Models of Minds: Reading the Mind Beyond the Brain

Nick Merrill
John Chuang
School of Information, University of California, Berkeley
Berkeley, CA, USA
ffff@berkeley.edu
john.chuang@berkeley.edu

ABSTRACT

Drawing on philosophies of embodied, distributed & extend cognition, this paper argues that the mind is readable from sensors worn on the body, and embedded in the environment. It contends that past work in HCI has already begun such work, introducing the term *models of minds* to describe it. To those who wish to develop the capacity to build models of minds, we argue that notions of the mind are entangled with the technologies that seek to sense it. Drawing on the racial and gendered history of surveillance, we advocate for future work on how models of minds may reinforce existing vulnerabilities, and create new ones.

CCS CONCEPTS

• **Human-centered computing → HCI theory, concepts and models**; *Ubiquitous computing*.

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland UK

ACM ISBN 978-1-4503-5971-9/19/05…$15.00
https://doi.org/10.1145/3290607.3310427

KEYWORDS

models of minds; biosensing; philosophy of mind; surveillance; security

ACM Reference Format:

Nick Merrill, John Chuang. 2019. Models of Minds: Reading the Mind Beyond the Brain. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI'19 Extended Abstracts), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3290607.3310427

INTRODUCTION

Figure 1: ***Ophiocordyceps unilateralis sensu lato*** **takes control of an ant's mind, without input from its brain. By constructing a network of sensors and actuators atop its muscles, the fungal complex forces the ant to chew on the underside of a twig, after which the ant's body will serve as a medium for fungal reproduction.**

Consider the ant. The fungal complex *Ophiocordyceps unilateralis sensu lato* overtakes the ant's behavior without acting on its brain at all. Instead, it uses the ant's body to navigate the world, constructing a network of coordinated sensing and actuation atop the ant's muscles [16]. By sensing the ant's environment and stimulating its muscles in response, it causes the ant to crawl beneath

What is the mind? What is its relationship to the body, and to the physical world? Philosophers' answers to this question fall into two basic categories. (For a slice of this debate, see [9]).

Dualism posits that the mind has non-physical components. Since the sensors discussed here sense only physical phenomena, the dualist perspective presents an impasse for our analysis (how can physical devices sense the non-physical?).

Physicalism proposes a mind of strictly physical composition. The physicalist interpretation lends itself naturally to scientific study—and to sensing. From the physicalist perspective, all phenomena in the mind can be reduced to descriptions of physical activity; thus, some physical theory will eventually explain the mind in entirety.

The physicalist perspective provides a natural route forward for our analysis. It implies that a sufficiently sensed world, combined with sufficient theories, could yield computational *models* of minds.

Sidebar 1: Top, fungal filaments surround an ant's mandible muscle [16]. Bottom, commercial sensing devices decorate the wrists of an enthusiastic self-tracker [14]. (Top image: ©2019 National Academy of Sciences.)

a twig and bite into it. Once affixed to the twig, the fungus paralyzes the ant, using its body as a breeding ground (Figure 1).

Setting aside its ability to control the ant, consider the degree of *sensing* the fungal complex must perform in order to utilize the ant's body. Using the ant's bodily infrastructure, the fungi create a *model* of ant-experience robust enough to control the organism completely. Although the *Ophiocordyceps* fungal complex cannot read the ant's brain (it has no physical presence there), it can read the ant's *mind* well enough to model its environment and body. The fungi's model of ant-experience may not be the same as—or even similar to—those used by the host ant. Regardless, the model has functional properties that allow the fungus to achieve its (reproductive) goals.

With this fungal complex in mind, consider the emerging class of internet of things (IoT) devices, which are increasingly embedded in the built environment or worn on the body (Figure 1). Though common, cameras too sense bodies, often in public and without subjects' knowledge [26]. All of these connected devices are endowed to some degree with the capacity to sense (and to build models of) human bodies in space. Past work has referred to this process broadly as *biosensing*, and these devices as *biosensors* [12].

While humans are significantly more complex than ants, the *Ophiocordyceps* fungal complex helps illustrate the possibility of creating *models of minds* with limited or no information from the brain. If fungi can do so, perhaps consumer sensors can, as well. As we review in the following section, contemporary philosophical theories engage seriously with the notion of a mind beyond the brain and beyond the body, raising the possibility that sensors worn or embedded in the environment can sense the mind. We argue that prior work in HCI has already begun this work. To describe it, we introduce the term *models of minds*. We discuss how these models might structure our notions of the mind itself. Finally, We draw on the racial and gendered history of surveillance to motivate future work on how models of minds may reinforce and create threats to autonomy, safety, and security online.

Material theories of mind

The remainder of this section outlines various physicalist theories of the mind. Beginning with cognitive science's computational accounts of the mind, we trace critiques of this field to the newer theories that have come to meet them. These theories motivate notions of a beyond-the-brain mind.

Cognitive science

Cognitive science has historically served as an influential source of physicalist theories about the mind. The field takes a computational account of the brain, understanding how it "processes information" [30] within the physical constraints of computational space and time [28]. This perspective offers computational *models* of cognition [28]. (These models informed the design of neural networks, before the relatively recent discovery of performant backpropogation algorithms made neural networks practical to deploy [21]).

However, cognitive scientific models of the mind have received considerable criticism [22, 30]. Two relevant critiques focus on cognitive science's "isolationist assumptions": a focus on the brain (isolated from the body), and a focus on the individual (isolated from social context, and from the environment). The following sections review major responses to these critiques: embodied cognition, distributed cognition, and extended cognition. These theories return later as we discuss prior work in affective computing and ubiquitous computing (*ubicomp*).

Cognitive Science: Influential in the early days of AI, cognitive science seeks to create computational models of cognition. Cognitive scientific theories seek computational or programmatic descriptions of cognitive activity, referring to fundamental limits in computational space and time to account for phenomena observed in experimental studies with living subjects.

Embodied Cognition: Proponents of embodied cognition critique cognitive science for its isolation of the brain as a unit of analysis, claiming this isolation undervalues the role of the body in performing cognition. Responding to critiques of cognitive science, embodied cognition seeks to demonstrate the role of the agent's beyond-the-brain body in performing cognition, and the role of this body in shaping the mind that cognitive science seeks to study.

Extended & Distributed Cognition: Pioneered by Clark & Chalmer's theoretical work [10] and by Hutchin's observational work in a naval vessel [20], extended and distributed cognition trouble the borders between the body and non-body, claiming that built artifacts, and other humans, together perform cognition *beyond* the body, as well as within it.

Mind extending into body: Embodied cognition

Cognitive science's study of the brain in isolation rests on the assumption that the brain is strictly equivalent to the mind (i.e., that all mental phenomena occur in the brain). This assumption has encountered two primary critiques. First, the dichotomy between the brain and body is unstable; neurons occur body-wide, running directly to the brain, such that it is difficult to evaluate the role of cerebral neural activity in cognition without also considering non-cerebral neural activity. Second, to quote Noë and Thompson (2004), "[t]he exact way organisms are embodied simultaneously constrains and prescribes certain interactions within the environment." [22]. Mind manifests as does due to the physical conditions of the body.

These critiques gave rise to the *embodiment thesis*: that an agent's beyond-the-brain body plays a causal role in that agent's cognitive processing. For example, Noë and O'Regan's analysis of vision recasts the "visual processing" of cognitive science, in which internal representations are built and manipulated within the brain, to an active, embodied process, in which the world is not simply waiting to be seen, but actively providing its representations; the body and brain must meet through an active process of co-adaptation [23].

Mind extending beyond body: Extended and distributed cognition

While the embodiment thesis prods at the causal relationship between mind and the physical conditions of the body, it glosses over the relationship between these bodies and the world in which they are situated. In response, Clark and Chalmer's *extended cognition* thesis argues that the environment at large can be considered as part of the mind; that "technological resources such as pens, paper, and personal computers are now so deeply integrated into our everyday lives that we couldn't accomplish many of our cognitive goals and purposes without them" [10]. This theory does not stop at tools in describing a mind beyond the body. Broadly, extended cognition refocuses the brain away from the individual body, and toward the "active role of the environment in shaping cognition" [10]. This theory paved the way toward a socially-extended cognition, or *distributed cognition*, as described in Hutchins' (1995) ethnography of sailors on a naval vessel [20]. In his analysis, multiple individuals, and the material environment play constituent roles in cognition, manifesting a mind that is distributed across multiple human and non-human actors.

In addressing some critiques levied against cognitive science, the theories in this section make various cases for a mind that extends beyond the confines of the brain, and even beyond the confines of the body. The following section argues these theories, perhaps unwittingly, make the mind amenable to modeling via sensors that are worn or embedded in the environment, and that past research has (also unwittingly) already begun to sense the mind from beyond the brain.

READING THE MIND IN HCI

The theories outlined in the previous section all propose that the mind is physically instantiated in the material world. Using these theories, this section argues that prior work in HCI has already attempted to sense aspects of mind from beyond-the-brain bodies.

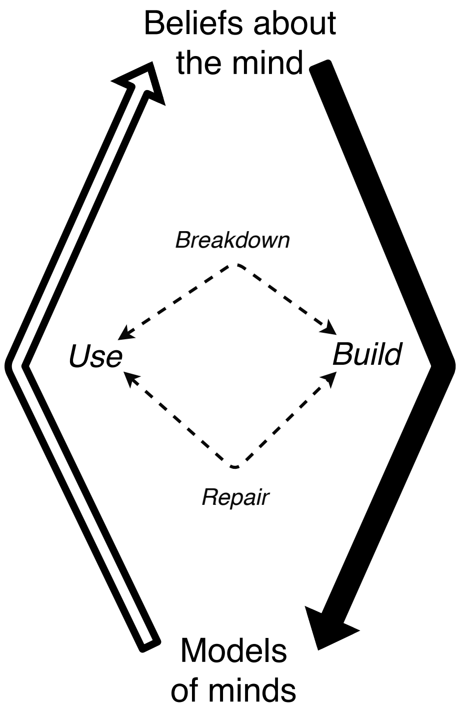
I read two strands of existing work in HCI and computer science through different accounts of mind: affective computing through embodied cognition, and ubiquitous computing (ubicomp) through distributed and embodied cognition. I argue that physical theories of the mind allow these two fields to claim that they sense the ground truth of mental phenomena. Thus, we argue that these fields have already begun using data from the beyond-the-brain-body to build *models of minds*—a term we will return to in the following section.

Models of minds: *Computational accounts of minds in the world, derived from sensor-mediated observation.*

This definition is specific about sensing minds in the world; it does not cover the synthesis of artificial minds through, for example, artificial intelligence. However, it is *not* specific about sensing brains, nor even about sensing bodies. Resting on the theories of extended and distributed cognition, *models of minds* provides room for a diversity of sensing modalities—and sensing targets.

Affective computing

Affective computing, pioneered by Rosalind Picard at the MIT Media Lab, seeks to use sensors to measure a users' affect, emotions, and mood in order to improve their interaction with machines. [24]. Two commercial examples of such sensing come directly from work in Rosalind Picard's research group. The Empatica wristband senses electrodermal activity, with the aim of correlating these data to emotional states [17]. This wristband has gone on to inspire cheaper consumer products, such as the Feel [15]. Also from Picard's lab, Affectiva classifies emotions from facial expressions, as detected by a camera. Their infrastructure works through a webcam, providing what they term "Emotion as a Service" [1]. In both of these examples, affect is framed as a bodily state, as in theories of embodied cognition. However, affective computing extends these claims further, positing that wearable sensors can measure, encode, and transmit emotions through their sensing of bodily states [19]. Although work in affective computing does not generally make explicit references to embodied cognition, it typically seeks to detect emotion via bodily phenomena, and does not consider these phenomena to be proxies from real emotions, indicating a general view of emotions as embodied primarily.



Sidebar 2: A big loop: beliefs about the mind inform the design of tools, and the use of these tools inform beliefs about the mind.

Ubiquitous computing

Ubiquitous computing (ubicomp) pioneered the use of mobile sensors as sources of data about human behavior, efforts that predate both commercial IoT devices and the general ubiquity of smartphones in the global north [3]. The *Social fMRI* provides a seminal example. A distributed, multimodal sensing infrastructure, implemented via mobile phones over more than a year, aimed at sensing "how things spread in [a] community, such as ideas, decisions, mood, or the seasonal flu" [2]. In this frame, both "ideas" and "the flu" are equated as properties not of individuals, but of communities and relationships. The Social fMRI study spawned numerous, similar projects, including one explicitly aimed at detecting "happiness" [4] or "creativity" [7], and, relevant to our discussion, one that aimed to diagnose depression from mobile phone traces [8]. In this study, longitudinal GPS traces were correlated with answers on questionnaires via machine learning and related statistical techniques.

As embodied cognition allows affective computing to present bodily phenomena as constituent of emotions, distributed and extended cognition allow this work to present extrabodily and multi-individual phenomena as constituent of mental states. If depression is an embodied phenomenon, then the phone senses depression's bodily correlates. However, if depression is an extended or distributed phenomenon, then the cellphone is, in fact, *a constituent of depression itself*; it reports on depression's ground truth. This example seeks to illustrate how distributed and extended cognition may trouble the boundaries of mental phenomena, dissolving our assumptions about how such phenomena may be sensed or detected.

MODELS OF MINDS

So far, this paper has argued that machines can know the inner workings of the human mind, from the brain and also beyond it, drawing from the body and the build environment. It further argues that past work in HCI has already begun in this program. How can we refer to such projects?

We propose the term *models of minds* (MoMs). This term borrows from philosophy's *theory of mind*, which refers to the (human) ability to reason about mental states [13, 31]. By substituting the word "theory" with the word "model," we emphasize formal or algorithmic representations. Turning this singular "model of mind" into a plural *models of minds* centers the diversity of minds in the world to model, and the diverse beliefs that could underlie these models' construction. For example, MoMs might draw on any, all, or none of the theories outlined previously. The term aims to cast a subtle doubt on models that appear too simple, or which (cl)aim to generalize too broadly.

While models of minds can include data about the brain, this paper argues that such data is not necessary. Given ubiquitous enough sensing, the world at large could be (re)purposed to sense the mind. Consider the minimal example of a lightswitch. Its state can be taken as a correlate of the beliefs and attitudes of the switcher(s): a request for light, a sense of darkness [29].

A Big Loop

Rather than presenting a theory of mind and a set of technologies that do or do not sense it, the term *models of minds* centers the relationship between beliefs about the mind, and the perceived or imagined capabilities of the technologies that seek to sense it. This relationship can be depicted as a big loop (Figure 2). In the right half of this loop, beliefs about the mind affect the technologies people build (and accept as working). The left half of this loop depicts existing technologies affecting beliefs about what the mind *is*.

This feed-forward loop raises the possibility that minds are not only readable because people believe they are, but because the very notion of personhood will change relative to existing claims about technical capabilities. How do we (re)make minds (and ourselves) through the things that sense them? An old question, “Are minds machines?” [30] comes under new light in this frame. Rather than asking what kinds of machines minds are, we may as well ask, how are machine-ness and mind-ness always already entangled, and what are the consequences?

The shifting of categorical boundaries, especially as it relates to shifts in technological infrastructures, has been the concern of philosophers of technology [5], feminist scholars [18], and disability scholars [27] for many years. Future work should integrate these perspectives in an examination of the other half of our big loop, or in an examination of the loop itself. Such work could complicate this notion of a loop, framing machines and minds as constantly co-constructed, always already entangled.

Security, privacy, surveillance, autonomy

How can HCI researchers hold themselves accountable to the impacts of their creations, such that their technical output does not inadvertently produce a technological *Ophiocordyceps* infection? MoMs provide a stark example of how desperately society requires practices and procedures for developing safe technologies, and for evaluating novel tools’ societal impacts. Future work at CHI and beyond should actively engage in developing such practices. Toward what theories might HCI researchers look as we do so?

Surveillance studies provides a rich sociocultural context, one which could provide theoretical footing for future practices to evaluate the impact of MoMs. In Simone Browne’s seminal history of surveillance in the United States [6], a racial, gendered and historical situatedness illuminates relationships between surveillance and power. While Browne’s history does not paint an optimistic picture for information technologies, Mcmillan Cottom’s work on black cyberfeminism [11] shows how the same tools of Browne’s surveillance can be repurposed to evade surveillance, and for activism. Future work should substantially engage with analyses such as these, so that we may better understand both what new power structures MoMs might create, and which existing ones it might reinforce. Future work must engage with these topics deeply.

So what? Broader impacts

Models of minds provide a new frontier for surveillance, challenging the essential privacy of inward thoughts and feelings. Who benefits, and who will suffers the consequences, is a fundamental concern for the CHI community. Through study of the history of surveillance [6], CHI researchers can sensitize themselves to the ways technologies may be differentially applied based on relational markers such as race, class, sex, gender, and disability [25]. Only through a critical study of such histories can the CHI community hope to develop MoMs responsibly, and to foresee and resist harmful use.

CONCLUSION

As sensors continue to saturate our environment, they will increasingly track bodies in space. Machines' purported ability to divine not just what bodies do, but what they think and feel, will prove to be a key concern for privacy, personal autonomy, and cybersecurity in the coming hundred years. It will also bring novel opportunities for communication, accessibility, business, and entertainment. These concerns and opportunities will likely exist not in opposition to each other, but in mutual reinforcement, entanglement, co-construction. MoMs will produce not only new technologies, but new theories about what the mind is, and who we are as thinking beings. By attending to these entangled processes, we can better anticipate how (and why) the development of these technologies may occur, and thus better prepare for an increasingly connected—and increasingly insecure—world, body, and mind.

ACKNOWLEDGEMENTS

We thank the BioSENSE group for their insightful conversations, and the reviewers for their helpful feedback. This work was supported in part by the Berkeley Center for Long Term Cybersecurity (CLTC), and the Hewlett Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors.

REFERENCES

[1] Affectiva Inc. 2016. Emotion as a Service - Affectiva. http://www.affectiva.com/solutions/emotion-as-a-service/
[2] Nadav Aharony, Wei Pan, Cory Ip, Inas Khayal, and Alex Pentland. 2011. The social fMRI:Investigating and shaping social mechanisms in the real world. *Proceedings of the 13th international conference on Ubiquitous computing - UbiComp '11* (2011), 445. https://doi.org/10.1145/2030112.2030171
[3] Genevieve Bell and Paul Dourish. 2007. Yesterday's tomorrows: Notes on ubiquitous computing's dominant vision. *Personal and Ubiquitous Computing* 11, 2 (2007), 133–143. https://doi.org/10.1007/s00779-006-0071-x
[4] Andrey Bogomolov, Bruno Lepri, and Fabio Pianesi. 2013. Happiness recognition from mobile phone data. In *Proceedings - SocialCom/PASSAT/BigData/EconCom/BioMedCom 2013*. IEEE, 790–795. https://doi.org/10.1109/SocialCom.2013.118
[5] Geoffrey C Bowker and Susan Leigh Star. 2000. *Sorting things out: Classification and its consequences*. MIT press.
[6] Simone Browne. 2015. *Dark Matters: On the Surveillance of Blackness*. Duke University Press.
[7] Winslow Burleson. 2012. Predicting Creativity in the Wild: Experience Sample and Sociometric Modeling of Teams. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)*. ACM, 1203–1212. https://doi.org/10.1145/2145204.2145386
[8] Luca Canzian and Mirco Musolesi. 2015. Trajectories of depression. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UBICOMP '15)*. 1293–1304. https://doi.org/10.1145/2750858.2805845
[9] David John Chalmers. 1998. The Conscious Mind: In Search of a Fundamental Theory. *International Journal of Quantum Chemistry* January 1998 (1998), xvii, 414. https://doi.org/DOI10.1207/s15327809jls0803&4_10 arXiv:arXiv:cond-mat/0402594v3
[10] Andy Clark and David Chalmers. 1998. The Extended Mind. *Analysis* 58, 1 (1998), 7–19. https://doi.org/10.1111/1467-8284.00096 arXiv:arXiv:1011.1669v3

[11] Mcmillan Cottom. 2016. Black CyberFeminism: Ways Forward for Intersectionality and Digital Sociology. (2016).
[12] Sophie Day and Celia Lury. 2016. Biosensing: Tracking Persons. *Quantified: Biosensing Technologies in Everyday Life* (2016), 43.
[13] Daniel Clement Dennett. 1989. *The intentional stance*. MIT press.
[14] Cory Doctrow. 2014. Rachel Kalmar's datapunk quantified self sensor array 2, Institute for the Future, Palo Alto, California, USA. https://www.flickr.com/photos/doctorow/15659135172
[15] Feel.co. 2018. Feel: The world's first emotion sensor & well-being advisor. https://www.myfeel.co/
[16] Maridel A. Fredericksen, Yizhe Zhang, Missy L. Hazen, Raquel G. Loreto, Colleen A. Mangold, Danny Z. Chen, and David P. Hughes. 2017. Three-dimensional visualization and a deep-learning model reveal complex fungal parasite networks in behaviorally manipulated ants. *Proceedings of the National Academy of Sciences* (2017), 201711673. https://doi.org/10.1073/pnas.1711673114
[17] Maurizio Garbarino, Matteo Lai, Dan Bender, Rosalind W. Picard, and Simone Tognetti. 2015. Empatica E3 - A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In *Proceedings of the 2014 4th International Conference on Wireless Mobile Communication and Healthcare - "Transforming Healthcare Through Innovations in Mobile and Wireless Technologies", MOBIHEALTH 2014*. 39–42. https://doi.org/10.1109/MOBIHEALTH.2014.7015904
[18] Donna Jeanne Haraway. 1988. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies* 14, 3 (1988), 575–599. https://doi.org/10.1177/0263276406069231 arXiv:arXiv:gr-qc/9809069v1
[19] Jennifer Healey. 2014. Physiological Sensing of Emotion. *The Oxford Handbook of Affective Computing* October (2014), 204. https://doi.org/10.1093/oxfordhb/9780199942237.013.023
[20] Edwin Hutchins. 1995. *Cognition in the Wild*. MIT press.
[21] Marvin Minsky and Seymour Papert. 1969. Perceptrons. (1969).
[22] Alva Noë and Evan Thompson. 2004. Are There Neural Correlates of Consciousness? *Journal of Consciousness Studies* 11, 1 (2004), 3–28.
[23] J. Kevin O'Regan and Alva Noë. 2001. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* 24, 05 (2001), 939–973. https://doi.org/10.1017/S0140525X01000115
[24] R. W. Picard and J. Healey. 1997. Affective wearables. *Personal and Ubiquitous Computing* 1, 4 (dec 1997), 231–240. https://doi.org/10.1007/BF01682026 arXiv:arXiv:1411.1164v1
[25] James Pierce, Nick Merrill, Richmond Y Wong, Sarah Fox, and Eric Paulos. 2018. An Interface without A User : An Exploratory Design Study of Online Privacy Policies and Digital Legalese. (2018).
[26] Elaine Sedenberg, Richmond Y Wong, and John C.-I. Chuang. 2017. A Window into the Soul: Biosensing in Public. *CoRR* abs/1702.0 (2017). arXiv:1702.04235 http://arxiv.org/abs/1702.04235
[27] Josephine A Seguna. 2015. Space and Affect: Using Heidegger to Re-interpret the Disability Experience. *Disability Studies Quarterly* 35, 4 (2015).
[28] Roger N Shepard and Jacqueline Metzler. 1971. Mental rotation of three-dimensional objects. *Science* 171, 3972 (1971), 701–703.
[29] Peter Tolmie, Andy Crabtree, Tom Rodden, James A Colley, and Ewa A Luger. 2016. "This has to be the cats": Personal Data Legibility in Networked Sensing Systems. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16*. 490–501. https://doi.org/10.1145/2818048.2819992
[30] Terry Winograd and Fernando Flores. 1987. On understanding computers and cognition: A new foundation for design. A response to the reviews. , 250–261 pages. https://doi.org/10.1016/0004-3702(87)90026-9
[31] Melanie Yergeau. 2013. Clinically significant disturbance: On theorists who theorize theory of mind. *Disability Studies Quarterly* 33, 4 (2013).