII coders. Thus, these "master workers" provide an opportunity to test our research questions about the impact of suggested codes on human

coders. Mechanical Turk Master workers, though different from SOII coders in that they do not have the same expertise and training, are expected to exhibit similar basic human judgment and decision making biases. While the task of coding SOII narratives is unique, we believe that the cognitive processes involved in using suggested codes are similar to the basic cognitive processes involved in general coding tasks.

SOII coders do a specialized job, which requires extensive training to use very specific kinds of information. Based on our observations of coders, we believe that coders often select some codes immediately after reading the case narratives; this fluency may play a critical role in how coders use any computer-suggested codes. Observing the general population code SOII data according to the same rules given to SOII coders would not be informative – the general population would have to look up codes and re-read rules of selection. We propose using a task that allows the general population to code as fluently as SOII coders code SOII data, approximating the effect of computer-suggested codes on highly-trained coders. Because members of the general public are expected to have some familiarity with prototypical jobs found in the U.S. economy (e.g., firefighter, teacher, waiter/waitress), we focused our task on a simplified version of SOC coding for which we could quickly train members of the general public to perform. The results of this study will be used to answer some basic questions about computer-suggested codes. Importantly, it will help inform the design of subsequent research studies in other phases of the project where we will assess the impact auto-suggested codes on coding quality, accuracy, and productivity using actual SOII state coders.

In this research, Mechanical Turk participants perform a coding task that mimics that of SOII coders. We will create a simplified environment where participants must make classification judgments about an employee's job title (or SOC code) when provided fictional narratives that include the company name, industry, typical job duties, and the injury or illness. The "correct" job title (or SOC code) was based on data from a set of 'gold standard' (Measure, 2014) cases wherein OSHA staff with expertise in SOC coding determined the best possible SOC code for each from which the fictional SOII narratives were developed. Each participant will code 24 cases.

A summary of the experimental design is included in Appendix A. The proposed experimental design manipulates three factors between subjects:

- Framing: Participants will either be told that the computer’s suggested codes can be wrong sometimes or that the computer’s suggested codes are likely to be correct. This framing will be carried out in the task instructions as well as at the time of suggestion, by using two different labels for the suggested codes: “Most Likely Codes” or “Suggested Codes”.

Framing of “Computer can be wrong sometimes”	<p>The information from each Injury Case Sheet has been analyzed by the Occupation Classifier program and the results will be shown to you as “Suggested Codes”.</p> <p>These suggestions are being provided to help you make your classification. However, the computer program is still being tested and improved so it may make mistakes.</p>
Framing of “Check your work against the computer”	<p>The information from each Injury Case Sheet has been analyzed by the Occupation Classifier program and the results will be shown to you as “Most Likely Codes”.</p> <p>These suggestions are provided to help you make your classification. The computer program has been used many times and has been shown to be highly accurate.</p>

- Information about the quality of suggested codes: One-third of participants will be shown probability information alongside the suggested codes; one-third of participants will be shown rank information; and one-third of participants will not be shown any additional information about the suggested codes. The rank information for the suggested codes is fixed for each set; in other words, in any given set of suggested codes, there is a “best” code.

Display probabilities	Next to each suggestion is the computer program’s calculation of how likely the job title is to be the correct one.
Display ranks	The suggestions are presented in order so that the job title most likely to be the correct one is first.
Display no additional information	[...]

- Timing of suggestions: Participants will either receive the suggested codes at the time of their initial code selection or after they submit their initial code selection with the opportunity to change their answer after seeing the suggested codes.

Suggestions shown AFTER participant makes selection	These codes will be shown on the screen after you make your code selection. You'll then have the chance to change your answer if you wish.
Suggestions shown BEFORE participant makes selection	These codes will be shown on the screen before you make your code selection.

The experimental design also manipulates three factors within subjects:

- Level of narrative information provided: For the first half of the trials they are asked to code, participants will be given all the information needed to successfully code the case. For the second half of the trials, the job duties information will be omitted.
- Congruency of the quality of the suggested codes with expectations: For half of the trials they are asked to complete, the suggested codes will be presented congruent to expectations about that case; in other words, code suggestions that appear to be correct will also be labeled as being likely to be correct. For the other half of the trials, the suggested codes will be presented incongruent to expectations. Those participants who are not shown any information about the quality of the suggested codes will not see these labels (e.g., 85% or 1st) but the suggested codes will be presented in the same vertical order.
- Ambiguity of the suggested codes: Half of the trials will be presented with sets of suggested codes that are both equally plausible by content (e.g., firefighter, emergency medical technician, and fire inspector) and by probability (the averages of the randomly assigned probabilities will be: 36.5%, 30.5%, and 25.5%). The other half of trials will be presented with sets of suggested codes for which there is an evident correct choice, both by content (firefighter, police officer, and lifeguard) and by probability (averages will be: 81%, 11%, and 5%). Half of the narratives will be randomly assigned for each participant to be ambiguous; therefore, for each narrative there is both a set of equally plausible codes and a set of obvious codes. These codes are shown in Appendix B.

Additionally, there will be a control condition, in which participants do not see suggested codes. These participants will not receive any of the experimental manipulations described except that the level of narrative information provided will vary in the same way as for the other participants.

Analysis Plan

This research is exploratory and designed to help determine the impact of using computer-suggested codes on human coding selections. We have identified a number of conditions that may impact human coding decisions, but do not have specific hypotheses because very little research has been done on this topic and SOC coding is a very specialized task. As such, results may vary. On the one hand, literature on satisficing suggests people may accept the computer suggestions to arrive at an acceptable answer as quickly as possible and continue with the task (e.g., Krosnick, Narayan & Smith, 1996). On the other hand, people may distrust computer algorithms, even when shown to outperform human judgments, and people may want to rely on their own judgments instead (Dietvorst et al., 2014). Thus, we plan to conduct the following analyses to assess the impact of suggested codes on coding selections:

1.) Mixed model Analysis of Variance (ANOVA) with the following factors:

- 3 between-subjects factors:
 - 3 Framing Conditions (Computer can be wrong vs. Check your work against the computer vs. Control condition)
 - Nested analysis for the experimental conditions (excluding the control group):
 - 3 Probability Display Conditions (Probabilities vs. Ranks vs. No information)
 - 2 Timing Conditions (Before Code Submission vs. After Code Submission)
- 3 within-subjects factors:
 - 2 Narrative Conditions (Job duties vs. No job duties)

- o 2 Congruency Conditions (Congruent Suggested Probabilities vs. Incongruent Suggested Probabilities)
- o 2 Similarity of Suggested Codes Conditions (Easy vs. Ambiguous)

This analysis will allow us to assess the impact of coding under each of these conditions on the following dependent variables:

- Coding accuracy (what percentage of coding selections were correct or incorrect as compared to the gold standard?)
- Coding productivity (duration of time participants spend making each coding selection)
- Confidence (participants' rating of their confidence in each code selection)
- Level of difficulty (participants' rating of how easy or difficult it was to select a code for each case)

In our analysis, we will also use regression and Analysis of Covariance (ANCOVA) models to assess and/or control for potential covariates of the dependent variables, including the amount of prior experience the participant had with categorization tasks on Mechanical Turk, and their level of familiarity with the job titles and duties described in the narratives for each case.

Procedure

Participants will be introduced to the survey (Appendix C) and first complete training. Training consists of reading about coding rules (Appendix D) and completing two training cases where they are led through the code selection process (Appendix E). After coding each case, participants will answer three questions about that case, regarding their confidence in their code selection, the difficulty of choosing a code, and a report of whether they were not familiar with any of the terms used in the case (Appendix F). After completing all 24 cases (all narratives are included in Attachment 1; each narrative can be linked by its Case ID to the 'correct' code and

suggested codes in Appendix B), participants will be asked to answer individual differences and demographic questions (Appendix G).

3. Participants

Participants will be recruited using a convenience sample of 'Master Categorization Workers' from Amazon Mechanical Turk of adult U.S. citizens (18 years and older); this study is focused on internal validity rather than representativeness of any population. This research design requires a large sample of 1260 participants in order to sufficiently explore the range of variables of interest and because we expect a very small effect size since, as the study manipulations are subtle for online surveys of this nature. These participants will be randomly assigned to the 12 groups described (a 2x3x2 design with 100 participants per group and an additional control group with 60 participants). The control group is smaller because this group will not be part of the main study manipulations and thus requires fewer participants to make comparisons. This sample size also takes into account break-offs, incomplete data, and participants who do not follow the task instructions, and was based off a similar study by Dietvorst et al., (2014), as reflected in the power analysis in the next section below.

An additional 10 participants will be recruited for an initial pilot test from TryMyUI.com. TryMyUI is an online testing website where respondents can complete a set of self-administered tasks while thinking "out loud" and respond to follow-up, scripted probes. TryMyUI provides a video recording of the output and each test can last up to 20 minutes. These pilot participants will be asked to think aloud while completing an abridged version of the task (the same instructions, training, and debriefing but only 6 of the 24 cases) and answer questions about the experience afterwards, which will help to confirm whether the training is effective, the task is clear, the questions are worded clearly, and the experimental manipulations work as intended. The pilot tests will be conducted iteratively so that any modifications can be tested with pilot participants before launching the full study.

3a. Power Analysis

The primary goal of the proposed research is to explore the effects of computer-suggested codes on human coding performance. We found no previous research that directly assessed this question, in particular with a task as complex as SOC coding. Further, we expect a very small effect size, as this is an online study that asks participants to imagine different scenarios and lacks the realism of the task that actual SOII state coders perform. Online studies such as these require a large sample size to even detect very small effects, as reflected in the power analyses.

Sample size estimation

A statistical power analysis was performed for sample size estimation, based on data from a similar study by Dietvorst et al., (2014) ($N= 400$, with $n=100$ per condition), that assessed people's confidence in an algorithm that predicted the rank of U.S. states in terms of number of airline passengers that departed from that state in 2011. Mean confidence ratings were 3.40 on a 5-point scale for the control group, who were not exposed to the algorithm's estimates; mean confidence ratings when participants were 3.34 when participants were exposed to the algorithm's estimates.

With an $\alpha = .05$ and power = 0.80, the projected sample size needed with this effect size is approximately $N = 27$ for detecting the main effects of presenting participants with computer-suggested codes. Extrapolating across the additional between-subjects variables in this study that were not included in the Dietvorst et al., (2014), we would require a minimum of four times that amount for approximately $n = 108$ per experimental condition. Thus, our proposed sample size of $N = 1260$ is very close to the amount needed for the main objective of this study and should also allow for expected attrition and our additional objectives of exploring possible covariates related to coding selections. In addition, this sample size matches the sample used in Study 3b by Dietvorst et al., (2014).

4. Burden Hours

Our goal is to obtain responses from 1260 participants recruited from Amazon Mechanical Turk. Each session is expected to take no more than 30 minutes to complete, for a total of 630 burden hours. In addition, the 10 pilot participants are expected to take no more than 20 minutes each, for an additional 3 burden hours. Total burden hours are expected to total no more than 633

hours. The survey will be administered completely online at the time and location of the participant's choosing.

5. Payment to Respondents

We will recruit 1260 participants from the Amazon Mechanical Turk database. Participants will be compensated \$2.35 for participating in the study, a typical rate provided by Mechanical Turk for similar tasks. A total of \$2961.00 will be paid to respondents for their participation in the study. The study will be advertised with a base pay of \$1.85, with a potential to earn a \$0.50 bonus for accuracy in selecting suggested codes. Dietvorst et al., (2014) offered a similar bonus as an incentive to stay motivated and put forth effort in the task. In actuality, all participants will receive the \$0.50 bonus, regardless of performance.

The pilot participants recruited from TryMyUI will receive the standard TryMyUI fee of \$20 each for their participation, regardless of their performance on the task. The payment for these additional 10 pilot participants will total \$200.

6. Data Confidentiality

Recruiting of participants will be handled by TryMyUI.com (for the pre-testing phase) and Amazon Mechanical Turk (for the experimental phase). All participants will be informed that the study is about their perceptions of different types of questions. Once participants are recruited into the study, they will be sent a link to the survey, which is hosted by Qualtrics. The data collected as part of this study will be stored on Qualtrics servers. Using the language shown below, participants will be informed of the voluntary nature of the study and they will not be given a pledge of confidentiality.

This voluntary study is being collected by the Bureau of Labor Statistics under OMB No. 1220-0141. We will use the information you provide for statistical purposes only. Your participation is voluntary, and you have the right to stop at any time. This survey is being administered by Qualtrics and resides on a server outside of the BLS Domain. The BLS cannot guarantee the protection of survey responses and advises against the inclusion of sensitive personal

information in any response. By proceeding with this study, you give your consent to participate in this study.

Appendix A: Design Summary

n = 1260 participants

Between Subjects Design			Within Subjects Design, across 24 cases			
<i>Framing for Suggested Codes</i>	<i>Probability Information for 3 Suggested Codes</i>	<i>Timing of Suggested Codes</i>	<i>Narrative Information</i>	<i>Congruency with Expectations</i>	<i>Similarity of Suggested Codes</i>	
Framing of “Computer can be wrong sometimes”	Display probabilities	Before Submit Code Choice	Job duties included	Congruent Suggested Probabilities	Easy	3 cases
		After Submit Code Choice			Ambiguous	3 cases
	Display ranks	Before		Incongruent Suggested Probabilities	Easy	3 cases
		After			Ambiguous	3 cases
	Display no information	Before		Congruent Suggested Probabilities	Easy	3 cases
		After			Ambiguous	3 cases
Framing of “Check your work against the computer”	Display probabilities	Before	Job duties excluded	Incongruent Suggested Probabilities	Easy	3 cases
		After			Ambiguous	3 cases
	Display ranks	Before		Congruent Suggested Probabilities	Easy	3 cases
		After			Ambiguous	3 cases
	Display no information	Before		Incongruent Suggested Probabilities	Easy	3 cases
		After			Ambiguous	3 cases
Control condition: No suggested codes						

The block of cases with job duties will always be shown first. The order of cases of congruent vs. incongruent and easy vs. ambiguous will be randomized for each participant. For participants in the Control condition, the suggested codes and the within-subjects manipulations regarding suggested codes will not be shown.

Appendix B: Summary of suggested codes

Case ID	Correct Code	Ambiguous Set of Suggested Codes			Easy Set of Suggested Codes		
59623	Automotive Service Technician/Mechanic	Automotive Service Technician/Mechanic	Bus and Truck Mechanic	Electric Motor Repairer	Automotive Service Technician/Mechanic	Industrial Truck and Tractor Operator	Automotive Glass Installer and Repairer
74521	Nursing Assistant	Nursing Assistant	Personal Care Aide	Physical Therapist Aide	Nursing Assistant	Medical Records Information Technician	Social Worker
45298	Electrician	Electrician	Electrical Equipment Assembler	Electrical Engineer	Electrician	Construction Worker	Computer Hardware Engineer
63952	Secretary/Administrative Assistant	Secretary/Administrative Assistant	File Clerk	Receptionist	Secretary/Administrative Assistant	Attorney	Correctional Officer
48972	Elementary School Teacher	Elementary School Teacher	Child Psychologist	School Counselor	Elementary School Teacher	Interpreter and Translator	School Nurse
19102	Heavy and Tractor-Trailer Truck Driver	Heavy and Tractor-Trailer Truck Driver	Freight Mover	Delivery Services Driver	Heavy and Tractor-Trailer Truck Driver	Sales Worker	Taxi Driver
19108	Butcher/Meat Cutter	Butcher/Meat Cutter	Supermarket Cashier	Food Preparer	Butcher/Meat Cutter	Baker	Delivery Services Driver
20451	Security Guard	Security Guard	Transportation Security Screener	Police and Sheriff's Patrol Officer	Security Guard	Delivery Services Driver	Computer Systems Analyst
53741	Stock Clerk	Stock Clerk	Retail Salesperson	Merchandise Displayer	Stock Clerk	File Clerk	Cashier
42856	Farmer	Farmer	Agricultural Inspector	Landscaping and Groundskeeping Worker	Farmer	Hunter and Trapper	Painter for Construction/Maintenance
78216	Firefighter	Firefighter	Medical Assistant	Hazardous Materials Removal Worker	Firefighter	Human Resources Specialist	Lifeguard
60321	Retail Salesperson	Retail Salesperson	Stock Clerk	Merchandise Displayer	Retail Salesperson	Driver	Advertising Sales Agent
85602	Police and Sheriff's Patrol Officer	Police and Sheriff's Patrol Officer	Security Guard	Detective and Criminal Investigator	Police and Sheriff's Patrol Officer	Crossing Guard	Attorney

Case ID	Correct Code	Ambiguous Set of Suggested Codes			Easy Set of Suggested Codes		
45820	Plumber	Plumber	Pipe layer	Heating, Air Conditioning, and Refrigeration Installer	Plumber	Roofer	Sheet Metal Worker
64587	School Bus Driver	School Bus Driver	Crossing Guard	Transportation Attendant	School Bus Driver	Childcare Worker	Delivery Services Driver
48752	Veterinarian	Veterinarian	Animal Caretaker	Zoologist and Wildlife Biologist	Veterinarian	Animal Breeder	Medical Records and Health Information Technician
42389	Garbage Collector	Garbage Collector	Heavy Truck Driver	Grounds Maintenance Worker	Garbage Collector	Construction Worker	Highway Maintenance Worker
29631	Telemarketer	Telemarketer	Retail Salesperson	Market Research Analyst	Telemarketer	Office Clerk	Door-to-Door Sales Worker
14329	Flight Attendant	Flight Attendant	Ticket Agent	Food Server	Flight Attendant	Commercial Pilot	Entertainment Attendant
59290	Hairdresser and Hairstylist	Hairdresser and Hairstylist	Makeup Artist	Skincare Specialist	Hairdresser and Hairstylist	Health Technologist	Maid/ Housekeeping Cleaner
15858	Photographer	Photographer	Film and Video Editor	Audio and Video Equipment Technician	Photographer	Writer	Computer Programmer
53279	Construction Worker	Construction Worker	Forklift Operator	Industrial Machinery Mechanic	Construction Worker	Sheet Metal Worker	Protective Service Worker
32875	Baker	Baker	Event Planner	Dishwasher	Baker	Head Chef/Cook	Cashier
27864	Postal Service Mail Carrier	Postal Service Mail Carrier	Delivery Services Driver	Billing and Posting Clerk	Postal Service Mail Carrier	Courier/Messenger	Traffic Clerk

Appendix C: Introduction to Survey

Welcome!

Thanks for your interest in our research. We're conducting this study to better understand how people make classifications. Although the scenarios you will read are fictional, they are similar to a real task that real people do.

We need your help making our injured employee database more complete. On the next several pages, you will read about employees who were injured on the job. We have information about the injuries and the company name and industry. But we only have generic job titles for each employee. Your task is to classify these employees into more specific job titles.

We will pay bonuses of \$0.50 to workers with high accuracy rates. You'll find out whether you earned the bonus at the end of the HIT.

Unlike some surveys or online tasks, we ask that you complete this task all at one time. Please begin only when you are in a quiet place where you won't be disturbed for about 30 minutes.

Please do not use your browser's back button.

This voluntary study is being collected by the Bureau of Labor Statistics under OMB No. 1220-0141. We will use the information you provide for statistical purposes only. Your participation is voluntary, and you have the right to stop at any time. This survey is being administered by Qualtrics and resides on a server outside of the BLS Domain. The BLS cannot guarantee the protection of survey responses and advises against the inclusion of sensitive personal information in any response. By proceeding with this study, you give your consent to participate in this study.

--- page break ---

Appendix D: Coding training

Instructions

For each employee, we will show you an Injury Case Sheet, which includes their generic job title, industry, job duties, and injury information, along with 10 potential job titles. Your task is to classify the employee into the specific job title that best fits the information from the Injury Case Sheet, following these rules:

(1) Select the job title based on the *job duties* the employee performs, as listed on the Injury Case Sheet.

(2) Select only one job title. If two or more seem to fit, select the job title that best matches the job duties the employee spends the *majority of time* doing.

(3) Consider all 10 of the potential job titles before making your final selection.

TIP: Read about what workers were doing when they got injured. This can sometimes help provide additional information about their job duties.

TIP: Read the company name and industry to narrow down what types of jobs are typically in that industry.

--- page break ---

[Insert condition-specific instructions regarding framing, probability information, and timing]

Now let's look at a practice case.

Appendix E: Training cases

Here is a practice case. We will guide you through the process of selecting job titles. Then you will complete some cases on your own.

Injury Case Sheet			
Company Name: <i>Bernie's Café & Restaurant</i>		Employee ID: <i>30700</i>	
Job Title	Industry		
<i>Restaurant Worker</i>	<i>Food Service</i>		
Job Duties			
<i>1) Take orders, serve food and beverage to customers at tables in dining establishment (70% of time)</i> <i>2) Prepare food and nonalcoholic beverages at counter (25% of time)</i> <i>3) Welcome customers, seat them at tables or lounges, and help ensure quality of facilities and services (5% of time)</i>			
What was the employee doing just before the incident occurred?	What happened?	What was the injury or illness?	What object or substance directly harmed the employee?
<i>Preparing drinks at counter</i>	<i>Liquid spilled on the floor, employee slipped</i>	<i>Broken arm</i>	<i>Slippery floor</i>

[Provide suggestions as appropriate. Allow participant to select a job title]

--- page break ---

You chose: [insert choice]

The best job title is *Waiter and Waitresses*. Let's walk through the selection process to see why.

Rule 1: *Select the job title based on the job duties the employee performs, as listed on the Injury Case Sheet.*

- The list of job duties for this employee includes taking orders, preparing drinks, and welcoming customers.
- *Note:* It is also sometimes helpful to look at the industry, company name, and description of the injury. This employee was injured while preparing drinks. However, the employee only does this task 25% of the time, so it is not the main job duty.
- *Note:* The company name and industry indicate the job is one found in the restaurant and food services.

Rule 2: *You may only select one job title. If two or more job titles seem to fit, select the job title that best matches the job duties the employee spends the majority of time doing.*

- This employee mainly takes orders and serves food to customers, and sometimes prepares food and greets customers. Since taking orders and serving food is the main job duty, select the job title that matches “taking orders and serving food” most closely.

Rule 3: Read all 10 of the potential job titles before selecting making your final selection.

- This employee takes orders and serves food to customers most of the time. Read all 10 job titles and eliminate any that have nothing to do with this job duty, such as: *Sales Worker, Stock Clerk, or Cashier.*

Best job title: *Waiter and waitress*

- Explanation: this is the best job title because it closely matches the main job duties (taking orders and serving food to customers).
- Note: *Host and Hostess* or *Bartender* were also possibilities, but didn't reflect the main job duties for this employee.

--- page break ---

Now let's try one more.

Injury Case Sheet			
Company Name:		<i>Peachy Clean</i>	
Employee ID:		<i>42037</i>	
Job Title	Industry		
<i>Cleaner</i>	<i>Janitorial Services</i>		
Job Duties			
<i>1) Perform heavy cleaning duties, such as cleaning floors, shampooing rugs, washing walls and glass, and removing trash (50% of time)</i> <i>2) Cleaning snow or debris from sidewalk (35% of time)</i> <i>3) Tending to the Furnace and boiler of building (15% of time)</i>			
What was the employee doing just before the incident occurred?	What happened?	What was the injury or illness?	What object or substance directly harmed the employee?
<i>Cleaning the floor</i>	<i>Wiped face, got chemicals in their eyes</i>	<i>Eye infection</i>	<i>Chemical cleaner</i>

[Provide suggestions as appropriate. Allow participant to select a job title]

--- page break ---

You chose: [insert choice]

The best job title is *Janitors and Cleaners*. Let's walk through the process one more time and each of the three selection rules.

Rule 1: Select the job title based on the job duties the employee performs, as listed on the Injury Case Sheet.

- This rule says to focus on the tasks the worker performs. The employee performs heavy cleaning duties most of the time.
- *Note:* The worker was cleaning the floor and got chemicals in their eyes. This is part of their main job duty --cleaning floors.
- The company name and industry indicate the job is one found in the janitorial services.

Rule 2: You may only select one job title. If two or more job titles seem to fit, select the job title that best matches the job duties the employee spends the majority of time doing.

- Since heavy cleaning is the main job duty, we'd select the job title that matches these duties most closely.

Rule 3: Read all 10 of the potential job titles before making your final selection.

- Eliminate job titles that having nothing to do with heavy cleaning, such as *Construction* and *Building Inspectors* and *Refuse and Recyclable Material Collectors*

Best job title: *Janitors and Cleaners*

- Explanation: this is the best job title because it closely matches the main job duties (heavy cleaning) and the job is in the janitorial services industry.
- *Note:* *Maids and Housekeeping Cleaners* might have also been possible, but the other job duties such as cleaning debris from the sidewalk or tending to furnaces are not typically performed for maids and housekeepers.

--- page break ---

You have completed the training! Now continue on to make the classifications on your own. Remember, if your accuracy rate is high enough, you will earn a bonus of \$0.50. You'll find out whether you qualify for the bonus at the end of the HIT.

Appendix F: Case-level debriefing questions

After each case, on a separate page, ask:

How confident are you that you selected the correct job title? If you guessed, then choose “not at all confident”.

- Not at all confident
- Slightly confident
- Moderately confident
- Very confident
- Extremely confident

Present this question to only those participants who saw suggested codes:

How confident are you in the job titles that the computer suggested to you?

- Not at all confident
- Slightly confident
- Moderately confident
- Very confident
- Extremely confident

How easy or difficult was it to make your selection?

- Extremely easy
- Very easy
- Somewhat easy
- Neither easy nor difficult
- Somewhat difficult
- Very difficult
- Extremely difficult

How many of the job titles, job duties, and activities were you at least moderately familiar with?

- None
- Fewer than half
- About half
- More than half
- All

Appendix G: Final debriefing questions

1. How many categorization HITs had you completed before this one?

- None
- 1-5
- 6-50
- 51 or more

2. What is your age? ____ [validate two digits]

3. What is your gender?

- Male
- Female

4. Which of the following best describes your highest level of education?

- Less than high school
- High school diploma or equivalent
- Some college
- Associate's degree or Bachelor's degree
- Master's degree or Doctoral degree

5. As a whole, how well did the computer suggestions perform in comparison to your expectations?

1

2

3

4

5

(much worse)

(much better)

6. What are your thoughts and feelings on the computer suggestions?

[open-ended]

7. Do you have any other comments on this study?

[open-ended]

References

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2014). Algorithm Aversion: People Erroneously Avoid Algorithms After Seeing Them Err. *Journal of Experimental Psychology: General* (144), 114-126. doi: <http://dx.doi.org/10.1037/xge0000033>

Krosnick, J. A., Narayan, S., & Smith, W. R. (1996). Satisficing in surveys: Initial evidence. *New Directions for Evaluation* (70), 29-44.

Measure, A C. (2014). Automated Coding of Worker Injury Narratives. *Joint Statistical Meetings, Government Statistics Section*.