


2015

Social Media and the Organization Man

Dylan E. Wittkower

Old Dominion University, dwittkow@odu.edu

Follow this and additional works at: https://digitalcommons.odu.edu/philosophy_fac_pubs

 Part of the [Organizational Behavior and Theory Commons](#), and the [Social Media Commons](#)

Repository Citation

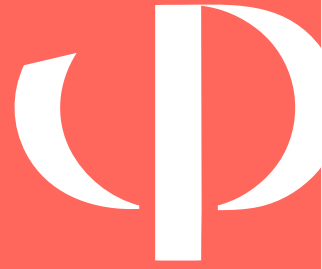
Wittkower, Dylan E., "Social Media and the Organization Man" (2015). *Philosophy Faculty Publications*. 15.
https://digitalcommons.odu.edu/philosophy_fac_pubs/15

Original Publication Citation

Wittkower, D.E. (2015). Social media and the organization man. *APA Newsletter on Philosophy and Computers*, 14(2), 16-20.

This Article is brought to you for free and open access by the Philosophy & Religious Studies at ODU Digital Commons. It has been accepted for inclusion in Philosophy Faculty Publications by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.

Philosophy and Computers



SPRING 2015

VOLUME 14 | NUMBER 2

FROM THE GUEST EDITOR

John P. Sullins

NOTES FROM THE COMMUNITY ON PAT SUPPES

ARTICLES

Patrick Suppes

Patrick Suppes Autobiography

Luciano Floridi

Singularitarians, Altheists, and Why the Problem with Artificial Intelligence is H.A.L. (Humanity At Large), not HAL

Peter Boltuc

First-Person Consciousness as Hardware

D. E. Wittkower

Social Media and the Organization Man

Niklas Toivakainen

The Moral Roots of Conceptual Confusion in Artificial Intelligence Research

Xiaohong Wang, Jian Wang, Kun Zhao, and Chaolin Wang

Increase or Decrease of Entropy: To Construct a More Universal Macroethics (A Discussion of Luciano Floridi's The Ethics of Information)

Social Media and the Organization Man

D. E. Wittkower
 OLD DOMINION UNIVERSITY

In an age of social media we are confronted with a problem novel in degree if not in kind: being called to account for the differences between presentations of self appropriate within a variety of group contexts. Business news in the post-Facebook era has been replete with stories about privacy fails, large and small—employees fired or denied promotion seemingly due to same-sex relationships revealed on social media, career advice to college students about destroying online evidence of having done normal college-student things, and so on. Keeping work and private lives separate has become more difficult, and difficult in different ways, and we are living in a new era of navigating self- and group-identities.

While social media in general tends to create these problems, Facebook, with its unitary profile, single Friend list, and real-name policy, has been central to creating this new hazardous environment for identity performance. Mark Zuckerberg is quoted in an interview with David Kirkpatrick saying, “You have one identity. . . . The days of you having a different image for your work friends or co-workers and for the other people you know are probably coming to an end pretty quickly. . . . Having two identities for yourself is an example of a lack of integrity.”¹ Many have critiqued this simplistic view of identity, but Michael Zimmer’s widely read blog post on the topic is particularly pithy and direct:

Zuckerberg must have skipped that class where Jung and Goffman were discussed. Individuals are constantly managing and restricting flows of information based on the context they are in, switching between identities and persona. I present myself differently when I’m lecturing in the classroom compared to when I’m having a beer with friends. I might present a slightly different identity when I’m at a church meeting compared to when I’m at a football game. This is how we navigate the multiple and increasingly complex spheres of our lives. It is not that you pretend to be someone that you are not; rather, you turn the volume up on some aspects of your identity, and tone down others, all based on the particular context you find yourself.²

And this view of the complexity of managing self-presentations within different organizational contexts, destructive as it already is to Zuckerberg’s—well, it’s hard to say . . . simplistic? Naïve? Unrealistic? Hetero- and Cis-privileged? Judgmental? All of these, I suppose—at any rate, to Zuckerberg’s *faulty* view of multiple identities as “a lack of integrity,” this view doesn’t even yet consider that different elements of identity may need to be not merely emphasized or toned down in different contexts, but that integral aspects of identity may need to be hidden entirely in some contexts and revealed only in others. Zimmer is aware of this too, and quotes an appropriately pseudonymous comment on Kieran Healy’s blog post on

the topic, that “Nobody puts their membership in Alcoholics Anonymous on their CV.”³ Surely we ought to say that if anything demonstrates integrity, it would be admitting a difficult truth about oneself and seeking support with others through a frank relationship of self-disclosure, making the AA example particularly apt, not least since the “anonymous” part of AA recognizes that this sort of integrity requires a safe separation of this organizational identity from other aspects of one’s life, of which the contents of a CV is only one particular example, dramatic in its absurdity.

Zuckerberg, for his part, seems to have started to think differently about this, stating in a 2014 interview that

I don’t know if the balance has swung too far, but I definitely think we’re at the point where we don’t need to keep on only doing real identity things [. . .] If you’re always under the pressure of real identity, I think that is somewhat of a burden.⁴

The 2010 comments are still important for us to take seriously, though. Not so much because Zuckerberg’s comments reveal a design trait in the Facebook platform that has changed how we think about and perform identity (although this is interesting as well!). But even more so because if Zuckerberg, mired as he is in thinking about how people manage self- and group identities, can fall into a way of thinking so disconnected from the actual conduct of lives, there must be something deeply intuitive, perhaps seductive, about this way of thinking about integrity.

At the heart of this intuition is a modern individualist notion of the self—the self which rights-bearing, with an individual and separable existence; the juridical self. We must assume an integral self logically prior to organizational and communal entanglement in order to pass judgment on whether it is limited, transformed, disfigured, hidden, or altered by its entrance into and representation within groups and contexts. We tend to take on a “correspondence theory” of integrity, parallel to the correspondence theory of truth, in which a self-representation is to have greater or lesser integrity depending upon the degree of similarity that it bears to some *a priori* “true” self. This view of an “unencumbered self” is deeply mistaken as Sandel (1984) among others has pointed out, but is logistically central to our liberal individualist conception of rights and community and thus hard to avoid falling into. Zuckerberg may do well to read philosophy in addition to the remedial Goffman (1959) to which Zimmer rightly wishes to assign him.

INTEGRITY AND SELF-PERFORMANCE

Turning to philosophical theories of personal identity seems at first unhelpful. Whether, for example, we adopt a body-continuity or mind-continuity theory of identity has only the slightest relevance to what might count as “integrity”—in fact, it seems any perspective on philosophical personal identity must view “integrity” as either non-optional or impossible; more a metaphysical state than a moral value. But even within, e.g., the Humean view that the self is no more than a theater stage on which impressions appear in succession⁵ fails to preclude that there may be some integral self—Hume’s claim applies only to the self as revealed by introspection, as Kant pointed out in arguing

for the idealism of the transcendental unity of apperception (1998); a grammatical necessity, as it were, corresponding in unknowable ways to the noumenal reality, which, however, is not necessarily less real for its unknowability. Indeed, when we look to Hume's (2012) theory of moral virtue, we see it is based upon sentiment and sympathy rather than following moral rules or calculation, implying that we have these acquired and habitual attributes which constitute our moral selves, even if they are not the "I" of the "I think" which accompanies all representations. Even reductive and skeptical positions within philosophical theories of personal identity make room for habit, character, and some sort of content to the self, inaccessible through introspection though it might be, which is subject to change and growth, and which is if not an origin then at least a conditioning factor in the determination of our thought and action.

We could do worse than to turn to Aristotle for an account of this.⁶ An Aristotelian view of character has the significant virtue of viewing identity as both real and consequential as well as also being an object of work. We have on his view a determinate character—e.g., we may, in fact, be a coward. But in this view we still need not fall into Sartrean bad faith, for a coward need not be a coward in the sense that Sartre's waiter is a waiter.⁷ A coward may be a coward, but may nevertheless be brave in this or that particular situation—and through an accretion of such instances of bravery, may become brave rather than cowardly. Aristotle, along with AA, tells us to "fake it 'til you make it," and both rightly view this "faking it" as a creation of integrity, not a mere demonstration of its absence.

On a correspondence theory of integrity, this self-conscious performance of a character which we do not possess appears as false representation, but this makes sense only when we assume a complete, settled, and coherent character. We say someone is "acting with integrity" when she takes an action in accordance with her values and principles, even or especially when it goes against her self-interest. Integrity, then, is not a degree of correspondence between character and behavior, but between values and behavior. One can even act with integrity by going against one's character, as in the case of the coward who nonetheless stands up for what she believes in a dangerous situation; the alcoholic entering recovery who affirms "I am intemperate" and concludes "therefore I will *not* drink."⁸

The sort of identity relevant to integrity, then, is not personal identity in a philosophical sense (for the mere unity of apperception is not a thing to which I can stay true), nor is it one's actual character or habits (for to reduce oneself to one's history and habits is bad faith, and acting according to our habits could well lead us away from integrity if our habits are vicious). Instead, the relevant sort of identity must be that with which we *identify*. Certainly we can recognize that we have traits with which we do not identify, and the process of personal growth is the process of changing our character in order to bring it into accordance with the values we identify with. As Suler has argued, disinhibition does not necessarily reveal some "truer self" that lies "underneath" inhibitions; disinhibition may instead make us unrecognized to ourselves.⁹ Our inhibitions—at the least the ones we value; which we *identify with*—are part of

the self that we recognize as ourselves, and inhibitions may themselves be the product of choice and work.

INTEGRITY IN AN ORGANIZATIONAL CONTEXT

We need not fall into a correspondence theory of integrity or adopt a liberal individualist conception of the self in order to recognize that organizational contexts present problems for personal integrity. Two primary sorts come immediately to mind: (1) that organizational contexts may exert influences rendering it more difficult to act with integrity, as in familiar cases such as conformity and groupthink; and (2) that organizational contexts may contain hostility towards certain self-identifications, making self-performance with integrity dangerous. The second kind of problem is the sort most obviously presented by social media in novel ways, and will be our focus here, but by the end of this chapter we'll have some insights on the first as well.

Conflicts between aspects of self-identity in different contexts certainly do not arise for the first time with social media, and are not limited to identities which are discriminated against. One does not, for the most part, discuss one's sex life in church, even if that sex life takes place within marriage—and within a straight marriage, and involves "vanilla sex" rather than BDSM, and so on. And yet it is not without reason that recent years have seen renewed and intensified discussion of managing boundaries between personal and professional life, and the tendency of social media to either blur or overlap contexts of identity performance has created a new environment of identity performance causing new requirements for thinking about and managing identities.¹⁰

In contemporary digital environments, we are frequently interacting simultaneously with persons from different personal and social contexts. Our friends and followers in social networking sites (SNS) are promiscuously intermixed. We have only a single profile in each, and we cannot choose which profile items—gender identity, religious identity, former employers, name—are viewable to which connections or groups of connections in our network. Nor can we choose to have different presentations for different connections or groups: we may portray ourselves differently in social or work contexts, but can choose only a single profile picture. There are work-arounds, of course, but they are onerous, difficult to maintain, and sometimes violate terms of service agreements requiring single accounts and real names. Even using built-in affordances intended to aid in maintaining contextual integrity,¹¹ such as private accounts (Twitter), friend lists (Facebook), or circles (Google+), is difficult and socially risky: difficult because managing such affordances requires significant upkeep, curation, memory, and attention; risky because members of groups of which we are members tend to have their own separate interconnections, online or off, and effective boundary enforcement must include knowledge of these interconnections and accurate prediction of information flows across them. If you wish to convince your parents that you've quit Facebook, how far out in their social networks must you go in excluding friends from viewing your posts? Aunts and uncles? Family friends? Friends of friends of family? Or, in maintaining separation of work and personal life, how are you to know whether a Facebook friend or

Twitter follower might know someone in your office well enough to mention that “Oh, I know a co-worker of yours! Sounds like you have some serious HR issues . . .” Social media is indeed connecting us more than ever before, but there are many significant silos the structural integrity of which we wish to maintain.

These social silos were previously maintained not only by non-simultaneous interactions with different groups and organizational contexts, but also by the mundane barriers of time and space, missing in digital and especially in SNS environments. In our offline lives, when one is in church one is not also simultaneously in the office, in one’s tennis partner’s car on a family vacation, in one’s adult children’s living rooms—and, similarly, when one is out on the town, it is not also simultaneously the morning after, next Monday at lunch break, and five years later while interviewing for a new position. Digital media do not limit information flows through time and space the same ways as do physically based interactions, and our ability to predict to where information may flow and how it may matter to others and in other contexts—and to project that prediction indefinitely into the future and in relation to concerns which our future selves may have—is obviously insufficient to inoculate ourselves against the “privacy virus” that SNS presents.¹² Worse still, in the absence of these mundane architectural barriers of time and space, and the social barriers to which they give rise, even our most thoughtful connections may not be able to accurately perceive and maintain the limits on information flows which we seek to maintain.

The co-worker who we run into at the gay bar, regardless of his sexual orientation, must have overcome potential social barriers by being sufficiently comfortable with presence in a context and location where a sexualized same-gender gaze is considered normal and proper rather than deviant. Given these mundane conditions, those who may bump into a co-worker at the gay bar—whether they be taking part in a community of common self-identification, or whether they be gay-friendly straights who are there to see a drag show, or because it’s just the best place in town to go dancing—can at least know that the other party has similarly passed through these social filters. Although it may not be known by either party what has brought the other there, both are “insiders” insofar as they have each met these conditions, and are thus aware that this knowledge of one another, conditioned by this limited mode of access, ought to be treated as privileged information to be transmitted only selectively.

By contrast, identification of sexual orientation through SNS profile data requires only a connection of any kind arising within any context in order to grant access to potentially sensitive information. But even without this self-disclosure, all contacts from all contexts are welcome in the virtual gay bar that may be overlaid on the SNS user’s page and feed. A vague work contact, made at a professional conference, is invited along to passively overhear conversations within communities which he might never have been invited and might never have made himself a party to—even if a user, for example, posts news of gay marriage legal triumphs and vacation pictures with her partner only to a limited “close friends” list, her page nonetheless remains a venue in which

conversations take place within overlapping contexts. A public post absent identity markers, a popular music video, for example, may receive a simple comment from an “in-group” friend (e.g., “Too bad she’s straight!”), and through such interactions a potentially sensitive social context may coalesce around all those participants and passive viewers present—and all this without the “in-group” friend having any cues that she has broken down a silo. How are we to know which of a friend’s user-defined groups we are in, and how they are organized?

These effects are related to prior theorizations of Meyrowitz’s “middle region,” Papacharissi’s “publicly private and privately public spaces,” and Marwick and Boyd’s “context collapse.”¹³ What is perhaps most distinctive about this particular case is the way these identity performances are tied to unitary SNS profiles and take place within shifting and interlocking publicities rather than across a public/private divide. We are not seeing the private leaking out into the public so much as we are seeing a variety of regional publics overlaid upon one another. In this, we are called to account for our contextual identities in a new way: our selves are displayed, through both our actions as well as through others’ interactions with us, simultaneously before a multiplicity of audience with which we may identify in different ways.

This is the most peculiar challenge to integrity in an age of social media: we can no longer work out our own idea of how our values and commitments can harmonize into an integral self. Siloed identity performances allow us to perform those aspects of our identity, understood as that version of ourselves with which we identify, which fit within one context and another context, variously and in sequence. We can be gay in one context, Muslim in another, and a soldier in another still, and whether and to what extent those identities can be integrated can largely be sequestered as an issue for our own moral introspection and self-labor. Once these identities must be performed before a promiscuously intermixed set of audiences, integrity in the sense of staying true to our values takes on a newfound publicity, for we can no longer gain acceptance within groups merely by maintaining the local expectations for values and behaviors within each group in turn, but instead must either (1) meet each and all local expectations globally, (2) argue before others for the coherence of these identities when they vary from expectations particular to each group with which we identify, or (3) rebuild and maintain silos where time, space, and context no longer create them.

Indeed, so striking is this change that some have worried whether we are losing our interiority altogether.

INTEGRITY AND THE “ORGANIZATION MAN”

The worry that maintaining multiple profiles, and with them multiple selves, reflects a lack of integrity is a Scylla in the anxieties of popular discourse about SNS to which there is a corresponding Charybdis: the fear that an emerging “let it all hang out” social norm will destroy the private self altogether, and ring in a new age of conformity, where all aspects of our lives become performances before (and by implication for) others.

There are, however, significant reasons to believe that, even if our lives become ubiquitously subject to surveillance and coveillance, this will not result in the exclusion of expressions of marginalized identities or unpopular views.¹⁴ First, we see tendencies towards formation of social and informational echo chambers, resulting in increasingly extreme views rather than an averaging-out to moderate and universally accepted views, as Sunstein has argued for and documented at length.¹⁵ But, secondly, even insofar as we do not separate ourselves out into social and informational “Daily Me”s becoming a virtual “city of ghettos,” the messy and contentious digital spaces in which we are called to account for the integration of our multiple selves may tend not only towards safe and “lowest-common denominator” versions of self-expression, but also towards greater visibility and impact of divergent views, and even a new impetus away from conformity.¹⁶

Thus far we have considered how limiting information flows across social and organizational contexts can promote integrity, but it is certainly true as well that such siloing of different self-performances can support a lack of integrity. Compartmentalization is a key tool in allowing diffusion of responsibility. The employee who takes an “I just work here” perspective in her professional life is more likely to encounter productive cognitive dissonance when participating in the mixed contexts of SNS in which discussions with co-workers about their employer’s actions are subject to viewing and commentary by other friends who may view a corporate triumph as an environmental disaster. The churchgoer who has come to a private peace with her personal rejection of some sectarian dogmas may be forced into a more vocal and public advocacy by having to interact simultaneously with various and divergent friends’ reactions to news of court rulings about abortion rights.

In these sorts of cases, there is a clear threat to identity performances, placing users into precarious positions wherein they must defend and attempt to reconcile seemingly incompatible group identifications—but this loss in the user’s tranquility, in some cases, may bring with it a gain in personal integrity and possibilities for organizational reform. While it is certainly a bad thing that intermixing of audiences may subject users to discrimination, and separate performances of identities proper to different groups and contexts need not be indicative of a lack of integrity, compartmentalization can also enable people to act against their own values and stifle productive criticism within organizations.

Luban et al. argue forcefully, with reference to the Milgram experiment, that bureaucracies create a loss of personal responsibility for collective outcomes, resulting in what Arendt called “rule by nobody.”¹⁷ They suggest that we should attempt to maintain adherence to our moral values—maintain our integrity in the sense of staying true to the version of ourselves with which we identify—by analogy to how we think of our responsibility for our actions when under the influence of alcohol. Just as we plan in advance for our impaired judgment later by taking a cab to the bar or designating a driver, so too, before we enter into an organizational context we should be aware

that our judgment will become impaired by groupthink and diffusion of responsibility, and work out ways in which we can avoid making poor judgments under that organizational influence. Social networks may metaphorically provide that more-sober friend who asks “are you sure you’re okay to drive?,” enabling our better judgment to gain a foothold.

Organizations may then have a similar relation to our integrity as does our character. Our character is formed by a history of actions and interactions, but we may not identify with the actions that it brings us to habitually perform. When we recognize our vices—e.g., intemperance—and seek to act in accordance with our values and beliefs, we act against our character and contribute thereby to reforming our habits and character to better align with the version of ourselves with which we identify. Organizations may similarly bring us, through their own form of inertia and habituation, to act in ways contrary to our values and beliefs. A confrontation with this contradiction through context collapse may help us to better recognize the organization’s vices and to act according to the version of ourselves, in that organizational context, with which we identify—and contribute thereby to reforming our organization to better align with our values, and with its values as well.

NOTES

1. D. Kirkpatrick, *The Facebook Effect*, 199.
2. M. Zimmer, “Facebook’s Zuckerberg,” n.p.
3. K. Healy, “Actually, Having One Identity for Yourself Is a Breaching Experiment,” n.p.
4. B. Stone and S. Frier, “Facebook Turns 10,” n.p.
5. D. Hume, *A Treatise of Human Nature*, I.4.6.
6. Aristotle, “Nicomachean Ethics,” 1729–1867.
7. J.-P. Sartre, *Existentialism and Human Emotion*; Sartre, *Being and Nothingness*, 101–03.
8. To forestall a possible misunderstanding, I do not mean to claim that alcoholism is a matter of character. As I understand it, the common view among those who identify as alcoholics is that it is a disease and a permanent condition—what is subject to change is whether the alcoholic is keeping sober or has relapsed. This is where character comes into play—specifically, the hard work of (re)gaining and maintaining the virtue of temperance through abstemiousness.
9. J. Suler, “The Online Disinhibition Effect.”
10. Discussion in the first part of this section covers material addressed more systematically in D. E. Wittkower, “Facebook and Dramauthentic Identity.”
11. H. Nissenbaum, “Privacy as Contextual Integrity.”
12. J. Grimmelmann, “The Privacy Virus.”
13. J. Meyrowitz, *No Sense of Place*; Z. Papacharissi, *A Private Sphere*; A. Marwick and D. Boyd, “I Tweet Honestly, I Tweet Passionately.”
14. S. Mann et al., “Sousveillance.”
15. C. Sunstein, *Republic.com 2.0*; Sunstein, *Going to Extremes*.
16. N. Negroponte, *Being Digital*; E. Pariser, *The Filter Bubble*; Wittkower, “Facebook and Dramauthentic Identity.”
17. D. Luban et al.; H. Arendt, *On Violence*, 38-39.

BIBLIOGRAPHY

Arendt, H. *On Violence*. New York: Harcourt, Brace & World, 1969.
 Aristotle. “Nicomachean Ethics.” In *The Complete Works of Aristotle*, edited by J. Barnes Princeton, NJ: Princeton University Press, 1984.

Grimmelmann, J. "The Privacy Virus." In *Facebook and Philosophy*, edited by D. E. Wittkower. Chicago: Open Court, 2010.

Goffman, E. *The Presentation of Self in Everyday Life*. New York: Doubleday, 1959.

Healy, K. "Actually, Having One Identity for Yourself Is a Breaching Experiment." *Crooked Timber*, May 14, 2010. Retrieved from <http://crookedtimber.org/2010/05/14/actually-having-one-identity-for-yourself-is-a-breaching-experiment/>.

Hume, D. *A Treatise of Human Nature*. Project Gutenberg, 2012. Retrieved from <http://www.gutenberg.org/files/4705/4705-h/4705-h.htm>.

Kant, I. *Critique of Pure Reason*. New York: Cambridge University Press, 1998.

Kirkpatrick, D. *The Facebook Effect*. New York: Simon & Schuster, 2010.

Luban, D., A. Strudler, and D. Wasserman. "Moral Responsibility in the Age of Bureaucracy." *Michigan Law Review* 90, no. 8 (1992): 2348–92.

Mann, S., J. Nolan, and B. Wellman. "Sousveillance: Inventing and Using Wearable Computing Devices for Data Collection in Surveillance Environments." *Surveillance & Society* 1, no. 3 (2003): 331–55.

Marwick, A., and D. Boyd. "I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience." *New Media & Society* 13, no. 1 (2011): 114–33.

Meyrowitz, J. *No Sense of Place: The Impact of Electronic Media on Social Behavior*. New York: Oxford University Press, 1986.

Negroponte, N. *Being Digital*. New York: Vintage, 1996.

Nissenbaum, H. "Privacy As Contextual Integrity." *Washington Law Review* 79, no. 1 (2004): 119–57.

Papacharissi, Z. *A Private Sphere: Democracy in a Digital Age*. Malden, MA: Polity Press, 2010.

Pariser, E. *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think*. New York: Penguin, 2012.

Sandel, M. "The Procedural Republic and the Unencumbered Self." *Political Theory* 12, no. 1 (1984): 81–96.

Sartre, J.-P. *Being and Nothingness*. New York: Washington Square Press, 1993.

Sartre, J.-P. *Existentialism and Human Emotion*. New York: Citadel, 2000.

Stone, B., and S. Frier. "Facebook Turns 10: The Mark Zuckerberg Interview." *Business Week*, January 30, 2014. Retrieved from <http://www.businessweek.com/printer/articles/181135-facebook-turns-10-the-mark-zuckerberg-interview>.

Suler, J. "The Online Disinhibition Effect." *CyberPsychology & Behavior* 7, no. 3 (2004): 321–26.

Sunstein, C. *Republic.com 2.0*. Princeton, NJ: Princeton University Press, 2009.

Sunstein, C. *Going to Extremes: How Like Minds Unite and Divide*. New York: Oxford University Press, 2011.

Wittkower, D. E. "Facebook and Dramathentic Identity: A Post-Goffmanian Model of Identity Performance on SNS." *First Monday*, 19, no. 4 (2014): n.p. Retrieved from <http://firstmonday.org/ojs/index.php/fm/article/view/4858/3875>.

Zimmer, M. "Facebook's Zuckerberg: 'Having Two Identities for Yourself Is an Example of a Lack of Integrity,'" May 5, 2014. *Michaelzimmer.org*. Retrieved from <http://www.michaelzimmer.org/2010/05/14/facebooks-zuckerberg-having-two-identities-for-yourself-is-an-example-of-a-lack-of-integrity/>.

The Moral Roots of Conceptual Confusion in Artificial Intelligence Research

Niklas Toivakainen

UNIVERSITY OF HELSINKI

INTRODUCTION

I gather that it would not be an overstatement to claim that the field of Artificial Intelligence (AI) research is perceived by many to be one of the most fascinating, inspiring, hopeful, but also one of the most worrisome and dangerous advancements of modern civilization. AI research and related fields such as neuroscience promise to replace human labor, to make it more efficient, to integrate robotics into social realities,¹ and to enhance human capabilities. To many, AI represents or incarnates an important element of a new philosophy of mind, contributing to a *revolution in our understanding of humans and life in general*, which is usually integrated with a vision of a new era of human and super human intelligence. With such grandiose hopes invested in a project it is not surprising that the same elements that invoke hope and enthusiasm in some, generate anxiety and disquietude in others.²

While I will have things to say about features of these visions and already existing technologies and institutions, the main ambition of this paper is to discuss what I understand to be a pervasive moral dimension in AI research. To make my position clear from the start, I do not mean to say that I will discuss AI from a moral *perspective*, as if it could be discussed from other perspectives *detached from morals*. I admit that thinking about morals in terms of a "perspective" is natural *if* one thinks of morality as corresponding to a theory about a separable and distinct dimension or aspect of human life, and that there are other dimensions or aspects, say, scientific reasoning for instance, which are essentially *amoral* or "neutral" with respect to morality. Granting that it is a common trait of modern analytical philosophy and scientific thinking to precisely presuppose such a separation between fact and morality (or "value" as it is usually perceived), I am quite aware that moral considerations enters into the discussion of AI (as is the case for all modern techno-science) as a distinct and separate consideration. Nevertheless, I will not be concerned here with a critique of moral evaluations relevant for AI research—as, for instance, an ethics committee would be—but rather with *radicalizing* the relationship between morality and techno-science.³ My main claim in this paper will be that the project of AI—as the project of any human endeavor—is *itself* inextricably a moral matter. Much of what I will be doing here is to try and articulate how this claim makes itself seen on many different levels in AI research. This is what I mean by saying that I will discuss the moral dimensions of AI.

AI AND TECHNO-SCIENTIFIC UNDERSTANDING OF NATURE

The term "Artificial Intelligence" invites three basic philosophical—i.e., conceptual—challenges: What *is* (the