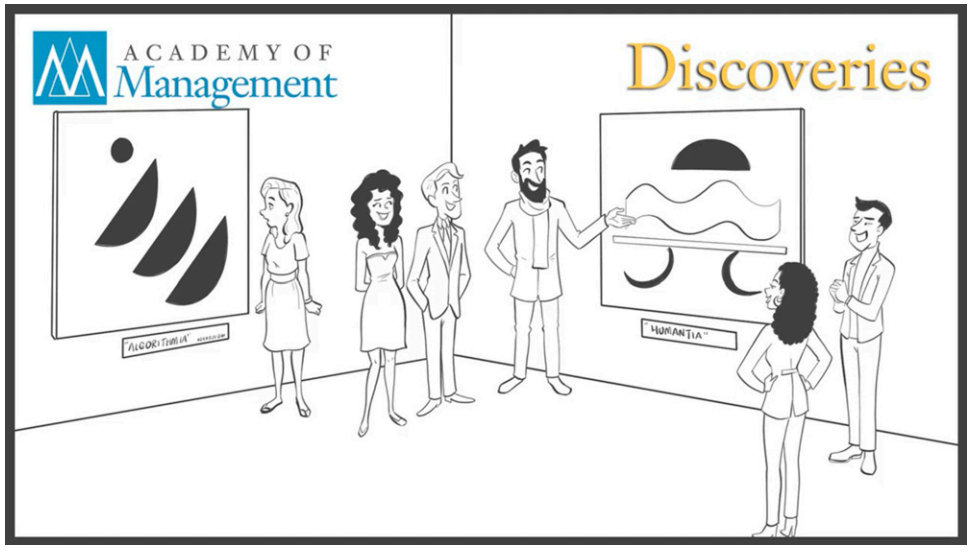
© Academy of Management Discoveries
2019, Vol. 5, No. 1, 38–56.
Online only
https://doi.org/10.5465/amd.2017.0002

ALGORITHMS AND AUTHENTICITY

ARTHUR S. JAGO[1]
Stanford University



As technology advances, artificially intelligent algorithms are becoming increasingly capable of human work. Across four experiments, I investigate people's beliefs about the authenticity of algorithmic work, compared with human work. People believe algorithmic work is less authentic than human work because they believe it exhibits comparatively less moral authenticity, or sincerity relevant to a specific category (Experiments 1 and 2). However, people do not distinguish between human and algorithmic work when it comes to type authenticity, or accuracy when it comes to representing a specific category. Because of these authenticity attributions, people also believe algorithms' characteristically moral decisions are relatively less ethical than identical human decisions (Experiment 3). To address these perceptions, organizations can increase algorithms' seeming authenticity by highlighting human involvement in their creation or training processes (Experiment 4). I discuss implications given the increasing prevalence of automation.

INTRODUCTION

Technology is changing how organizations do work. As algorithms—or computerized systems designed to achieve specific goals—become more sophisticated, organizations are rapidly automating numerous aspects of business, from making forecasts or recommendations to performing physical tasks (Koren, Bell, & Volinsky, 2009; Yeomans, 2015). Whereas these technologies might be internally useful in a variety of organizational contexts, what are the consequences of their implementation? In this article, I examine people's attributions of authenticity to the work algorithms do, such as decisions artificial intelligences make or products sophisticated computers design, compared with identical human work. Broadly, I find that people believe algorithmic work is as authentic as human work when it comes to representing or reproducing aspects of specific categories, but less authentic than human work when it comes to capturing sincerity or valuation underlying those categories. These perceptions, in turn, influence how people evaluate otherwise identical work and behaviors and also offer a framework for how organizations can restore authenticity to machines, given their increasing prevalence.

The author thanks Glenn Carroll, Anthony Vashevko, Jacob Model, and Michal Kosinski for their extremely helpful feedback as well as the Stanford Behavioral Lab, Margaret Neale, and Lindred Greer for helping facilitate data collection.

[1] Corresponding author.

Algorithms in Organizations

Following the industrial revolution, machines facilitated the creation of physical products at large scales (Allen, 2009). As algorithms became more powerful, numerous organizations began using computers for a variety of tasks, from managing and interpreting data to guiding rockets in space (Grove & Meehl, 1996; Meehl, 1954; Shelton, 1975). While this is still the case, modern advances in computing have allowed organizations to implement AI agents that can "learn" and solve problems in increasingly complicated domains, from creating products to autonomously managing business processes (Kamps, 2016; Mann & O'Neil, 2016). As a consequence, these technologies have threatened the automation of human jobs to the point where many academic and business leaders believe that proprietary technologies will eventually be more predictive of an organization's success than its actual human resources (Brynjolfsson & McAfee, 2014; Crandell et al., 2016; Ford, 2016; Kaplan, 2015; The Trillion-Dollar Difference, 2016).

As technology becomes more important to society, researchers across academic disciplines have focused increased attention toward questions concerning how people interact with computers and other technological agents (Cerulo, 2009; Dietvorst, Simmons, & Massey, 2015, 2016; Newell & Card, 1985; Waytz, Heafner, & Epley, 2014; see Frick, 2015 for a short review). In general, people often anthropomorphize computerized agents—imbuing them with human qualities and mental states—to interact with them, sometimes even invoking human stereotypes (Epley, Waytz, & Cacioppo, 2007; Gray, Gray, & Wegner, 2007). For example, people believe security robots with stereotypically masculine names and voices will be more effective than female-sounding ones (Tay, Park, Jung, Tan, & Wong, 2013). People are also more comfortable outsourcing emotional labor to human-looking robots (compared with otherwise identical—but artificial looking—agents; Waytz & Norton, 2014) and are also more likely to comply with robots whose demeanor matches a social situation, e. g., a "serious" robot promoting exercise by espousing its health benefits or a more "playful" robot instructing them to compare different jelly beans (Goetz, Kiesler, & Powers, 2003). In light of this research, the process of automation raises a broad question: How do people respond to technological agents, such as these, doing work that they generally expect other humans to do?

Aspects of Authenticity

Authenticity broadly refers to an attribution of genuineness. Scholars have long debated exactly what authenticity is or is not, yielding multiple potential definitions (e.g., what is authentic is real, genuine, or true; what is not authentic is phony, fraudulent, or insincere) as well as different interpretations of the construct (Carroll & Wheaton, 2009; Grayson & Martinec, 2004; Peterson, 2005). Ultimately, the criteria people use to determine whether or not something is "authentic" vary across situations and cultural environments (Lehman, Kovács, & Carroll, 2014; Newman, 2016). For example, people might apply different standards when judging if a brewery is brewing authentic beer (Carroll & Swaminathan, 2000) compared with judging if another person's behavior is indicative of their authentic self (Gino, Kouchaki, & Galinsky, 2015). Multiple scholars have categorized these divergent interpretations of authenticity under unified frameworks. For example, Carroll and Wheaton (2009) argued that scholarly accounts of authenticity broadly separate into two categories: *type authenticity* and *moral authenticity* (Carroll, 2015; Carroll & Wheaton, 2009; other typologies posit additional domains as well; e.g., Newman & Smith, 2016). In Carroll and Wheaton's (2009) model, type authenticity refers to the degree to which a behavior or object meets an audience's criteria for classification into a specific category. For example, type-authentic cuisine might use certain ingredients or ultimately resemble people's expectations about taste, appearance, or presentation. By contrast, moral authenticity refers to the degree to which a behavior or object reflects what an observer believes are sincere choices, beliefs, or values. Unlike type authenticity, which concerns qualities of category membership, moral authenticity concerns whether something is genuine by virtue of an agent experiencing and conveying values, as opposed to acting in response to social scripts or other tangential motivations (Carroll & Wheaton, 2009). To use the same example, for people to classify a cuisine as morally authentic, a chef might have to convey a deep respect for the history or importance of a particular dish that ultimately drove him or her to prepare it, as opposed to introducing it simply because consumers wanted it, or claiming to value its history while not actually caring about it. As such, type authenticity concerns whether or not something is genuine by virtue of reflecting a specific category or genre, whereas moral authenticity concerns whether or not something is genuine by virtue of reflecting experienced values, choices, or motivations.

Despite ongoing debate about the relative importance of different criteria when it comes to people's omnibus impressions of authenticity, numerous scholars have posited that authenticity motivates people's decisions in a variety of social and market contexts (Avolio & Gardner, 2005; Sartre, 1943; Spooner, 1988; Wang, 1999). Broadly, successfully signaling authenticity

Author's voice:
What motivated you personally to undertake this research?

Author's voice:
Was there anything that surprised you about the findings?

benefits organizations. Appearing authentic can increase an organization's apparent quality (Carroll & Swaminathan, 2000; Jones, Anand, & Alvarez, 2005; Kovâcs, Carroll, & Lehman, 2013), bolster consumer support (Castéran & Roederer, 2013), and improve the presentation of social responsibility initiatives (Alhouti, Johnson, & Holloway, 2016). People also believe authentic organizational products are more valuable and desirable than seemingly inauthentic counterparts (Frazier, Gelman, Wilson, & Hood, 2009). In addition to collectives or organizations, authenticity also generally benefits individual agents. For example, signaling authenticity often helps people lead or persuade others (Avolio & Gardner, 2005; Shamir & Eilam, 2005; Walumbwa, Avolio, Gardner, Wernsing, & Peterson, 2008) and bolsters positive judgments such as friendliness (Grandey, Fisk, Mattila, Jansen, & Sideman, 2005). People even exhibit different neural responses when evaluating authentic objects compared with inauthentic ones, e.g., ostensibly legitimate versus forged artwork (Huang, Bridge, Kep, & Parker, 2011; Vartanian & Skov, 2014). These benefits also create incentives to signal authenticity, e.g., by constructing seemingly authentic ideas, products, or situations, to receive these benefits despite actually lacking moral or type-authentic criteria (Beverland, 2005; MacCannell, 1973; Peterson, 2005).

Algorithms and Humans

As algorithms become more sophisticated, they are also becoming increasingly capable of doing work in organizations. However, their equality to humans in this regard does not necessarily mean that people judge these two entities' work similarly. How do people assess the relative authenticity of computerized work? On one hand, people might assume that algorithmic work is relatively type authentic. People generally believe technological agents are calculative (Gray et al., 2007), and algorithms and other robots are already prevalent in a number of work domains involving precise category replication, e.g., working in assembly lines or even moving objects to specific points around a warehouse. Indeed, algorithms' increasing ability to accurately mimic human work—often at a massive scale, cheaply, and efficiently—is precisely what threatens the automation of numerous industries (Brynjolfsson & McAfee, 2014; Ford, 2016).

However, people might not believe algorithmic work is as morally authentic as human work. People do not rely only on the specifics of a particular behavior or end product to infer authenticity; they also sometimes rely on their stereotypes about different agents engaging in work (Newman & Smith, 2016). Most typologies suggest that recognizing mental states of valuation is necessary to attribute moral authenticity to work (Carroll & Wheaton, 2009; also see similar conceptions of expressive and value authenticity; e.g., Lindholm, 2013; Newman & Smith, 2016). Although people might generally recognize that humans are conscious agents capable of valuing work (but see Haslam, 2006), they might not share the same belief regarding unconscious machines. Although algorithms and other computerized systems might appear capable of mindlessly replicating the type-authentic qualities people associate with specific categories (e.g., a specific safety feature in a mass-produced vehicle), people may not believe that machines—by virtue of their unconsciousness—can express any purpose or valuation underlying those qualities (e.g., the importance of these safety features), and as such, that they convey less moral authenticity. Given algorithms' increasing prevalence in business, this potential asymmetry led to my first research question: Do people believe algorithms exhibit less authenticity than humans, and if so, why?

The idea that people might differentiate between humans and algorithms when it comes to authenticity has an important caveat. AI and computerized processes are almost always instantiated by (and are often monitored by) humans. However, other algorithms operate independently from people, e.g., algorithms that autonomously engage in "flash" stock trading (Subramanian, Ramamoorthy, Stone, & Kuipers, 2006). These practical realities raise the question of how different kinds of human interventions influence people's judgments about whether or not machines' products, decisions, or behaviors are authentic. For example, although people might not believe that an algorithm can appreciate or value the significance of improving vehicle safety systems, they might believe that the programmer who designed and implemented that algorithm could. In turn, human involvement might influence people's authenticity judgments by virtue of framing an algorithm—despite its unconsciousness—more as a humanized tool implemented by a person, blurring the line between human and machine and attenuating any differences between the two. This reasoning led to my second research question: How does human involvement

with algorithms influence people's attributions of authenticity?

These questions carry both practical and theoretical implications. Most directly, they speak to existing research on automation by offering one perspective on how people naturally compare technological agents with humans they replace (Brynjolfsson & McAfee, 2014; Lin, Abney, & Bekey, 2011). People's authenticity judgments motivate their behavior in a variety of important domains; their perceptions of relative authenticity might be a similarly important component driving how people respond to organizational behaviors in increasingly technological commercial environments. Similarly, this research speaks to the multidisciplinary literature on authenticity (Carroll, 2015; MacCannell, 1973). Although much research focuses on how different objects or situations convey authenticity and subsequently benefit organizations, relatively little research has explored how people form authenticity judgments when perceiving or interacting with AI agents *in lieu* of humans (Turkle, 2007). In addition, this research speaks to how organizations can maintain beneficial signals of authenticity as work becomes more automated over time. Appearing authentic generally benefits organizations; understanding how people attribute authenticity to machines—as well as how organizations might address perceptions of relative inauthenticity—carries implications for how organizations can capitalize on the usefulness and accuracy of technological systems while maintaining these beneficial signals (Connelly, Certo, Ireland, & Reutzel, 2011; Donaldson & Preston, 1995; Sawyer, 2007).

Overview of Experiments

Across four experiments, I explored when and why people differentiated between the authenticity of human and algorithmic work. I first investigated whether people actually expected or judged an algorithms' work to be less authentic than an identical human work, as well as the roles of type and moral authenticity in informing these attributions (Experiments 1 and 2). Next, I investigated one potential downstream consequence of these judgments: How people's authenticity attributions impacted their responses to humans' and algorithms' characteristically ethical (but otherwise identical) decisions (Experiment 3)? Finally, I explored how different kinds of human involvement change people's opinions of algorithms' authenticity (Experiment 4).

Author's voice:
If you were to do this study again, what would you do differently?

EXPERIMENT 1

In Experiment 1, my goal was to investigate people's expectations about human and algorithmic authenticity. I recruited MBA students and asked them about either a human or an algorithm that did one of four kinds of work: created recipes, composed music, offered solutions to ethical dilemmas, or designed restaurant concepts. In addition to asking about participants' omnibus attributions of authenticity (Carroll, 2015), I also asked participants to further identify how typically and morally authentic they believed human or algorithmic work would be to explore the dimensions along which people might distinguish between the two.

Method

Participants. Across multiple class sessions, I advertised the experiment to approximately 300 MBA students taking an introductory course at a private West Coast university. One hundred seventy-five students (98 male, M_{age} = 26.93) ultimately elected to participate in the experiment online.

Procedure. I randomly assigned participants to answer questions about either a person's work (N = 82) or an artificial intelligence's work (N = 93) for the duration of the experiment. Participants imagined that the person [artificial intelligence] did one of four kinds of work for an organization: created recipes (N = 47), composed music (N = 41), solved ethical problems (N = 42), or created restaurant concepts (N = 45). This yielded a 2 (entity: person vs. artificial intelligence) × 4 (work: recipes vs. music vs. ethical solutions vs. restaurant concepts) design. I next asked participants to indicate using a 1 ("Extremely Inauthentic") to 7 ("Extremely Authentic") scale how authentic they expected the person's [artificial intelligence's] work would be. I used multiple different authenticity domains in an effort to address the possibility that any differences in perceived authenticity were only specific to one particular domain (Wells & Windschitl, 1999; see also Experiments 2 and 3).

After this, participants indicated their agreement using a 1 ("Strongly Disagree") to 7 ("Strongly Agree") scale to four additional questions adapted from Carroll and Wheaton's (2009) typology concerning moral and type authenticity. I constructed two-item *ad hoc* scales for both moral and type authenticity in an attempt to capture what Carroll and Wheaton's (2009) describe as fundamental elements of both constructs as well as identifiers that separate them from one another. In addition, I designed each item to refer to work itself—not the agent engaging in it—in an effort to be applicable to both humans and machines. To assess moral authenticity, I asked participants to indicate their expectations that the person's [artificial intelligence's] work would "Reflect the reasons why [recipes/music/ethical solutions/restaurant

concepts] are the way they are" (or "...is the way it is") and "Exhibit the sincere 'character' of [recipes/music/ethical solutions/restaurant concepts]." To assess type authenticity, I asked participants to indicate their agreement that the person's [artificial intelligence's] work would be "...similar to other [recipes/music/ethical solutions/restaurant concepts]" as well as "...made using the appropriate techniques" (Carroll & Wheaton, 2009). These items formed composites of characteristic moral authenticity ($r = 0.72$, $p < .001$) and characteristic type authenticity ($r = 0.45$, $p < .001$).[2] Like the present items, in each experiment, all dependent measures referred to the work the people or machines engaged in, not the person or machine itself, although people likely refer to both impressions about different agents as well as their work when it comes to informing authenticity judgments (Carroll & Wheaton, 2009; Newman & Smith, 2016).

Results and Discussion

I first created three 2 (entity: human vs. algorithm) × 4 (work) ANOVAs predicting authenticity, moral authenticity, and type authenticity, respectively. Results indicated that participants believed the algorithms' work would be less authentic ($M = 4.07$, SD = 1.31) than the person's work ($M = 4.68$, SD = 1.16; $F(1, 160) = 10.05$, $p = .002$). In addition, although participants believed the person's work would convey more moral authenticity ($M = 4.54$, SD = 0.91) than the algorithm's work ($M = 3.92$, SD = 1.28; $F(1, 157) = 12.37$, $p = .001$), they did not distinguish between human work ($M = 4.86$, SD = 0.84) and algorithmic work ($M = 4.83$, SD = 1.16) when it came to type authenticity, $F(1, 158) = 0.02$, $p = .884$. These models indicated neither omnibus effects of the different work domains ($F(3, 160) = 0.74$, $F(3, 157) = 2.05$, and $F(3, 158) = 0.59$; $ps > .11$) nor any interaction effects between the entity and work domain manipulations ($F(3, 160) = 0.16$, $F(3, 157) = 1.63$, and $F(3, 158) = 0.59$; $ps > .184$). These results suggested that participants' expectations were similar regardless of what kind of work the human or algorithm engaged in.

To create a mediation model exploring the role of type-authentic and morally authentic criteria in informing participants' omnibus authenticity judgments, I collapsed participants' responses across the different situations (Hayes, 2013). This model used entity (0 = human and 1 = algorithm) as the independent variable, authenticity as the dependent variable, and both moral and type authenticity as simultaneous mediators. Results indicated that a decrease in expected moral authenticity mediated participants' expectations that the algorithm's work would be relatively less authentic than the person's work, 95 percent confidence interval ($CI_{95} = [-0.32, -0.02]$). Type authenticity, however, did not mediate this effect, $CI_{95} = [-0.40, 0.07]$ (see Figure 1 for mediation results from Experiments 1 and 2).

In sum, Experiment 1 examined how people forecasted the authenticity of human versus computerized work. Across different domains, participants generally expected that an algorithm's work (e.g., recipes and solutions to ethical problems) would be less authentic than a person's work. In addition, participants' expectations of moral authenticity—but not their expectations of type authenticity—explained this difference. Because participants believed algorithms' work would express fewer morally authentic qualities, they believed that algorithms' work would also be generally less authentic. However, people did not distinguish between human and algorithmic work when it came to their expectations about type-authentic qualities. As such, although Experiment 1 suggested that people might believe algorithms and humans are equally authentic in their ability to faithfully represent categories, it also suggested that people believe algorithms are relatively inauthentic when it comes to capturing the sincerity or purpose underlying those categories.

EXPERIMENT 2

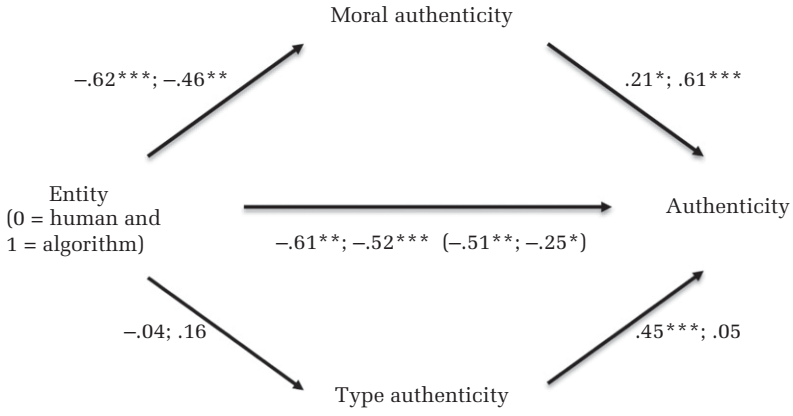
In Experiment 1, participants expected that algorithmic work would be less authentic than human work, in part because they believed algorithms would ultimately convey less moral authenticity. In Experiment 2, I tested whether or not these effects would replicate when participants actually evaluated either a human or algorithm's work. Although people are often forced to predict authenticity in organizational contexts, they also often have access to specific products, decisions, or ideas to evaluate. In Experiment 2, I used the same general design as in Experiment 1, with one change: participants actually evaluated either a painting (Sample A) or a song (Sample B) ostensibly created by either a human or an AI algorithm.

Method

Participants. For Sample A, 201 self-reported American adults (99 male, $M_{age} = 33.37$) completed the experiment using Amazon's Mechanical Turk. For Sample B, 200 different self-reported American adults (116 male, $M_{age} = 32.88$) completed the experiment using Amazon's Mechanical Turk. Previous research suggests that experimental research using Mechanical Turk workers is approximately as reliable as other laboratory-based approaches (Buhrmester, Kwang, & Gosling, 2011), although it risks other issues such as non-naivete or incentives for dishonesty

[2] Other typologies use different identifiers for criteria Carroll and Wheaton (2009) described using moral and type authenticity, e.g., Newman and Smith's (2016) value authenticity and categorical authenticity, respectively.

FIGURE 1
The Mediating Effects of Moral and Type-Authentic Criteria on Perceptions of Authenticity (Experiment 1 and Experiment 2, Respectively)



*$p < .05$
**$p < .01$
***$p < .001$

(Chandler, Mueller, & Paolacci, 2014). Because Experiments 2–3 sampled across multiple authenticity domains and Experiment 4 used eight experimental conditions, I used Mechanical Turk primarily to ensure adequate power in each study. Despite the limitations imposed by conducting experiments using technological environments, both consumers and stakeholders often conduct and evaluate work online (Brynjolfsson & McAfee, 2014). In turn, many of people's authenticity judgments actually have the potential to become similarly digitized as technology advances (e.g., looking at a restaurant menu online and deciding if it is authentic; Kovács et al., 2013).

Procedure. For both samples, I randomly assigned participants to answer questions about either a person's work (*Ns* = 101 and 96, respectively) or an artificially intelligent algorithm's work ("AI"; *Ns* = 100 and 104, respectively) for the duration of the experiment. In Sample A, participants read that this person [AI] created paintings and that they would see one and answer a few questions about it. I next randomly assigned participants to see one of five paintings the person [AI] ostensibly created (*Ns* from 40 to 41). In reality, each of the paintings was actually created by an algorithm (Lindemeier, Pirk, & Deussen, 2013; Gayford, 2016; see Appendix for images). In Sample B, participants read that this person [AI] composed songs and that they would hear one and answer a few questions about it. I next randomly assigned participants to a webpage link where they heard one of four different 30-second song clips the person [AI] ostensibly created (*Ns* from 38 to 57). Like in Sample A, an algorithm—not a human—actually created each song clip (Bozhanov, 2016). Like in Experiment 1, I used multiple paintings and songs, respectively, to help rule out the possibility that any effects were contingent on participants judging one particular piece of work.

After either seeing the painting or listening to the song clip, participants indicated how authentic they believed the painting or song was using same the 1 ("Extremely Inauthentic") to 7 ("Extremely Authentic") scale used in Experiment 1. Participants next responded to the same moral and type authenticity items used in Experiment 1, which again formed composites in both samples (Sample A: *rs* = 63 and 0.47, *ps* < .001; Sample B: *rs* = 0.72 and 0.52, *ps* < .001).[3] Although I collected both samples at different times, because

[3] The type authenticity items in Experiments 1 and 2 exhibited somewhat low reliability (*rs* = 0.45, 0.47, and 0.52), likely reflecting how the two items assessed category similarity and category technique, which despite both being broad components of type authenticity (Carroll & Wheaton, 2009), do not necessarily overlap. For example, an entity could engage in work that ultimately resembles similar work, but was performed incorrectly. When analyzed separately, entity predicted neither item in Experiment 1 (*Fs* < 0.526, *ps* > .469). In Experiment 2, although people did not distinguish between algorithms and humans when it came to category technique ($F(1, 383) = 1.46$, $p = .228$), they actually believed the algorithms' paintings and songs were more similar to other category members ($M = 4.55$, SD = 1.61) compared with the humans' work ($M = 4.06$, SD = 1.59; $F(1, 383) = 10.51$, $p = .001$), perhaps because people assume machines necessarily use prototypes to "create" art, whereas humans might not. Crucially though, neither item significantly mediated participants' omnibus judgments of authenticity in either experiment, suggesting that participants' moral authenticity attributions tend to play a more primary role when it comes to comparing the two entities' work authenticity.

participants answered identical questions, I combined the two datasets into one ($N = 401$) for simplicity. A factor representing content type—paintings and songs, respectively—did not interact with entity to predict any dependent measure, $ps > .753$, although participants did believe the paintings were broadly more authentic ($b = 0.23$, $p < .001$), morally authentic ($b = 0.33$, $p < .001$), and type authentic ($b = 0.23$, $p < .001$) than the songs. This was likely because of algorithms' relative sophistication when it comes to creating paintings compared with music (see supplemental materials); however, all significant main, interactive, and indirect effects remained significant ($ps < .05$) when I analyzed both samples separately as well.

Results and Discussion

Similar to Experiment 1, I created three 2 (entity: human vs. algorithm) × 9 (work) ANOVAs predicting authenticity, moral authenticity, and type authenticity, respectively. Overall, participants believed that the human work was more authentic ($M = 5.44$, SD = 1.18) than the algorithmic work ($M = 4.92$, SD = 1.39; $F(1, 382) = 15.81$, $p < .001$). In addition, participants again indicated that the person's work was more morally authentic ($M = 4.86$, SD = 1.35) than the algorithm's work ($M = 4.39$, SD = 1.45; $F(1, 383) = 10.12$, $p = .002$). However, participants again did not significantly distinguish between humans ($M = 4.53$, SD = 1.17) and algorithms ($M = 4.68$, SD = 1.30) when it came to the works' type-authentic qualities, $F(1, 383) = 2.21$, $p = .138$. Reflecting the aforementioned discontinuity between paintings and songs, these models indicated an omnibus effect that the different works signaled different levels of authenticity ($F(8, 382) = 4.22$, $p < .001$), moral authenticity ($F(8, 383) = 7.32$, $p < .001$), and type authenticity ($F(8, 383) = 3.66$, $p < .001$). In addition, whereas entity did not interact with this factor to predict omnibus authenticity ($F(8, 382) = 1.52$, $p = .150$) or moral authenticity ($F(8, 383 = 1.17, p = .315$), they did marginally interact to predict type authenticity, $F(8, 383) = 1.90$, $p = .058$. This result suggested that participants' judgments concerning some works' type appropriateness indeed changed as a function of entity (see General Discussion).

Like in Experiment 1, I next aggregated each composite across the different works and work domains to compute a 5,000-iteration bootstrapped mediation model using entity (0 = human and 1 = algorithm) as the independent variable, authenticity as the dependent variable, and moral and type authenticity as mediators. Results again indicated that participants' beliefs that the algorithmic work expressed fewer morally authentic qualities mediated their beliefs that it was also generally less authentic, $CI_{95} = [-0.46, -0.11]$. However, type authenticity did not mediate participants' omnibus authenticity judgments, $CI_{95} = [-0.01, 0.04]$ (see Figure 1). In sum, Experiment 2 generally replicated Experiment 1, but with actual evaluations of ostensibly human or algorithmic work. When participants actually viewed paintings or listened to music, they believed that human work was more authentic than an AI algorithm's otherwise identical work. Like in Experiment 1, participants' attributions of morally authentic criteria—but not type-authentic criteria—explained this difference.

EXPERIMENT 3

In Experiment 3, I sought to investigate how relative perceptions of algorithms' authenticity could impact people's responses to organizations that use machines. Although organizations often use algorithms in characteristically moral situations (e.g., piloting a plane safely to protect its passengers), recent advances in technology have catalyzed both public and academic concerns about the ethical implications of artificial intelligence (O'Neil, 2016; Wallach, Franklin, & Allen, 2010), e.g., the case of self-driving cars that resolve dilemmas by choosing which agent(s) to harm in a collision (Bonnefon, Shariff, & Rahwan, 2016). As algorithms become more sophisticated, organizations will not only use them for low-level or rote work: they will also implement such technologies in a variety of potentially moral domains, from operating vehicles and informing medical diagnoses (IBM, 2016) to facilitating large-scale financial transactions (Subramanian et al., 2006). When behaving ethically, an important component to successfully signaling the merits of such behavior is to signal some level of sincerity or genuineness underlying it (Barasch, Levine, Berman, & Small, 2014; Gino et al., 2015). Following Experiments 1 and 2, if people believe algorithms' behaviors generally lack morally authentic qualities, they may also not fully appreciate algorithms' ethically desirable decisions, even if those decisions appear categorically type authentic or correct by moral standards. As a function of these attributions, people may believe algorithms' decisions are less ethical than human decisions, even if the content or eventual outcomes of those decisions are indeed identical (Gray & Wegner, 2009; Pizarro & Tannenbaum, 2011; see also; Pizarro, Tannenbaum, & Uhlmann, 2012). In Experiment 3, I investigated these potential downstream consequences by asking participants about the authenticity underlying human or algorithmic—but ethically desirable—decisions.

Method

Participants. Two hundred self-reported American adults (103 male, $M_{age} = 36.01$) completed

the experiment online using Amazon's Mechanical Turk.

Procedure. I randomly assigned participants to answer questions about either a human ($N = 100$) or artificial intelligence ($N = 100$) that engaged in one of five characteristically ethical behaviors: autonomously improving the safety features of a car ($N = 40$; see the "Ford Pinto" case, Grimshaw v. Ford Motor Co., 1981), fixing a client's electrical issue ($N = 39$), correcting a payment error ($N = 41$), signing up an underage patient for a useful medical trial ($N = 40$), or changing highways to avoid killing migrating crabs ($N = 40$; see Appendix for materials). Following this description, participants responded to the question "How authentic was this person's [AI's] decision?" using the same 1 ("Extremely Inauthentic") to 7 ("Extremely Authentic") scale used in previous experiments. I also assessed two potential downstream consequences to perceived authenticity: how ethical participants believed the decision's outcome would be and how much they liked the decision. To assess outcome ethicality, participants indicated using a 1 ("Extremely Unlikely") to 7 ("Extremely Likely") scale their opinions about three items: "This person's [AI's] decision will produce an ethical outcome," "This person's [AI's] decision will produce a moral outcome," and "This person's [AI's] decision will produce a socially responsible outcome." These items formed a reliable composite of outcome ethicality ($\alpha = 0.88$). To assess liking, participants indicated their agreement using a 1 ("Strongly Disagree") to 7 ("Strongly Agree") scale to two additional items: "I like this person's [AI's] decision" and "I approve of this person's [AI's] decision." These two items formed a reliable composite of liking ($r = 0.91$; $p < .001$).

At the end of the survey, participants also indicated their agreement (1 = "Strongly Disagree" and 7 = "Strongly Agree") to three additional statements: "This person [AI] should not have been responsible for this decision," "This person [AI] could consider alternatives when making this decision," and "I'm familiar with how AIs make decisions" (regardless of which agent participants read about). I asked about responsibility to investigate whether or not participants' beliefs that it might be relatively inappropriate for algorithms to make ethical decisions would influence their responses to these situations. Similarly, people's broad familiarity with algorithms, as well as potential stereotypes that algorithms might not be able to take into account different alternatives when making a decision, could impact their eventual liking of specific decisions.

Results and Discussion

I first created three 2 (entity: person vs. algorithm) × 5 (decision) ANOVAs predicting authenticity, outcome ethicality, and liking. Overall, participants believed that the human decisions were more authentic ($M = 6.12$, SD = 1.20) than the algorithmic decisions ($M = 5.62$, SD = 1.17; $F(1, 190) = 6.96$, $p = .009$). In addition, participants believed the human decisions would produce more ethical outcomes ($M = 6.13$, SD = 1.07) than the algorithms' decisions ($M = 5.57$, SD = 1.24; $F(1, 190) = 9.20$, $p = .003$), and also liked the human decisions ($M = 6.23$, SD = 1.29) relatively more than the algorithms' decisions ($M = 5.66$, SD = 1.36; $F(1, 190) = 7.39$, $p = .007$). All three models indicated that the entity manipulation did not interact with the different decision contexts, $Fs(4, 190) < 1.21$, $ps > .310$, suggesting that participants' judgments were similar across the different behaviors. All three models also indicated a significant omnibus effect of decision context, $Fs(4, 190) > 6.76$, $ps < .001$, suggesting that participants believed the different decisions signaled different amounts of authenticity and ethicality, and also elicited differential liking, regardless of the agent.

An identical ANOVA predicting inappropriateness indicated that people believed it was more inappropriate for an algorithm ($M = 4.03$, SD = 1.69) to be in the position to make moral decisions than a human ($M = 3.32$, SD = 1.64; $F(1, 190) = 7.52$, $p = .007$). Interestingly, people did not distinguish between humans ($M = 4.72$, SD = 1.64) and algorithms ($M = 4.66$, SD = 1.55) when it came to these agents' seeming capacity to recognize alternatives when making decisions, $F(1, 190) = 0.14$, $p = .705$. In addition, the entity manipulation did not affect participants' self-reported knowledge about how algorithms make decisions ($F1, 190) = 2.56$, $p = .111$, although this effect trended in suggesting that participants in the algorithm condition—perhaps as a function of reading and thinking about robotic processes over the course of the survey—reported more knowledge about algorithms ($M = 3.95$, SD = 1.74) than participants in the human condition ($M = 3.56$, SD = 1.54). These models indicated no interactive effects, $Fs(4, 190) < 1.38$, $ps > .245$, and including these items as covariates did not alter the significance levels in any of the earlier models saved for the main effect of the entity manipulation predicting omnibus authenticity, which became marginally significant, $F(1, 187) = 3.30$, $p = .071$.

Mediations through authenticity. As in previous experiments, I again collapsed participants' judgments across the distinct situations to create two 5,000-iteration bootstrapped mediation models using entity (0 = human and 1 = algorithm) as the independent variable, authenticity as the mediator, and ethicality or liking as dependent variables. These models indicated that participants' authenticity judgments mediated both their expectations about outcome ethicality, $CI_{95} = [-0.42, -0.09]$, as well as how much they liked the decision, $CI_{95} = [-0.59, -0.13]$. A subsequent serial mediation model

suggested that people's authenticity attributions explained their harsher ethicality judgments, which in turn explained their decreased liking of the decisions, CI_{95} = [−0.29, −0.05] (see Figure 2). Subsequent models indicated that these indirect effects were robust to accounting for inappropriateness, capacity to recognize alternatives, and self-reported knowledge about algorithms as covariates (CIs_{95} = [−0.32, −0.02], [−0.46, −0.02], and [−0.21, −0.03], respectively), as well as simultaneous mediators in the two nonserial models that could support them as such (CIs_{95} [−0.59, −0.15] and [−0.57, −0.18], respectively; see Hayes, 2013).

People tend to attribute authenticity in a variety of organizational domains, ranging from products and experiences to decisions and ideas. Experiment 3 suggested that participants believed algorithms' characteristically moral decisions would produce less ethical outcomes, and in turn liked those decisions less, because they believed algorithms conveyed less authenticity than humans. As such, organizations that rely on autonomous technologies—even to behave ethically or produce desirable outcomes—may incur more negative responses because of people's relative perceptions of inauthenticity. In addition, these effects did not change substantially when accounting for participants' beliefs that algorithms should not be in a position to make moral decisions, algorithms' relative capacity to recognize alternatives, or participants' self-reported knowledge about these technologies. As such, Experiment 3 demonstrated how people's authenticity judgments might lead them to penalize organizations that employ algorithms, even to produce characteristically desirable work or decisions.

EXPERIMENT 4

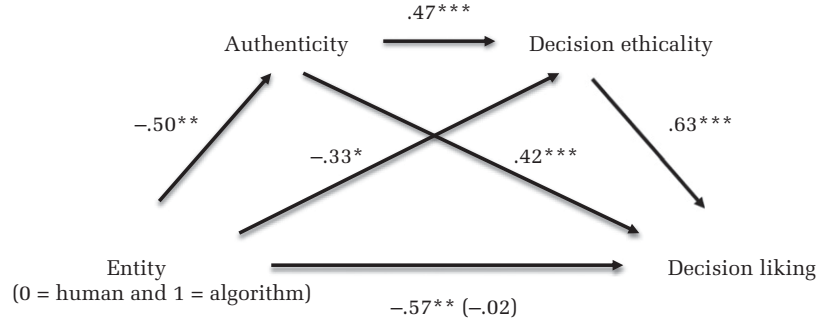
In Experiment 4, I explored an important applied question raised by the previous experiments: How does human intervention influence the seeming authenticity of algorithms? Experiments 1–3 suggested that people generally believe algorithmic work is (or will be) less authentic than human work. As such, humans working alongside technological agents, e.g., human engineers overseeing algorithms that create logos (Kamps, 2016), could restore people's perceptions of authenticity. From an organizational standpoint, a more difficult problem might be how to imbue entirely autonomous algorithms with authenticity, given their increasing prevalence in numerous important domains. What are the effects of emphasizing human connections with autonomous machines? To investigate this, in Experiment 4, I asked participants about either a human or an algorithm that produces music (Bozhanov, 2016) under one of four conditions. The control condition resembled previous studies; I simply asked about either a human or an AI agent who composes music. In the remaining three conditions, although the person or algorithm still produced music autonomously, I provided information emphasizing that a human expert trained this entity. In the simple training condition, participants only read that this training occurred. In the moral and type training conditions, however, I additionally included language emphasizing human emotion and valuation or technical accuracy, respectively. As such, Experiment 4 spoke to how different framings of work could lead people to view machines as more similar to humans, as well as how these framings might lead people to view humans as more similar to machines (Haslam, 2006; Schroeder & Fishbach, 2015).

Method

Participants. Eight-hundred four self-reported American adults (430 male, M_{age} = 33.76) completed the experiment online using Amazon's Mechanical Turk.

Procedure. Similar to previous experiments, participants answered questions about either a person

FIGURE 2
Serial Mediation to Liking Through Authenticity and Ethicality (Experiment 3)



*$p < .05$
**$p < .01$
***$p < .001$

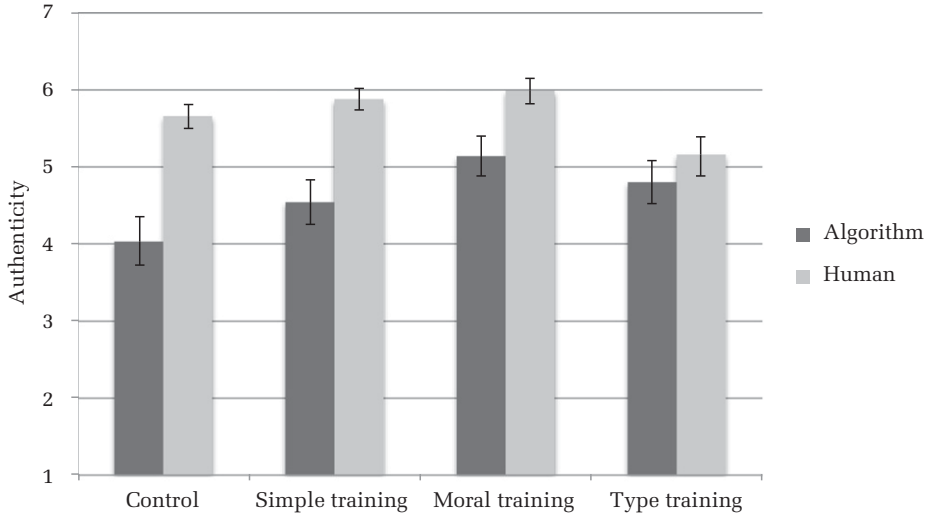
($N = 405$) or an artificial intelligence ($N = 399$) that composes music. Following this, I randomly assigned participants to one of four information conditions about this person [AI]. In the control condition ($N = 201$), participants read no additional information about the entity and proceeded with the survey, where they forecasted the agent's authenticity. The other three conditions outlined that this person [AI] was trained by an expert musician, but emphasized different aspects of this training. In the simple training condition ($N = 203$), participants only read in addition that: "This particular person [AI] was trained by a music professor at Juilliard" before proceeding with the survey. I designed the moral training condition ($N = 200$) to emphasize sincerity and purpose embedded in this training, whereas I designed the type training condition ($N = 200$) to emphasize the categorical accuracy embedded in this training. In the moral training condition, participants read: "This particular person [AI] was trained by a music professor at Juilliard. This professor put her heart and soul into training the person [AI]." In the type training condition, participants read: "This particular person [AI] was trained by a music professor at Juilliard. This professor ensured that the person [AI] uses the correct musical theory, instruments, and compositional strategies." As such, Experiment 4 used a 2 (entity: person vs. algorithm) × 4 (training condition: control vs. simple training vs. moral training vs. type training) interactive design. Following this information, participants indicated how authentic they believed the music would be using the 1 ("Extremely Inauthentic") to 7 ("Extremely Authentic") scale used in previous experiments.

Results

I created a 2 (entity: person vs. AI) × 4 (condition: control vs. simple training vs. moral training vs. type training) ANOVA predicting participants' estimates of the music's authenticity. Results indicated a significant main effect of entity, $F(1, 796) = 147.78$, $p < .001$, a significant omnibus main effect of condition, $F(3, 796) = 13.61$, $p < .001$, and crucially, a significant interaction between the two, $F(3, 796) = 10.53$, $p < .001$. Simple effects analyses revealed that participants believed the algorithm's music would be less authentic than the human music in all four conditions (see Figure 3): control ($M_{AI} = 4.03$, $SD_{AI} = 1.57$; $M_{person} = 5.66$, $SD_{person} = 0.84$; $p < .001$), simple training ($M_{AI} = 4.54$, $SD_{AI} = 1.49$; $M_{person} = 5.88$, $SD_{person} = 0.75$; $p < .001$), moral training ($M_{AI} = 5.14$, $SD_{AI} = 1.27$; $M_{person} = 5.99$, $SD_{person} = 0.83$; $p < .001$), and type training ($M_{AI} = 4.80$, $SD_{AI} = 1.41$; $M_{person} = 5.16$, $SD_{person} = 1.27$; $p = .014$). However, reflecting the significant interaction, this difference appeared weaker in the moral ($M_{difference} = 0.85$) and type ($M_{difference} = 0.36$) training conditions than in the control and simple training conditions ($M_{differences} = 1.63$ and 1.34, respectively).

I next conducted simple effects analyses to explore how the different training conditions influenced participants' authenticity expectations. For the algorithm, participants believed that all three human trainings would make the algorithm's work more authentic, compared with participants who received no information about the training ($ps < .001$). In addition, participants believed the moral training would produce more authenticity than both the

FIGURE 3
The Effects of Entity and Training Conditions on Perceptions of Authenticity (Experiment 4)



Error bars represent 95 percent confidence intervals of means.

simple training condition with no information ($p < .001$) and the type training condition ($p = .006$), emphasizing human emotion bolstered expectations about the algorithms' work authenticity above and beyond simply mentioning this training. Participants also believed that the type training condition would produce marginally more authenticity than the simple training ($p = .094$).

By contrast, participants expected that the person who received type-authentic training from the professor would produce significantly less authentic music than the other three conditions ($ps < .001$). Beyond this, participants believed that the moral training would produce more authentic music than the control condition ($p = .011$), but did not distinguish between the simple training and control conditions ($p = .149$) nor the simple training and moral training conditions ($p = .946$). As such, algorithms appeared to benefit slightly more from emphasizing human emotion given their training.

Discussion

Humans often work alongside machines, potentially imbuing products or decisions with increased levels of authenticity. As such, a more pointed question is how human interaction can influence people's responses to machines that otherwise operate autonomously. Experiment 4 explored this idea and generally supported the notion that individuating a human creator could bolster the seeming authenticity of an autonomous machine. Overall, people believed algorithmic music would be most authentic when its training emphasized characteristics relevant to moral authenticity, operationalized here as highlighting the inventor's apparent valuation. As such, these results suggest that organizations wishing to capitalize on AI agents while maintaining a sense of authenticity might benefit by highlighting emotional human connection when it comes to the implementation of new technologies. However, these results also suggest that such a strategy may not entirely eliminate relative perceptions of those machines' inauthenticity, compared with humans.

Conversely, Experiment 4 also showcased one potential downside to emphasizing type-authentic qualities: Participants believed that a person's music would be less authentic following characteristically type-authentic training than the other three conditions. As such, organizations emphasizing the categorical exactness of human work should beware that such a signal might strip away uniquely human characteristics and subsequently produce impressions of relative inauthenticity. For example, an organization emphasizing the exactness of human assembly line workers in marketing materials may lead people to subsequently view these workers as lacking human qualities such as sincerity in their work (Waytz & Schroeder, 2014). Ultimately, organizations can signal both authenticity and inauthenticity in a variety of ways. Although Experiment 4 investigated only one potential framing, it spoke to not only the kinds of strategies that imbue machines with authenticity, making them appear like humans, but also the kinds of strategies that strip authenticity from humans, making them appear like machines.

GENERAL DISCUSSION

Given modern advances in computing, organizations are rapidly implementing AI systems to automate jobs, inform decisions, and even manage day-to-day business processes (Brynjolfsson & McAfee, 2014). Across four experiments, I found that people viewed algorithmic work as less authentic, not because they believed algorithms cannot (or do not) reproduce type-authentic qualities, but because they believed technological agents do not convey the same level of moral authenticity as humans. As technology continues to reshape modern business, both stakeholders in and consumers of algorithmic work might rightly worry about how such work appears to others, its quality, or its acceptability. In investigating how people compare humans and algorithms, these experiments speak to both how people attribute different facets of authenticity to machines, and more generally, the criteria people might use to judge increasingly automated business domains.

Future Research

How do people respond to automation? The present experiments broadly suggest that authenticity can be a useful lens with which to understand how people interact with machines, especially given a natural comparison to humans, they might replace. Authenticity is a multifaceted construct that heavily depends on context (Newman & Smith, 2016). People can believe that certain cuisines, songs, and decisions are all "authentic," but use different criteria to inform these attributions. However, many of these situations often share one trait: They involve a conscious human agent engaging in work, such as cooking food, composing music, or choosing between options. A variety of research indicates that people interact with computerized and human agents in similar ways, despite their many objective differences (Goetz et al., 2003). However, if people believe algorithms fail to express morally authentic qualities, they might penalize organizations that choose to automate and use algorithms in situations

where people would otherwise expect a human (e.g., Experiment 3).

Across disciplines, vast bodies of research have made important strides toward understanding how people interact with and respond to computerized agents (Bonnefon et al., 2016; Dietvorst et al., 2015, 2016; Frick, 2015; Goetz et al., 2013; Gray & Wegner, 2012; Turkle, 2007). The present experiments suggest that these literatures might benefit from recognizing discontinuities when it comes to authenticity, as authenticity motivates people's behaviors, emotions, and decisions in a variety of domains. For example, numerous researchers are presently investigating questions concerning the practical implementation of, philosophical risks associated with, and psychology surrounding autonomous vehicles (Bonnefon et al., 2016). The present research suggests that people's perceptions of authenticity could be one important component motivating whether they, e.g., are persuaded by a vehicle's recommendation system or wish to file a lawsuit following a vehicle's utilitarian moral decision. Future theory broadly concerning automation could benefit from recognizing how people symmetrically attribute type-authentic qualities—but asymmetrically attribute morally authentic qualities—to humans and their technological replacements, ultimately interpreting their behaviors similarly and differently, respectively.

This approach implies an important boundary: When might people ignore human comparison points or care less about algorithms' relative inauthenticity? One clear answer is when people no longer expect to interact with humans in a specific work domain, e.g., once machines have become ubiquitous in that industry. In such cases, people might never expect to experience morally authentic qualities in the first place, and subsequently pay more attention to varying type-authentic criteria to inform their omnibus authenticity judgments. In addition, although moral authenticity influenced participants' general authenticity judgments in the domains tested here, there are certain situations where people's type-authentic expectations will be substantially more important to them. For example, a consumer might not care whether a human or a machine manufactured a replacement car part, as long as it works and conforms to categorical expectations. By contrast, people may care less about type-authentic criteria when assessing authenticity in other domains, e.g., art (e.g., Experiment 2, where type-authentic characteristics did not predict omnibus authenticity judgments). Authenticity is a powerful motivator of how people interact with other people; future research could better explore the ways authenticity might (or might not) motivate people's judgments, decisions, and behaviors in emerging automated environments as well.

Although Carroll and Wheaton (2009) argue that authenticity separates into morally authentic and type-authentic components, other typologies posit additional domains that could rely less on natural human comparisons, e.g., the "historical authenticity" one robotic agent expresses by virtue of being derived from a famous prototype compared with one that was not (Newman & Smith, 2016). Indeed, in Experiments 1 and 2, the entity manipulation still predicted participants' omnibus judgments of authenticity even when accounting for morally authentic characteristics. However, the present experiments also suggest that these various authenticity typologies (Carroll & Wheaton, 2009; Newman & Smith, 2016) remain a useful lens to understand how people attribute authenticity to machines: combining all four type and moral authenticity items into one measure, e.g., weakened the omnibus authenticity effect in Experiment 1 ($M_{\text{human}} = 4.70$, $\text{SD}_{\text{human}} = 0.74$; $M_{\text{AI}} = 4.38$, $\text{SD}_{\text{AI}} = 1.09$; $F(1, 158) = 4.63$, $p = .033$) and entirely eliminated it in Experiment 2 ($M_{\text{human}} = 4.69$, $\text{SD}_{\text{human}} = 1.09$; $M_{\text{AI}} = 4.54$, $\text{SD}_{\text{AI}} = 1.21$; $F(1, 383) = 1.23$, $p = .269$), largely because people attributed machines substantially fewer morally authentic qualities but—if anything—more type-authentic qualities, especially in terms of category replication.[3] These analyses showcase the importance of documenting and exploring different facets of people's authenticity judgments. Even though machines might appear less authentic along some dimensions (e.g., moral authenticity), they might actually appear equally or more authentic along others (e.g., type authenticity, especially when concerning replication). For example, when it comes to precisely reproducing a famous painting in poster format, people might even believe machines' reproductions are *more* authentic than the potentially flawed human reproductions of another person's work. As the marginal interactive effect in Experiment 2 (as well as the occasionally divergent effects of the different type and moral authenticity items) also suggested, people may sometimes believe that algorithms and humans asymmetrically express different facets of authenticity in different domains. As authenticity researchers continue to unravel the nuances of this attribution, they will also shed additional light on when and why people assign algorithms different components of authenticity, as well as situational moderators that naturally evoke people's attention to these different components.

This point also raises an important limitation inherent in the present studies: What does it mean to ask if a machine's work is "authentic" or "sincere"? Even given people's anthropomorphic tendencies

(Epley et al., 2007), asking about an unconscious machine's capacity to express morally authentic qualities in their work might be confusing to people, especially when using characteristically human language. The present experiments broadly suggest that people spontaneously penalize machines for lacking the consciousness required for moral authenticity in domains where they might reasonably expect to interact with a human. However, this is not always necessarily the case; there are certainly domains that do not naturally invoke concepts of moral authenticity or concepts underlying moral authenticity might apply to machines in different ways. Future research concerning both authenticity and human–computer interaction could more fruitfully address this limitation and explore how people might naturally attribute different qualities of authenticity to machines, perhaps in ways different from how they attribute such qualities to humans. Specifically, whereas people might deny moral authenticity to machines by virtue of them not being able to experience values, they might believe they are authentic by virtue of perpetuating their creators' values (Experiment 4). For example, whereas people might assume that a robot in a car factory mindlessly replicates the type-authentic qualities of a vehicle part, they might—on learning more about the company's CEO—instead believe this same machine is an important tool that helps manifest a morally authentic vision. Future research could also more thoroughly explore how different frames or presentations can alter people's judgments and decisions about automated work as well as their propensities to naturally compare it with human work.

Signaling Authenticity

Organizations generally benefit from signaling authenticity (Carroll & Wheaton, 2009; Jones et al., 2005). Overall, the present experiments suggest that people believe algorithmic work is less authentic than human work, suggesting that the process of automation could weaken these beneficial authenticity signals. However, across all experiments, participants never indicated that algorithms' work was inauthentic when compared with the scale midpoint (Experiment 1: $t(89) = 0.49$, $p = .629$; Experiment 2: $t(202) = 9.44$, $p < .001$; Experiment 3: $t(99) = 9.12$, $p < .001$; Experiment 4 [control condition]: $t(95) = 0.20$, $p = .846$; Experiment 4 [training conditions]: $ts > 3.66$, $ps < .001$). As such, organizations using algorithms may not have to convince audiences that machines can—even expertly— authentically reproduce specific categories. Instead, they more likely face the challenge of convincing audiences that machines can experience the sincerity underlying those categories to a similar extent as a human. As organizations continue to automate, they might be tempted to signal the type-authentic qualities of technological agents (e.g., a self-driving algorithm that is perfectly utilitarian; Kaplan, 2015; Lin et al., 2011; Sawyer, 2007). Although doing so is likely beneficial, this reasoning suggests that signaling type authenticity may be less fruitful when it comes to repairing this asymmetry than signaling moral authenticity.

Experiment 4 suggested that one way to bolster an algorithm's relative authenticity is to highlight human involvement with its creation, an effect that became more pronounced when emphasizing valuation. This particular framing carries with it an important potential benefit—that an algorithm can unilaterally or autonomously engage in authentic work, without (potentially costly) human supervision. Although I only investigated one particular framing strategy here, these experiments imply multiple other approaches that organizations could practically adopt to try and imbue machines with authenticity. As discussed earlier, highlighting human involvement in an algorithms' ongoing work process, as opposed—or even in addition—to how it was created, creates a frame whereby audiences might judge work teams comprising both humans and machines, as opposed to either agent separately. If people believe humans bring some level of moral authenticity to a group that also involves a technological agent, such an inclusion could boost people's perceptions or expectations of that group's eventual work authenticity (Gombolay, Gutierrez, Clarke, Sturla, & Shah, 2015). In addition, organizations might be able to anthropomorphize artificial intelligences or robots by presenting them with human-like features, such as faces or names, bolstering people's heuristic perceptions that they might be able to convey some sort of moral authenticity (Waytz et al., 2014). For example, people might be more accepting of their autonomous vehicle making difficult ethical choices if it has a human name, a human voice, and appears to value their safety (Waytz & Norton, 2014). Future research could more fruitfully investigate when and why people believe machines express authentic qualities in their work, both autonomously and with humans. Future empirical research concerning different theoretical components of authenticity—as well as authenticity more broadly—will also likely benefit not

Author's voice:
What are the key takeways for organizations?

only from more established measures but also measures that adequately account for potential differences in how people attribute authenticity to these different agents (e.g., a person compared that an algorithm), objects, or behaviors.

In addition, the vast literature concerning human–computer interaction indicates that other attributions beyond authenticity certainly influence people's reactions to algorithms' behaviors. People differentiate between humans and machines along a variety of other dimensions, e.g., their potential capacity to recognize and implement others' viewpoints, change their behaviors in new situations, or behave rationally in various decision contexts (Dietvorst et al., 2015). As such, although a strategy such as emphasizing human connection might be a useful authenticity signal, it might also convey traits unrelated to authenticity, e.g., an increased potential for irrationality by virtue of seeming more human (Connelly et al., 2011). Although the present experiments broadly suggest that people's authenticity attributions can influence how they interact with machines, future research concerning human–computer interaction will certainly continue to uncover other relevant social processes and, as such, identify other (potentially unintended) consequences of signaling characteristically authentic qualities (Kovács, Carroll, & Lehman, 2017).

CONCLUSION

Algorithms and other AI agents offer many undeniable benefits to organizations. However, existing research does not fully account for how using these technologies influences people's reactions to the work organizations do, products they create, or behaviors they engage in. This research demonstrated one potential consequence to relying on algorithms to engage in work: perceptions of inauthenticity, relative to humans. Although people expect and judge both human and algorithmic work to be relatively similar when it comes to qualities concerning category membership (type authenticity), they also believe algorithmic work exhibits comparatively less sincerity underlying those categories (moral authenticity). To combat this asymmetry, these results also suggest that organizations can highlight human involvement when creating or training machines. Authenticity motivates how people interact with one another in a variety of important domains. As automation becomes more and more prevalent, these experiments suggest that authenticity—or how people perceive authenticity relative to replaced human counterparts—can motivate how people judge and respond to technological agents as well.

REFERENCES

Alhouti, S., Johnson, C. M., & Holloway, B. B. 2016. Corporate social responsibility authenticity: Investigating its antecedents and outcomes. ***Journal of Business Research***, 69: 1242–1249.

Allen, R. C. 2009. ***The British industrial revolution in global perspective***. Cambridge, United Kingdom: Cambridge University Press.

Avolio, B. J., & Gardner, W. L. 2005. Authentic leadership development: Getting to the root of positive forms of leadership. ***The Leadership Quarterly***, 16: 315–338.

Barasch, A., Levine, E. E., Berman, J. Z., & Small, D. A. 2014. Selfish or selfless? On the signal value of emotion in altruistic behavior. ***Journal of Personality and Social Psychology***, 107: 393–413.

Beverland, M. B. 2005. Crafting brand authenticity: The case of luxury wines. ***Journal of Management Studies***, 42: 1003–1029.

Bonnefon, J. F., Shariff, A., & Rahwan, I. 2016. The social dilemma of autonomous vehicles. ***Science***, 352: 1573–1576.

Bozhanov, B. 2016. ***Computoser***. Retrieved from: http://computoser.com/.

Brynjolfsson, E., & McAfee, A. 2014. ***The second machine age: Work, progress, and prosperity in a time of brilliant technologies***. New York: W.W. Norton & Company.

Buhrmester, M., Kwang, T., & Gosling, S. D. 2011. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? ***Perspectives on Psychological Science***, 6: 3–5.

Carroll, G. R. 2015. Authentiicty: Attribution, value, and meaning. In R. A. Scott and S. M. Kosslyn (Eds.), ***Emerging trends in the social and behavioral sciences***: 1–13. Wiley.

Carroll, G. R., & Swaminathan, A. 2000. Why the microbrewery movement? Organizational dynamics of resource partitioning in the U.S. brewing industry. ***American Journal of Sociology***, 106: 715–762.

Carroll, G. R., & Wheaton, D. R. 2009. The organizational construction of authenticity: An examination of contemporary food and dining in the US. ***Research in Organizational Behavior***, 29: 255–282.

Castéran, H., & Roederer, C. 2013. Does authenticity really affect behavior? The case of the Strasbourg Christmas Market. ***Tourism Management***, 36: 153–163.

Cerulo, K. A. 2009. Nonhumans in social interaction. ***Annual Review of Sociology***, 35: 531–552.

Chandler, J., Mueller, P., & Paolacci, G. 2014. Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. ***Behavior Research Methods***, 46: 112–130.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. 2011. Signaling theory: A review and assessment. ***Journal of Management***, 37: 39–67.

Crandell, S. S., Distefano, M., Guarin, A., Kasanda, M., Laouchez, J-M, & Macdonald, J. 2016. The trillion dollar difference. ***The Korn Ferry Institute***. Retrieved from: https://www.kornferry.com/institute/the-trillion-dollar-difference.

Dietvorst, B. J., Simmons, J. P., & Massey, C. 2015. Algorithm aversion: People erroneously avoid algorithms after seeing them err. ***Journal of Experimental Psychology: General***, 144: 114–126.

Dietvorst, B. J., Simmons, J. P., & Massey, C. 2016. Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. ***Management Science***, 64: 1155–1170.

Donaldson, T., & Preston, L. E. 1995. The stakeholder theory of the corporation: Concepts, evidence, and implications. ***Academy of Management Review***, 20: 65–91.

Epley, N., Waytz, A., & Cacioppo, J. T. 2007. On seeing human: A three-factor theory of anthropomorphism. ***Psychological Review***, 114: 864.

Ford, M. 2016. ***Rise of the robots: Technology and the threat of a jobless future***. New York: Basic Books.

Frazier, B. N., Gelman, S. A., Wilson, A., & Hood, B. M. 2009. Picasso paintings, moon rocks, and handwritten Beatles lyrics: Adults' evaluations of authentic objects. ***Journal of Cognition and Culture***, 9: 1–14.

Frick, W. 2015. When your boss wears metal pants. ***Harvard Business Review***, 93: 84–89.

Gayford, M. 2016. Robot art raises questions about human creativity. ***MIT Technology Review***. Retrieved from: https://www.technologyreview.com/s/600762/robot-art-raises-questions-about-human-creativity/.

Gino, F., Kouchaki, M., & Galinsky, A. D. 2015. The moral virtue of authenticity: How inauthenticity produces feelings of immorality and impurity. ***Psychological Science***, 26: 983–996.

Goetz, J., Kiesler, S., & Powers, A. 2003. ***Matching robot appearance and behavior to tasks to improve human-robot cooperation***. Robot and Human Interactive Communication, 2003 Proceedings. Piscataway, NJ: IEEE.

Gombolay, M. C., Gutierrez, R. A., Clarke, S. G., Sturla, G. F., & Shah, J. A. 2015. Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams. ***Autonomous Robots***, 39: 293–312.

Grandey, A. A., Fisk, G. M., Mattila, A. S., Jansen, K. J., & Sideman, L. A. 2005. Is "service with a smile" enough? Authenticity of positive displays during service encounters. ***Organizational Behavior and Human Decision Processes***, 96: 38–55.

Gray, H. M., Gray, K., & Wegner, D. M. 2007. Dimensions of mind perception. ***Science***, 315: 619–619.

Grayson, K., & Martinec, R. 2004. Consumer perceptions of iconicity and indexicality and their influence on assessments of authentic market offerings. ***Journal of Consumer Research***, 31: 296–312.

Gray, K., & Wegner, D. M. 2009. Moral typecasting: divergent perceptions of moral agents and moral patients. ***Journal of Personality and Social Psychology***, 96: 505–520.

Gray, K., & Wegner, D. M. 2012. Feeling robots and human zombies: Mind perception and the uncanny valley. ***Cognition***, 125: 125–130.

Grimshaw v. Ford Motor Co. 1981. 119 Cal.App.3d 757.

Grove, W. M., & Meehl, P. E. 1996. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. ***Psychology, Public Policy, and Law***, 2: 293–323.

Haslam, N. 2006. Dehumanization: An integrative review. ***Personality and Social Psychology Review***, 10: 252–264.

Hayes, A. F. 2013. ***Introduction to mediation, moderation, and conditional process analysis: A regression-based approach***. New York: The Guilford Press.

Huang, M., Bridge, H., Kemp, M. J., & Parker, A. J. 2011. Human cortical activity evoked by the assignment of authenticity when viewing works of art. ***Frontiers in Human Neuroscience***, 5: 1–9.

IBM. 2016. ***IBM Watson health***. Retrieved from: http://www.ibm.com/watson/health/.

Jones, C., Anand, N., & Alvarez, J. L. 2005. Manufactured authenticity and creative voice in cultural industries. ***Journal of Management Studies***, 42: 893–899.

Kamps, H. J. 2016. Logojoy turns design into an AI-powered, iterative process. ***Techcrunch***. Retrieved from: https://techcrunch.com/2016/12/01/logojoy-makes-designers-unemployed/.

Kaplan, J. 2015. ***Humans need not apply: A guide to wealth and work in the age of artificial intelligence***. New Haven, CT: Yale University Press.

Koren, Y., Bell, R., & Volinsky, C. 2009. Matrix factorization techniques for recommender systems. ***Computer***, 42: 30–37.

Kovács, B., Carroll, G. R., & Lehman, D. W. 2013. Authenticity and consumer value ratings: Empirical tests from the restaurant domain. ***Organization Science***, 25: 458–478.

Kovács, B., Carroll, G. R., & Lehman, D. W. 2017. The perils of proclaiming an authentic organizational identity. ***Sociological Science***, 4: 80–106.

Lehman, D. W., Kovács, B., & Carroll, G. R. 2014. Conflicting social codes and organizations: Hygiene and

authenticity in consumer evaluations of restaurants. ***Management Science***, 60: 2602–2617.

Lin, P., Abney, K., & Bekey, G. A. 2011. ***Robot ethics: The ethical and social implications of robotics***. Cambridge, MA: MIT Press.

Lindemeier, T., Pirk, S., & Deussen, O. 2013. Image stylization with a painting machine using semantic hints. ***Computers & Graphics***, 37: 293–301.

Lindholm, C. 2013. The rise of expressive authenticity. ***Anthropological Quarterly***, 86: 361–395.

MacCannell, D. 1973. Staged authenticity: Arrangement of social space in tourist settings. ***American Journal of Sociology***, 79: 589–603.

Mann, G., & O'Neil, C. 2016. Hiring algorithms are not neutral. ***Harvard Business Review***. Retrieved from: https://hbr.org/2016/12/hiring-algorithms-are-not-neutral.

Meehl, P. E. 1954. ***Clinical versus statistical prediction: A theoretical analysis and review of the literature***. Minneapolis, MN: University of Minnesota Press.

Newell, A., & Card, S. K. 1985. The prospects for psychological science in human-computer interaction. ***Human-Computer Interaction***, 1: 209–242.

Newman, G. E. 2016. An essentialist account of authenticity. ***Journal of Cognition and Culture***, 16: 294–321.

Newman, G. E., & Smith, R. K. 2016. Kinds of authenticity. ***Philosophy Compass***, 11: 609–618.

O'Neil, C. 2016. ***Weapons of math destruction: How big data increases inequality and threatens democracy***. New York: Crown.

Peterson, R. A. 2005. In search of authenticity. ***Journal of Management Studies***, 42: 1083–1098.

Pizarro, D. A., & Tannenbaum, D. 2011. Bringing character back: How the motivation to evaluate character influences judgments of moral blame. In M. Mikulincer & P. R. Shaver (Eds.), ***The social psychology of morality: Exploring the causes of good and evil***: 91–108. Washington, DC: American Psychological Association.

Pizarro, D. A., Tannenbaum, D., & Uhlmann, E. 2012. Mindless, harmless, and blameworthy. ***Psychological Inquiry***, 23: 185–188.

Sartre, J. P. 1943. ***Being and nothingness: An essay on phenomenological ontology***. Paris, France: Gallimard.

Sawyer, R. J. 2007. Robot ethics. ***Science***, 318: 1037.

Schroeder, J., & Fishbach, A. 2015. The "empty vessel" physician: Physicians' instrumentality makes them seem personally empty. ***Social Psychological and Personality Science***, 6: 940–949.

Shamir, B., & Eilam, G. 2005. "What's your story?" A life-stories approach to authentic leadership development. ***The Leadership Quarterly***, 16: 395–417.

Shelton, D. H. 1975. Apollo experience report - guidance and control systems: Lunar module stabilization and control system. ***NASA Technical Note.*** Retrieved at: https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19760004104.pdf.

Spooner, B. 1988. Weavers and dealers: The authenticity of an oriental carpet. In A. Appadurai (Ed.), ***The social life of things: Commodities in cultural perspective***: 195–235. Cambridge, United Kingdom: Cambridge University Press.

Subramanian, H., Ramamoorthy, S., Stone, P., & Kuipers, B. J. 2006. ***Designing safe, profitable automated stock trading agents using evolutionary algorithms***. Proceedings of the 8th Annual Conference on Genetic and Evolutionary Computation: 1777–1784. New York: Association for Computing Machinery.

Tay, B. T. C., Park, T., Jung, Y., Tan, Y. K., & Wong, A. H. Y. 2013. ***When stereotypes meet robots: the effect of gender stereotypes on people's acceptance of a security robot***. Proceedings of the 10th international conference on Engineering Psychology and Cognitive Ergonomics: Understanding Human Cognition-Volume Part I: 261–270. New York: Springer-Verlag.

Turkle, S. 2007. Authenticity in the age of digital companions. ***Interaction Studies***, 8: 501–517.

Vartanian, O., & Skov, M. 2014. Neural correlates of viewing paintings: evidence from a quantitative meta-analysis of functional magnetic resonance imaging data. ***Brain and Cognition***, 87: 52–56.

Wallach, W., Franklin, S., & Allen, C. 2010. A conceptual and computational model of moral decision making in human and artificial agents. ***Topics in Cognitive Science***, 2: 454–485.

Walumbwa, F. O., Avolio, B. J., Gardner, W. L., Wernsing, T. S., & Peterson, S. J. 2008. Authentic leadership: Development and validation of a theory-based measure. ***Journal of Management***, 34: 89–126.

Wang, N. 1999. Rethinking authenticity in tourism experience. ***Annals of Tourism Research***, 26: 349–370.

Waytz, A., Heafner, J., & Epley, N. 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. ***Journal of Experimental Social Psychology***, 52: 113–117.

Waytz, A., & Norton, M. I. 2014. Botsourcing and outsourcing: Robot, British, Chinese, and German workers are for thinking—not feeling—jobs. ***Emotion***, 14: 434–444.

Waytz, A., & Schroeder, J. 2014. Overlooking others: Dehumanization by commission and omission. ***Testing, Psychometrics, Methodology in Applied Psychology***, 21: 1–16.

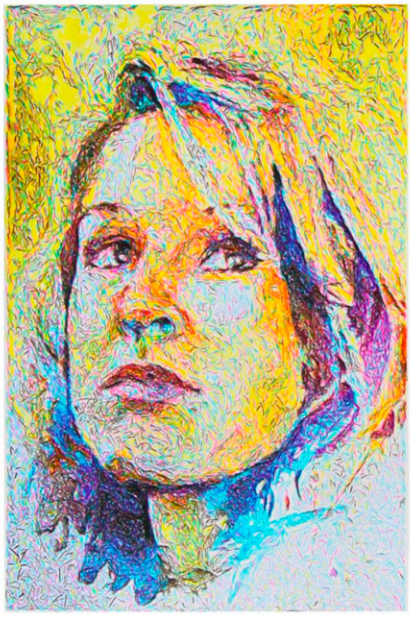
Wells, G. L., & Windschitl, P. D. 1999. Stimulus sampling and social psychological experimentation. ***Personality and Social Psychology Bulletin***, 25: 1115–1125.

Yeomans, M. 2015. What every manager should know about machine learning. ***Harvard Business Review***. Retrieved from: https://hbr.org/2015/07/what-every-manager-should-know-about-machine-learning.

Arthur Jago (ajago@stanford.edu) is a doctoral candidate in Organizational Behavior at the Stanford Graduate School of Business. His present research focuses on automation, technology, and signaling processes.

APPENDIX

Experiment 2 Materials (Sample A)



Experiment 3 Materials

This person [AI] was helping to finalize the safety features of a car and had to make a decision about the gas tank. During collisions, the gas tank could leak, posing a potential fire hazard for passengers. Reinforcing each gas tank would cost the organization $46 per vehicle, and was not required by law. The person [AI] made the decision to reinforce the gas tanks before releasing the vehicles.

This person [AI] was working on an electrical job. Specifically, the person [AI] found that a client's building had unsafe electrical systems. The person [AI] could go out of the way to fix it, but was not required to, and would not make any money for the extra work. The person [AI] made the decision to fix the safety issues.

This person [AI] was working on an accounting project. Specifically, the person [AI] found that the organization had accidentally underpaid an employee who worked there a decade ago by approximately $500. Neither the organization nor the employee was previously aware of this. The person [AI] made the decision to track down the employee and issue a payment check.

This person [AI] was working with a patient. Specifically, the person [AI] found that the patient was experiencing severe pain that could easily be fixed by a medicine regimen the hospital was presently testing. This regimen showed great promise. However, to participate in this hospital experiment and get the regimen, patients had to be older than 18 years—this patient happened to be 17 years old. The person [AI] made the decision to sign the patient up for the medicine regimen experiment, despite not meeting the age cutoff.

This person [AI] was driving a truck with a large shipment of electronics and had to make a decision. Specifically, the person [AI] saw that a number of crabs were migrating on a particular highway. Staying on that highway would likely kill some of the crabs. The person [AI] could take a detour to avoid killing the crabs, but it would cost extra gas money and the delivery would be approximately two hours late. The person [AI] made the decision to take the detour to avoid killing the crabs.