

# TWO CHEERS FOR CORPORATE EXPERIMENTATION: THE A/B ILLUSION AND THE VIRTUES OF DATA-DRIVEN INNOVATION

MICHELLE N. MEYER\*

INTRODUCTION: TWO FRAMES FOR THINKING ABOUT CORPORATE EXPERIMENTATION .....	274
I. BACKGROUND TO THE FACEBOOK EXPERIMENT .....	279
<i>A. Facebook’s Algorithmic News Feed Practice</i> .....	279
<i>B. The Emotional Valence of News Feed Posts</i> .....	282
<i>C. Four Hypotheses about the Effects of News Feed</i> .....	283
<i>D. The Facebook “Emotional Contagion” Experiment</i> .....	285
II. FRAME ONE: HUMAN SUBJECTS RESEARCH.....	287
<i>A. Subjects Gave No Ethically Meaningful Consent to the     Experiment</i> .....	288
<i>B. Why Consent Is Not Always an Ethical Requirement of     Human Subjects Research</i> .....	292
<i>C. Was Informed Consent an Ethical Requirement of the     Facebook Experiment?</i> .....	298
1. Infeasibility.....	298
2. Incompletely Informed Consent.....	298
3. Debriefing.....	300
4. Minimal Risk Analysis.....	302
III. FRAME TWO: RESPONSIBLE INNOVATION .....	310
<i>A. The Ethical Relevance of the Distribution of Research     Risks and Benefits</i> .....	310
<i>B. Mini Case Study in the A/B Illusion: OkCupid’s Matching</i>	

---

\* J.D., Ph.D.; Assistant Professor and Director of Bioethics Policy, Union Graduate College-Icahn School of Medicine at Mount Sinai Bioethics Program. For helpful questions and comments on various versions of this project, I thank seminar participants at Carnegie Mellon University’s Philosophy Department and Center for Ethics and Policy, Emory University’s Center for Ethics, and The Wharton School’s Legal Studies & Business Ethics Department and audience members and fellow panelists at MIT’s Conference on Digital Experimentation, Public Responsibility in Medicine and Research (PRIM&R)’s Advancing Ethical Research Conference, and the Silicon Flatirons Center’s Conference on When Companies Study Their Customers: The Changing Face of Science, Research, and Ethics.

<i>Algorithm Experiment</i> .....	312
<i>C. How the Research/Practice Distinction Fosters the A/B Illusion (And Overprotects Subjects and Underprotects Users)</i> .....	321
<i>D. A Brief Note on the Problem of Unethical Underlying Practices</i> .....	324
CONCLUSION: RESPONSIBLE INNOVATION & A CULTURE OF CONTINUOUS TESTING .....	327

#### INTRODUCTION: TWO FRAMES FOR THINKING ABOUT CORPORATE EXPERIMENTATION

Ten years ago, if you wanted to feel badly about how your life was going compared to the lives of your friends, family, and various acquaintances—or, conversely, if you wanted to be vicariously uplifted by their successes or depressed by their failures—you had to wait for the annual onslaught of holiday cards or a decennial school reunion. All that changed in September of 2006, when Facebook, the online social networking service launched two years earlier, introduced News Feed, a continually updated, selective aggregation of the accomplishments, tragedies, complaints, political musings, and cat pictures posted by users’ friends.<sup>1</sup>

Some ten years on, it is difficult for us to recognize it as such, but the advent of News Feed marked a major shift in how we allocate our time and in the way we observe and interact with others.<sup>2</sup> As Facebook noted on the occasion of its launch, News Feed is “not only different from anything we’ve had on Facebook before, but . . . quite unlike anything you can find on the web.”<sup>3</sup>

But what was a novel Internet experience in 2006 quickly became the primary means by which Facebook users both actively engage with their friends (for instance, by “liking,” sharing, or commenting on friends’ news items) and passively receive their friends’ content. By 2011, twenty-seven percent of all time spent on

---

1. Ruchi Sanghvi, *Facebook Gets a Facelift*, FACEBOOK (Sept. 5, 2006, 2:03 AM), <https://www.facebook.com/notes/facebook/facebook-gets-a-facelift/2207967130>. Facebook launched News Feed just a few weeks before opening up access to its service to anyone with a valid email address on September 26 of the same year. Facebook, *Our History*, <http://newsroom.fb.com/company-info/>.

2. See Robert E. Wilson, Samuel D. Gosling & Lindsay T. Graham, *A Review of Facebook Research in the Social Sciences*, 7 PERSPECTIVES ON PSYCHOLOGICAL SCI. 203, 204 (2012) (Facebook is “spawning new [social processes] by changing the way hundreds of millions of people relate to one another and share information”).

3. Matt McGee, *EdgeRank Is Dead: Facebook’s News Feed Algorithm Now Has Close to 100K Weight Factors*, MARKETING LAND (Aug. 16, 2013, 9:00 AM), <http://marketingland.com/edgerank-is-dead-facebooks-news-feed-algorithm-now-has-close-to-100k-weight-factors-55908>.

Facebook—at that time, over 4.7 million person-hours per day, not including Facebook use via the company's mobile application—was spent on News Feed. One year later, that number had grown to forty percent,<sup>4</sup> or about seven million person-hours per day.

Meanwhile, Facebook itself quickly became the largest online social media platform in the world. By July of 2014, Facebook boasted one-fifth of the world's population—1.35 billion users, and growing—as active monthly users;<sup>5</sup> 864 million of those (also growing) log on daily.<sup>6</sup> In the United States, more than half the population (129.5 million) logs into Facebook at least monthly,<sup>7</sup> and about forty percent logs in daily. The average U.S. adult user spends forty minutes per day on Facebook,<sup>8</sup> and more time on News Feed, in particular, than on the online news sites of ABC, MSNBC, Yahoo!, CNN, the *New York Times*, and the Huffington Post combined.<sup>9</sup>

What are the effects of this massive shift in how a sizable and growing portion of human beings engage with and learn about one another? We don't know, and if some critics have their way, we may never find out.

Academic studies have suggested two contradictory hypotheses about the risks of Facebook use to its 1.35 billion users: that exposure to friends' positive posts is psychologically risky (through a social comparison mechanism) and that exposure to negative posts is psychologically risky (through an emotional contagion mechanism).<sup>10</sup> But these contradictory studies were mostly small and observational. The company alone was in a position to sort this out and rigorously determine the mental health effects of News Feed. And for one week in January of 2012, Facebook—with some help

---

4. Pamela Vaughan, *Demystifying How Facebook's EdgeRank Algorithm Works*, (April 23, 2013, 9:00 AM), <http://blog.hubspot.com/marketing/understanding-facebook-edgerank-algorithm-infographic> (republishing an infographic developed by the now-defunct Facebook optimizer startup PostRocket with the help of Facebook's News Feed Product Manager Will Cathcart).

5. *Number of Monthly Active Facebook Users Worldwide as of 4th Quarter 2014 (in Millions)*, STATISTA (2015), <http://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>.

6. *Number of Daily Active Facebook Users Worldwide from 1st Quarter 2011 to 3rd Quarter 2014 (in Millions)*, STATISTA (2015), <http://www.statista.com/statistics/346167/facebook-global-dau/>.

7. *Facebook's US Ad Revenues Outpace Users' Average Daily Time Spent on the Site*, EMARKETER (Sept. 18, 2014), <http://www.emarketer.com/Article/Facebooks-US-Ad-Revenues-Outpace-Users-Average-Daily-Time-Spent-on-Site/1011215>.

8. Joshua Brustein, *Americans Now Spend More Time on Facebook Than They Do on Their Pets*, BLOOMBERG BUSINESS (July 23, 2014), <http://www.businessweek.com/articles/2014-07-23/heres-how-much-time-people-spend-on-facebook-daily>.

9. Vaughan, *supra* note 4.

10. See *infra* Part I.C.

from Cornell academics with expertise in studying how social interactions are mediated by information and communication technology—conducted an experiment in which it attempted to do just that.<sup>11</sup>

When, in June of 2014, the researchers published the results of the experiment—titled, perhaps regrettably, *Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks*<sup>12</sup>—and the media reported it, reaction was swift and fierce.<sup>13</sup> Criticism by both the public and some prominent ethicists centered on the fact that the 700,000 or so users involved had not consented to participate in what appeared to be a study designed to psychologically harm users by manipulating their emotions.<sup>14</sup> Critics charged Facebook with exploiting its position of power over users, treating users as mere means to the corporation’s (or scientists’) ends, and depriving users of information necessary to make a considered judgment about what was in their best interests.

But the considerable discussion of the experiment has paid scant attention to the experiment’s relationship to Facebook’s underlying practice of algorithmically curating users’ News Feeds and *its* risks and uncertainties. “Practitioners”—whether business people, lawmakers, clinicians, or other actors—are constantly innovating, in the broad sense of introducing new products, services, policies, or practices. In some cases, we have decided that the risks of such innovations require that they be introduced into small populations under carefully controlled conditions, and their safety and efficacy measured, before they are introduced into the general population.<sup>15</sup> But for the vast majority of innovations, no such *ex ante* regulation requiring evidence of safety and efficacy does—or

---

11. See *infra* Part I.D.

12. Adam D.I. Kramer, Jamie E. Guillory & Jeffrey T. Hancock, *Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks*, 111 PROC. NAT’L ACAD. SCI. 8788, 8788 (2014).

13. See *infra* notes 66–69 and accompanying text.

14. Although human beings who are studied as part of a research project are traditionally called “subjects,” it is increasingly common to refer to them instead as “participants.” Proponents of this change in terminology argue that it is more respectful of the important and active role that these individuals play in knowledge production. I am a research subject myself in numerous studies and I do not find “subject” to be pejorative (research “object” would be a different matter). Moreover, because many parties participate in the research enterprise (including investigators, sponsors, subjects, and subjects’ legal representatives), “participant” is often ambiguous. Finally, it is perhaps especially unfitting to refer to subjects in non-consensual research, which is the subject of this article, as “participants.”

15. See, e.g., 21 U.S. §§ 351-60 (providing that new drugs may not be introduced into interstate commerce without FDA approval and describing that approval process, including clinical trials to establish the drug’s safety and efficacy).

feasibly could—exist. In these cases, how should practitioners responsibly innovate?

Much of the time, innovation is ad hoc. A practitioner comes upon an idea (call it A) and because she has the power within the relevant institution (say, a company) to do so, she implements A by fiat. If those subject to A (say, end-users or employees) are very lucky, the practitioner will observe A's apparent effects and make any necessary adjustments. But without a control, the effects will be just that: apparent.<sup>16</sup>

Other practitioners do attempt to rigorously determine the effects of A before universally implementing it by first comparing it to an alternative (call this B, which may be an alternative innovation or the status quo) and randomizing half of people to receive A and half to receive B. But, not infrequently, if those people are told about the exercise, that knowledge will effect their behavior and bias the attempt to measure the comparative effects of A and B. In those cases, the practitioner may undertake the exercise more or less in secret, at least initially.

Practices that are subjected to such A/B testing (as marketers and data scientists refer to it) or experimentation (as scientists in other fields call it) generally have a far greater chance of being discovered to be unsafe or ineffective, potentially leading to substantial welfare gains if practitioners act on their newfound knowledge. Yet the conventional wisdom of many academics, members of Congress, and members of the public is that “human experimentation” is inherently dangerous and human experimentation without informed consent is absolutely unethical.<sup>17</sup> As Facebook learned, practitioners known to engage in A/B testing may find themselves at the center of a public relations nightmare, facing calls for federal agency investigations, injunctions, and more.<sup>18</sup>

In this article, using the Facebook experiment as a case study,<sup>19</sup> I

---

16. See note 46, *infra*.

17. I use “absolutely” here in the philosophical sense of a duty that applies in every case; such absolute moral duties are in contrast to *prima facie* duties, which may be overridden by other considerations.

18. See Sen. Warner Raises Questions About Facebook Experiment to Influence Users' Emotions, MARK R. WARNER, (July 9, 2014), [http://www.warner.senate.gov/public/index.cfm/pressreleases?ContentRecord\\_id=0c84f1af-5025-46bc-b118-87123cd745e1](http://www.warner.senate.gov/public/index.cfm/pressreleases?ContentRecord_id=0c84f1af-5025-46bc-b118-87123cd745e1) (calling on the Federal Trade Commission to investigate Facebook); In the Matter of Facebook, Inc. (2014) EPIC Complaint, Request for Investigation, Injunction, and Other Relief, <https://epic.org/privacy/internet/ftc/facebook/psycho/Facebook-Study-Complaint.pdf> (same). See also *infra* note 130.

19. Facebook's conundrum shares many features faced by—and so it is relevant for—not only other companies involved in digital experimentation, but also many others who seek to engage in data-driven practice, not least, clinicians and administrators

explore two frames through which we can think about that and similar field experiments. The first frame is the familiar one used by ethicists and regulators to govern human subjects research. Contrary to popular belief and even some expert commentary offered in the immediate wake of the Facebook experiment, this framework, articulated in the *Belmont Report* and codified in the federal Common Rule, does not absolutely prohibit non-consensual human subjects research. Rather, it appropriately permits *prima facie* duties to obtain subjects' informed consent to be overridden when obtaining consent would be infeasible and risks to subjects are no more than minimal.

This first framework, which I discuss in Part II (after providing relevant background information about Facebook's experiment and underlying practices in Part I), is designed for IRBs, who must decide whether to approve a protocol or not. As such, it is understandably limited to the threshold question of ethical permissibility. A second framework provides an additional reason to conclude that field experiments like the one Facebook conducted are ethically permissible: namely, the tight fit between the population upon whom (no more than minimal) risks are imposed and the population that stands to benefit from the knowledge produced by a study.

Instead of narrowly focusing on the ethics of a field experiment, this second framework contextualizes the experiment against the backdrop of the underlying practice, asking how a practitioner ought to responsibly innovate and the appropriate role, if any, of experiments in that innovation process. This framework, which I sketch in Part III, allows us to ask not only whether an experiment is ethically permissible but also whether it is ethically laudable or even obligatory. This framework may be especially fruitful for corporate managers and other practitioners who typically operate outside of the Common Rule.

Most commentators saw the Facebook experiment through the first framework of human subjects research and not through the second framework of responsible innovation. Why? I dub the "A/B illusion" the widespread tendency to view a field experiment designed to study the effects of an existing or proposed practice as more morally suspicious than an immediate, universal implementation of an untested practice. Critics of Facebook's

---

working in modern healthcare systems. The (comparative) effects on patients of many medical and healthcare delivery practices are uncertain, imperiling patient welfare and potentially squandering scarce resources. Healthcare systems are in a unique position to rigorously field test the consequences of their services, yet obtaining explicit informed consent for participation in learning activities (whether "research" or quality improvement (QI) / quality assurance (QA)) is often infeasible. See *infra* Part III.C.

“emotional contagion” experiment charged the company with exploiting its position of power over users, treating them as mere means to the corporation’s ends, and depriving them of information necessary for them to make a considered judgment about what was in their best interests. But the Facebook experiment simply created conditions—a somewhat more positive or negative news week than these particular users would otherwise have experienced that week—that are almost certain to fall within the normal range of their Facebook experience. Doing so allowed Facebook—and, because they published the results, the rest of us—to better understand the mental health effects of News Feed on its 1.35 billion users. Seen through the framework of responsible innovation, the criticisms of Facebook not only fall short of demonstrating that the experiment was unethical; they should be inverted. It is not the practitioner who engages in A/B testing but the practitioner who simply implements A who is more likely to exploit her position of power over users or employees, to treat them as mere means to the corporation’s ends, and to deprive them of information necessary for them to make a considered judgment about what is in their best interests.

## I. BACKGROUND TO THE FACEBOOK EXPERIMENT

### A. *Facebook’s Algorithmic News Feed Practice*

For reasons that will become clear, before considering the ethics of the Facebook experiment, it is necessary to understand Facebook’s underlying practice of aggregating and algorithmically curating friends’ posts, the effects of which on users the experiment sought to better understand.

Every Facebook user has her own page—or “Timeline”<sup>20</sup>—on which appears a combination of her profile information and a reverse chronological list of her posts (such as plain text “status updates,” with or without Internet links, others’ status updates that she has shared, photos, videos, and app activity). A Facebook user has always been, and remains, able to click on a friend’s Timeline to see everything that person has posted (to whatever extent the friend’s privacy settings permit). But navigating to each friend’s Timeline to every time a user wants to know whether that friend has posted something new is tedious (and likely makes users feel a bit

---

20. Facebook introduced Timeline in 2012 to replace and combine user profiles and “Walls.” Jill Duffy, *12 Things You Should Know About Facebook Timeline*, PCMag (Jan. 25, 2012), <http://www.pcmag.com/article2/0,2817,2393464,00.asp>.

stalkerish), and so in September of 2006, Facebook launched “News Feed”: “a personalized list of news stories throughout the day,” where “news items” refers, somewhat grandiosely, to anything that a friend (or group) posts.<sup>21</sup> As Facebook explained, “[n]ow, whenever you log in, you’ll get the latest headlines generated by the activity of your friends and social groups.”<sup>22</sup>

News Feed has never been a chronological list of all of the news items that a user’s friends produce. The average active user is eligible to see about 1500 items in their News Feed in any given session,<sup>23</sup> and Facebook, quite plausibly, does not believe that most users have time to read all of these. Nor does Facebook believe that the average user is equally interested in all 1500 posts, from his sister’s wedding photos to the announcement that his boss unlocked a new level in Candy Crush. Indeed, Facebook has found in “tests” that when they “stop ranking and instead show posts in chronological order, the number of stories people read and the likes and comments they make decrease.”<sup>24</sup>

From the beginning, then, Facebook has curated users’ News Feeds “in the interest of showing viewers the content they will find most relevant and engaging.”<sup>25</sup> The company uses a ranking

---

21. Sanghvi, *supra* note 1.

22. *Id.*

23. Facebook randomly sampled what it calls “daily active users” during one week in July of 2013 and found that the median such user is eligible to see about 1500 news items in any given session (i.e., when a user logs into her Facebook account, opens the Facebook application on her mobile device, or refreshes Facebook on her Internet browser). Lars Backstrom, *News Feed FYI: A Window Into News Feed*, FACEBOOK (Aug. 6, 2013), <https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed>. Some users are eligible to see tens of thousands of news items. Matt McGee, *Facebook Updates News Feed Algorithm With New “Story Bumping” & “Last Actor” Factors*, MARKETING LAND (Aug. 6, 2013, 3:14 PM), <http://marketingland.com/facebook-story-bumping-last-actor-54804>.

24. Backstrom, *supra* note 23.

25. Kramer et al., *supra* note 12, at 8788. Critics will rightly note that Facebook’s explanation of the purpose of News Feed is rather self-serving. Needless to say, the interests of companies like Facebook and their users are not identical. All else equal, most Facebook users would probably prefer not to see any ads in their News Feed, for instance, or at least not to have their data used to target ads to them. Of course, this is the price users must pay to access a “free” social media platform, without which that platform would not exist (and so such ads may be consistent with users’ overall preferences). Most efforts to maximize user engagement, however, seem to serve the interests of both Facebook and users. It is almost tautologically true that creating an enjoyable Facebook experience is in the interest of users (save, perhaps, masochistic users). If you’re the kind of person who tends to comment and “like” lots of photos, then the News Feed algorithm is configured to prioritize photos in your feed, the operating assumption being that you are, based on historical practice, relatively more likely to engage with photos than with other kinds of new items. But in addition, the more enjoyable using Facebook is, the more time users will spend on the site. The more time they spend on the site, the more likely they are to see (and, perhaps, engage with) the



algorithm to select and show the user about 300 items in any given session.<sup>26</sup> According to the company, its:

goal . . . is to deliver the right content to the right people at the right time so they don't miss the stories that are important to them. Ideally, we want News Feed to show all the posts people want to see in the order they want to read them.<sup>27</sup>

In pursuing this goal, Facebook “continually develops and tests” the News Feed algorithm.<sup>28</sup> Initially, the algorithm was crude and one-size-fits-all: Facebook likened it to “turning knobs” on different kinds of News Feed content for all users— “[t]urn up photos a little bit, turn down platform stories a little bit.”<sup>29</sup> Different users surely had different preferences, but the company relied on an unscientific squeaky-wheel test of the algorithm’s success: they adjusted the knobs based on “often angry emails and conversations with users outside the Facebook office.”<sup>30</sup>

The launch of Facebook Ads and Pages in 2007 produced significantly more eligible News Feed content for most users, increasing the odds that important content would get lost in the mix. So Facebook got serious about its proprietary ranking algorithm. EdgeRank, as the algorithm was referred to internally, ranked eligible news items according to three primary criteria: affinity (the extent to which the potential viewing user has previously interacted with the same type of content—photos, videos, status updates, sponsored content—and with the source user), weight (the amount of interaction with the item by others in the form of, e.g., likes, comments, clicks, and shares), and decay (time passed since the content was posted).<sup>31</sup> For instance, the more Cameron has previously interacted with Tyler’s posts, the more Cameron has

---

sponsored content that is Facebook’s lifeblood. Whatever actually motivated Facebook to create an algorithmically curated News Feed—a desire to create an enjoyable product, a desire to maximize profits, or some combination of the two—to a large extent, corporate and user interests in maximizing user engagement coincide. This is not to say that these sets of interests coincide perfectly, that all users have identical preferences (or that all Facebook employees have identical motivations), or that there is nothing else ethically wrong with the News Feed algorithm. But it is to say that glib descriptions of the News Feed algorithm as simply designed to sell users things—even if this accurately describes the intentions of the individuals responsible for the News Feed—neglects the fact that maximizing user engagement undeniably also produces benefits for users.

26. Backstrom, *supra* note 23.

27. *Id.*

28. Kramer et al., *supra* note 12, at 8788.

29. McGee, *supra* note 3.

30. *Id.*

31. *Id.*; see also Vaughan, *supra* note 4.

previously interacted with photos (posted by anyone), the more third party users have interacted with the particular photo Tyler posted, and the more recently Tyler posted that photo, the more likely Cameron is to see Tyler's photo in his News Feed.

By early 2010, Facebook had reportedly abandoned EdgeRank—in name, at least.<sup>32</sup> By then, the company had turned to machine learning, in which the News Feed algorithm “learns” from users' behavior to determine the optimal presentation.<sup>33</sup> Affinity, weight, and decay apparently remain important, but now contain “categories and subcategories.”<sup>34</sup> In August of 2013, Facebook's Engineering Manager for News Feed Ranking could only estimate that the increasingly complex algorithm was based on about 100,000 “weights.”<sup>35</sup> A team of seventeen Facebook computer engineers adjusts the News Feed algorithm about once per week.<sup>36</sup>

### B. *The Emotional Valence of News Feed Posts*

The emotional valence of eligible News Feed items will surely vary from user to user and over time, depending on the mix of Debbie Downers and Peppy Patties the user has befriended on the platform and what is happening in their lives and in the broader world. But overall, items posted to Facebook may be predominantly positive. A small survey (N=82) of users' motivations for using Facebook revealed that seventy-eight percent of respondents said they used Facebook, among other purposes, “to share good things with friends,” while only thirty-six percent reported using the platform “to share bad things with friends.”<sup>37</sup>

On top of this possible “positive bias” in what users post, the News Feed algorithm almost certainly affects the emotional valence of those items that users are most likely to see—again, probably in the direction of prioritizing positive content. Because Facebook's News Feed algorithm is proprietary, it is not clear whether any of its 100,000 or so weights directly prioritize (or deprioritize) posts on the basis of whether they contain words with a positive or negative

---

32. McGee, *supra* note 3.

33. *Id.*

34. *Id.*

35. *Id.*

36. Ravi Somaiya, *How Facebook Is Changing the Way Its Users Consume Journalism*, N.Y. TIMES (Oct. 26, 2014), [http://www.nytimes.com/2014/10/27/business/media/how-facebook-is-changing-the-way-its-users-consume-journalism.html?\\_r=0](http://www.nytimes.com/2014/10/27/business/media/how-facebook-is-changing-the-way-its-users-consume-journalism.html?_r=0).

37. Ethan Kross et al., *Facebook Use Predicts Declines in Subjective Well-Being in Young Adults*, 8 PLOS ONE e69841 (2013), <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0069841>.

emotional valence, as did the experimental algorithm.<sup>38</sup> But this pro-engagement algorithm, combined with the limited ways in which users can engage with posts to express negative reactions, likely results in News Feeds that are dominated by emotionally positive content. For instance, given that Facebook features a “like” button but, to date, has resisted calls to add a “dislike” feature,<sup>39</sup> to the extent that the algorithm prioritizes items that other users have previously interacted with, the algorithm may prioritize items with positive emotional valence.<sup>40</sup>

### C. Four Hypotheses about the Effects of News Feed

In the wake of the rapid emergence of online social media, several competing hypotheses emerged about the relationship between Facebook and well-being. Some studies find that Facebook use correlates with well-being,<sup>41</sup> while others find the opposite.<sup>42</sup> Others suggest it depends on one or more additional variables, like the user’s personality or how she uses Facebook.<sup>43</sup> Still other

---

38. If Facebook *has* deliberately configured the News Feed algorithm to maximize the emotionally positive content users see, this would hardly be surprising. After all, Facebook’s business model depends on user engagement on the platform, and intuition suggests that users will be relatively attracted to emotionally positive content (and relatively repelled by emotionally negative content). Of course, intuition is often a poor guide to such matters, as some research has suggested. *See infra* notes 171–79 and accompanying text.

39. Will Oremus, *You Can’t Dislike This Article*, SLATE (Dec. 15, 2014, 4:14 PM), [http://www.slate.com/articles/technology/future\\_tense/2014/12/facebook\\_dislike\\_but\\_ton\\_why\\_mark\\_zuckerberg\\_won\\_t\\_allow\\_it.html](http://www.slate.com/articles/technology/future_tense/2014/12/facebook_dislike_but_ton_why_mark_zuckerberg_won_t_allow_it.html).

40. Since August of 2012, Facebook has rolled out several features that allow users to negatively interact with News Feed items in other ways, such as by reporting an item as spam or in violation of Facebook’s “Community Standards,” instructing Facebook to “hide” an item from their feed, and instructing Facebook to show them fewer items posted by a particular friend. But the algorithm *deprioritizes* items that have received such user or network interactions, so that too seems likely to lead to disproportionately positive News Feeds. Vaughan, *supra* note 4.

41. Sebastian Valenzuela, Namsu Park & Kerk F. Kee, *Is There Social Capital in a Social Network Site?: Facebook Use and College Students’ Life Satisfaction, Trust, and Participation*, 14 J. COMPUTER-MEDIATED COMM. 875 (2009).

42. Chiungjung Huang, *Internet Use and Psychological Well-Being: A Meta-Analysis*, 13 CYBERPSYCHOL., BEHAV., & SOC. NETWORKING 241 (2010); Hui-Tzu, Grace Chou & Nicholas Edge, *‘They Are Happier and Having Better Lives than I Am’: The Impact of Using Facebook on Perceptions of Others’ Lives*, 15 CYBERPSYCHOL., BEHAV., & SOC. NETWORKING 117 (2012).

43. *See* Amanda L. Forest & Joanne V. Wood, *When Social Networking Is Not Working: Individuals with Low Self-Esteem Recognize but Do Not Reap the Benefits of Self-Disclosure on Facebook*, 23 PSYCHOL. SCI., 295, 295–302 (2012); Adriana M. Manago, Tamara Taylor & Patricia M. Greenfield, *Me and My 400 Friends: The Anatomy of College Students’ Facebook Networks, Their Communication Patterns, and Well-Being*, 48 DEV. PSYCHOL. 369, 369–80 (2012); Junghyun Kim, Robert LaRose & Wei Peng, *Loneliness as the Cause and the Effect of Problematic Internet Use: The Relationship Between Internet*

researchers dismiss all of this data as mere “noise.” And for each camp, there is empirical research to back them up.<sup>44</sup>

Two hypotheses in particular are relevant to the “emotional contagion” experiment. First, several academic studies have found that increased Facebook use correlates with increases in a variety of negative psychological conditions, including stress, jealousy, loneliness, and depression. The unflattering theory that has emerged—called “social comparison”—is that exposure to the happiness of others on Facebook (whether that happiness is genuine or contrived) depresses users by making them feel worse about their own lives.<sup>45</sup> These studies, however, were generally small, observational,<sup>46</sup> cross-sectional, and based on self-reporting by subjects of their mood.

Meanwhile, a second line of research has found that emotional states can spread in social networks, much like viruses. That phenomenon is known as “emotional contagion”—“the tendency to automatically mimic and synchronize expressions, vocalizations, postures, and movements with those of another person[] and, consequently, to converge emotionally.”<sup>47</sup> In other words, physically interacting with happy people will tend to make you happy, while similar exposure to unhappy people will tend to make you unhappy.

---

*Use and Psychological Well-Being*, 12 CYBERPSYCHOL. & BEHAV.: THE IMPACT OF THE INTERNET, MULTIMEDIA & VIRTUAL REALITY ON BEHAV. & SOC'Y 451, 451-55 (2009).

44. For a good lay overview of the “starkly divided” empirical literature about the psychological effects of online social media in general, and of Facebook in particular, in which “opposite argument[s] [are] equally prominent,” see Maria Konnikova, *Why Facebook Is Making Us Unhappy*, THE NEW YORKER (Sept. 10, 2013), <http://www.newyorker.com/tech/elements/how-facebook-makes-us-unhappy>.

45. See, e.g., SHERRY TURKLE, ALONE TOGETHER: WHY WE EXPECT MORE FROM TECHNOLOGY AND LESS FROM EACH OTHER (2011); Kross et al., *supra* note 37; Edson C. Tandoc, Patrick Ferrucci & Margaret Duffy, *Facebook Use, Envy, and Depression Among College Students: Is Facebooking Depressing?*, 43 COMPUTERS IN HUM. BEHAV. 139 (2015); Hanna Krasnova, et al., *Envy on Facebook: A Hidden Threat to Users' Life Satisfaction?* (2013), available at [https://ara.cat/xarxes/facebook\\_ARAFIL20130128\\_0001.pdf](https://ara.cat/xarxes/facebook_ARAFIL20130128_0001.pdf).

46. An observational study draws conclusions on the basis of correlations among two or more variables of interest (say, varying amounts of time spent by different subjects on Facebook and the extent to which subjects report negative or positive effects). Unlike an interventional study, in which subjects are randomized to different conditions (say, using Facebook a lot, a little, or not at all during the study period), an observational study simply records the apparent effects of the behaviors in which subjects naturally choose to engage. This design does not enable researchers to reliably infer that variable A caused variable B; correlations between two variables are also consistent with the hypothesis that variable B caused variable A or that a third, unobserved variable caused both A and B. For instance, it could be that certain personality traits cause people with those traits both to spend more time on Facebook and to report negative affect.

47. Elaine Hatfield, John T. Cacioppo & Richard L. Rapson, *Emotional Contagion*, 2 CURRENT DIRECTIONS PSYCHOL. SCI. 96, 96 (1993).

Large longitudinal field studies have found that happiness,<sup>48</sup> depression,<sup>49</sup> and loneliness<sup>50</sup> strongly correlate with the presence of these effects in one's face-to-face social network, up to three degrees of separation (one's friends' friends' friends).

But these studies, too, were correlational.<sup>51</sup> Moreover, they were limited to in-person social networks, where emotion was hypothesized to spread via exposure to others' movements, facial expressions, postures, and vocalizations, not through online networks, where the "vector" through which emotions spread would be emotionally evocative text. Indeed, researchers who conducted an exceptionally large, longitudinal study of correlations between mobile social networks and technology adoption behavior found that most of that correlation was caused by homophily, the tendency to associate with people who are already like you, rather than contagion, the tendency to take on the traits of those around you.<sup>52</sup>

If the social comparison hypothesis is correct, then many Facebook posts with a positive emotional valence are psychologically risky for users who see them, since the items could cause negative emotions. On the other hand, if the emotional contagion hypothesis is correct and extends to online social networks, then negative Facebook items are risky, since they may spread negative emotion, while positive items might actually carry a psychological benefit by spreading positive affect.

#### D. The Facebook "Emotional Contagion" Experiment

To determine the effects on users of positive and negative

---

48. James H. Fowler & Nicholas A. Christakis, *Dynamic Spread of Happiness in a Large Social Network: Longitudinal Analysis Over 20 Years in the Framingham Heart Study*, 337 BRIT. MED. J. a2338 (Dec. 5, 2008), available at <http://www.bmj.com/content/337/bmj.a2338>.

49. J.N. Rosenquist, J.H. Fowler & N.A. Christakis, *Social Network Determinants of Depression*, 16 MOLECULAR PSYCHIATRY 273 (2011).

50. John T. Cacioppo, James H. Fowler & Nicholas A. Christakis, *Alone in the Crowd: The Structure and Spread of Loneliness in a Large Social Network*, 97 J. PERSONALITY & SOC. PSYCHOL. 977 (2009).

51. See, e.g., Ethan Cohen-Cole & Jason M. Fletcher, *Is Obesity Contagious? Social Networks vs. Environmental Factors in the Obesity Epidemic*, 27 J. HEALTH ECON. 1382 (2008) (noting, of a similar longitudinal study by Christakis and Fowler of the spread through social networks of obesity, that there are at least three other explanations, besides contagion, for why traits like obesity—or happiness—cluster in reference groups over time).

52. Sinan Aral, Lev Muchnik & Arun Sundararajan, *Distinguishing Influence-Based Contagion from Homophily-Driven Diffusion in Dynamic Networks*, 106 PROC NAT'L ACAD. SCI. 21544 (2009) (examining a global instant messaging network of more than 27.4 million users and finding that "previous methods over-estimate peer influence in product adoption decisions in this network by 300–700%").

words in News Feeds, Facebook data scientist Adam Kramer—with Cornell professor of communications and information science Jeffrey Hancock and his graduate student, Jamie Guillory—designed an experiment.<sup>53</sup> For one week in 2012, Facebook applied an additional algorithm to the News Feeds of 689,003 users randomly selected (by Facebook ID number) from among those who view Facebook in English. For approximately 155,000 of these users, Facebook removed varying proportions (between ten and ninety percent) of News Feed posts containing one or more positive-sounding words.<sup>54</sup> For another 155,000 or so users, Facebook removed varying proportions (again, between ten and ninety percent) of posts containing one or more negative-sounding words.<sup>55</sup> Each treatment condition was compared to a control condition in which the same proportion of posts was filtered out randomly (i.e., without regard to the emotional valence of their content) for the same number of users.<sup>56</sup>

Whether a post contained a “positive” or “negative” word was determined by automated text analysis software;<sup>57</sup> researchers had no access to users’ posts. A News Feed item did not have to have a strong overall emotional valence for the software to code it as positive or negative: if a post contained a single word that the software categorizes as positive (such as “love,” “nice,” or “sweet”) or negative (such as “hurt,” “ugly,” or “nasty”), it was coded accordingly.<sup>58</sup> As always, users could see non-prioritized eligible items by going to the relevant friend’s Timeline or group Page, and items filtered out of a user’s News Feed for one viewing may have appeared in a subsequent viewing.

As Kramer would later explain:

The reason we did this research is because we care about the emotional impact of Facebook and the people that use our product. We felt that it was important to investigate the common

---

53. Kramer et al., *supra* note 12. Technically, two parallel experiments were conducted, each with its own control group, because, pre-experiments, eligible News Feed items did not contain an equal number of positive and negative items. The composition of eligible items was 46.8% positive posts and 22.4% negative posts.

54. *Id.* at 8789.

55. *Id.*

56. *Id.*

57. *Id.* (“Posts were determined to be positive or negative if they contained at least one positive or negative word, as defined by . . . [Researchers used Linguistic Inquiry and Word Count (LIWC) 2007]”).

58. *Id.*; *Table 1: LIWC2007 Output Variable Information*, LINGUISTIC INQUIRY & WORD COUNT, <http://www.liwc.net/descriptiontable1.php> (last visited March 16, 2015) (listing these words among those in LIWC2007’s “positive emotion” and “negative emotion” categories, respectively).

worry that seeing friends post positive content leads to people feeling negative or left out. At the same time, we were concerned that exposure to friends' negativity might lead people to avoid visiting Facebook.<sup>59</sup>

The researchers characterize their results as evidence for the emotional contagion hypothesis and against the social comparison hypothesis.<sup>60</sup> Compared to control subjects, subjects exposed to fewer positive posts used 0.1 percent fewer positive words and 0.04 percent more negative words in their own subsequent posts, and produced only 96.7 percent as many words overall.<sup>61</sup> Compared to control subjects, subjects exposed to fewer negative posts used 0.07 percent fewer negative words and 0.6 percent more positive words in their own subsequent posts and produced only 99.7 percent as many words overall.<sup>62</sup>

These effects, while statistically significant, are extremely small, and discernible at all only because the sample size was so large. As Kramer later characterized them, "people produced an average of one fewer emotional word, per thousand words, over the following week."<sup>63</sup> What was significant about the results was not the size of these miniscule effects, but their direction: if the researchers' interpretation of their results is to be believed,<sup>64</sup> on Facebook, positivity begets positivity—not, as some had worried, negativity.

## II. FRAME ONE: HUMAN SUBJECTS RESEARCH

The main line of criticism about the experiment by both the public and scholars can be summarized as follows<sup>65</sup>: The experiment

---

59. Adam D.I. Kramer, FACEBOOK (June 29, 2014, 2:05 PM), <https://www.facebook.com/akramer/posts/10152987150867796>.

60. Kramer et al., *supra* note 25, at 8790.

61. *Id.*

62. *Id.*

63. Adam D.I. Kramer, *supra* note 59.

64. LIWC 2007 is not intended for lengthy text, like some Facebook posts. Moreover, the instrument cannot handle quirks of linguistics, such as sarcasm, negatives, and slang. The following Facebook posts—"Oh great," "I'm not having a great day," and "That's sick!"—likely would have been incorrectly coded. Finally, it is not obvious that very slight changes in users' word choices entail any changes in users' emotions. A plausible alternative explanation is that seeing friends post negative (or positive) things frees the user to express the negative feelings she already had.

65. A secondary criticism was procedural rather than substantive—namely, that no IRB reviewed the study (other than whatever internal Facebook review procedure existed at the time). Federal law did not require such review—see Michelle N. Meyer, *Everything You Need To Know About Facebook's Controversial Emotion Experiment*, WIRED (June 30, 2014), <http://www.wired.com/2014/06/everything-you-need-to-know-about-facebooks-manipulative-experiment/>—but whether some sort of prospective review of the study should have occurred as a matter of ethics and sound policy is another matter,

was unethical because Facebook and Cornell researchers (1) intentionally psychologically harmed subjects,<sup>66</sup> (2) without their consent, (3) thereby abusing their power over users,<sup>67</sup> treating them as mere means to Facebook's corporate ends or the ends of "science," and depriving them of information necessary for them to make a considered judgment about what was in their best interests—and all this (4) in order to make more money for Facebook, to satisfy researchers' intellectual curiosity about an abstract scientific question,<sup>68</sup> or "just to see what happens."<sup>69</sup>

#### A. *Subjects Gave No Ethically Meaningful Consent to the Experiment*

The second premise of this critique—that the consent of subjects was not obtained—seems true (or so I shall conclude by the end of this section, in any case). The philosophical, ethical, and legal literature on consent is vast and I cannot review it here. But on one plausible account, the immediate purpose,<sup>70</sup> ethically speaking, of

one I bracket in this article in order to focus on substantive objections.

66. See, e.g., Katy Waldman, *Facebook's Unethical Experiment*, SLATE (June 28, 2014), [http://www.slate.com/articles/health\\_and\\_science/science/2014/06/facebook\\_unethical\\_experiment\\_it\\_made\\_news\\_feeds\\_happier\\_or\\_sadder\\_to\\_manipulate.html](http://www.slate.com/articles/health_and_science/science/2014/06/facebook_unethical_experiment_it_made_news_feeds_happier_or_sadder_to_manipulate.html) ("intentionally made thousands upon thousands of people sad"); Alex Hern, *Facebook deliberately made people sad. This ought to be the final straw*, GUARDIAN (June 30, 2014), <http://www.theguardian.com/commentisfree/2014/jun/30/facebook-sad-manipulating-emotions-socially-responsible-company>; Electronic Privacy Information Center, *In re: Facebook (Psychological Study)*, EPIC (July 17, 2014), <https://epic.org/privacy/internet/ftc/facebook/psycho/> ("purposefully messed with people's minds").

67. See, e.g., Laurie Penny, *Facebook Can Manipulate Your Mood. It Can Affect Whether You Vote. When Do We Start To Worry?*, NEWSTATESMAN (June 30, 2014), <http://www.newstatesman.com/internet/2014/06/facebook-can-manipulate-your-mood-it-can-affect-whether-you-vote-when-do-we-start> ("Nobody has ever had this sort of power before. No dictator in their wildest dreams has been able to subtly manipulate the daily emotions of more than a billion humans so effectively. There are no precedents for what Facebook is doing here. Facebook itself is the precedent. What the company does now will influence how the corporate powers of the future understand and monetise human emotion. . . . If Facebook is a country, then it is a corporate dictatorship. This is not a metaphor.").

68. Kashmir Hill, *Facebook Added 'Research' To User Agreement 4 Months After Emotion Manipulation Study*, FORBES (June 30, 2014), <http://www.forbes.com/sites/kashmirhill/2014/06/30/facebook-only-got-permission-to-do-research-on-users-after-emotion-manipulation-study/> (referring to users as "guinea pigs made to have a crappy day for science").

69. See, e.g., Jaron Lanier, *Should Facebook Manipulate Users?*, N.Y. TIMES (June 30, 2014), [http://www.nytimes.com/2014/07/01/opinion/jaron-lanier-on-lack-of-transparency-in-facebook-study.html?\\_r=0](http://www.nytimes.com/2014/07/01/opinion/jaron-lanier-on-lack-of-transparency-in-facebook-study.html?_r=0) (likening the Facebook experiment to a "pharmaceutical firm . . . randomly, secretly sneak[ing] an experimental drug . . . into the drinks of hundreds of thousands of people, just to see what happens").

70. Such notice and opportunity, in turn, may serve a number of deeper purposes,



obtaining someone's consent is to provide notice of a planned action affecting that individual and an opportunity for her to agree to that action or reject it. No one has suggested that in the Facebook case, subjects received any notice whatsoever of, nor any opportunity to agree or decline to participate in, this *particular* study.

But to be ethically meaningful, consent need not always be fully informed or contemporaneous. Sometimes, a disclosure of limited information about a proposed action before soliciting agreement suffices ethically or legally, as may a one-time, "blanket" consent to be subject to categories of action in the future.<sup>71</sup> Terms of Service and similar agreements sometimes serve this purpose, for example. But Facebook's Data Use Policy in effect at the time of the experiment, to which users must agree when they sign up for a Facebook account, made no reference to research, as critics were quick to point out.<sup>72</sup>

On the other hand, that Data Use Policy did tell users that the company might use their data "as part of our efforts to keep Facebook products, services and integrations safe and secure."<sup>73</sup> As I will argue below, the experiment in fact had the potential to serve

---

such as respecting an individual's right to self-determination (a deontological end) or promoting individuals' welfare by allowing those best positioned to decide whether participation in an event will further or set back their interests—the affected individuals themselves—to choose (a consequentialist end).

71. Blanket consent is commonly used in biobank research, for example, because it is difficult for researchers to know in advance what research questions the stored tissue samples (or the genomic data set derived therefrom) may be used to study in the future and re-contacting and re-consenting subjects to each new use may be infeasible or intrusive. Emerging evidence also suggests that in some scenarios, subjects are comfortable with blanket consent. See Susan E. Kelly et al., *Evaluating the Consent Preferences of UK Research Volunteers for Genetic and Clinical Studies*, 10 PLOS ONE e0118027 (2015). Federal regulators have recently proposed amending human subjects research regulations to permit biospecimens researchers to use "a brief standard consent form" in which "consent need not be study-specific, and could cover open-ended future research." Human Subjects Research Protections: Enhancing Protections for Research Subjects and Reducing Burden, Delay, and Ambiguity for Investigators, 76 Fed. Reg. 44512, 44515 (proposed July 26, 2011) (to be codified at 45 C.F.R. pts. 46, 160, 164 and 21 C.F.R. pts. 50, 56).

72. Hill, *supra* note 68. The revised version of Facebook's Data Use Policy released some four months after the "emotional contagion" experiment adds that user data may be used "for internal operations, including troubleshooting, data analysis, testing, research and service improvement." *Id.* The researchers themselves offer a curious argument for informed consent, noting that "no text was seen by the researchers. As such, it was consistent with Facebook's Data Use Policy, to which all users agree prior to creating an account on Facebook, constituting informed consent for this research." Kramer et al., *supra* note 12, at 8789. The fact that the study did not violate agreed-upon terms pertaining to data privacy does not mean that subjects gave affirmative informed consent to participate in any study.

73. Hill, *supra* note 68.

the end of ensuring that Facebook is (psychologically) safe.<sup>74</sup> But did this Data Use Policy statement serve the ethical purposes of consent—to put users signing up for Facebook accounts on notice that by agreeing to the policy, they were agreeing to participate in something like the emotional contagion study, and to give them the opportunity (by declining to open an account) to refuse to participate? There are two reasons to be very skeptical that it did.

First, it is well known that very few people read click-through user agreements.<sup>75</sup> This point was brought memorably home in one recent study, which resulted in several Londoners, in exchange for free Wi-Fi service, giving click-through “consent” to forfeit their first-born child to the service provider.<sup>76</sup>

That most Facebook subjects likely did not read the Data Use Policy might not spell the end of the case for there having been *some* sort of ethically meaningful consent to participate in the emotional contagion study. Even if a user fails to read particular end-user licensing agreements (EULA) or terms of service (ToS) documents, by giving click-through acknowledgement that such terms exist and proceeding to use the product or service, they may have implicitly consented to any terms that were consistent with a reasonable user’s expectations. For instance, in one survey of online gamers—which, not surprisingly, found that the vast majority of users provide click-through agreement to terms they have not read—one respondent explained, “While I don’t read the EULA or ToS, I expect that they have the right to run the game, change it and do what they need to keep it growing.”<sup>77</sup> Gaming companies and social media companies are just that—companies—and even unsavvy users, if they are reasonable, should anticipate that companies will behave in ways that are consistent with their for-profit status, such as declaring

---

74. For a discussion of the distinction between researchers’ motives and the ends that the experiment served (whether or not they were intended by researchers), see *infra* note 139.

75. See, e.g., Victoria C. Plaut & Robert P. Bartlett, *Blind Consent? A Social Psychological Investigation of Non-Readership of Click-Through Agreements*, 36 L. HUM. BEHAV. 293 (2007).

76. Rachel Feltman, *Londoners Accidentally Pay for Free Wi-Fi with a Firstborn, Because No One Reads Anymore*, WASH. POST (Sept. 29, 2014), <http://www.washingtonpost.com/news/speaking-of-science/wp/2014/09/29/londoners-accidentally-pay-for-free-wi-fi-with-a-firstborn-because-no-one-reads-anymore/>. Incidentally, this is a good example of a behavioral study that could not have been conducted with subjects’ informed consent, a point I shall return to below.

77. Suzanne de Castell, “I AGREE”: *Informed Consent and the Ethics of Third Party Access to Game Player Data*, IPOSGOODE (Apr. 23, 2011), <http://www.iposgoode.ca/2011/04/i-agree-internet-research-informed-consent/#sthash.ZjnwUhh.dpuf>.

certain property rights in their products or services, periodically and unilaterally redesigning those products and services (within limits), and taking steps to assure or improve the quality of their products or services, including ensuring that they do not risk users' safety or security. On these grounds it might be argued that Facebook users implicitly consented to participate in research designed "to keep Facebook products, services and integrations safe and secure" by investigating credible claims that positive or negative News Feed posts are psychologically harmful.

But this brings us to the second reason to be skeptical that we can derive ethically meaningful consent from users' agreement to the Data Use Policy, notwithstanding that it, unlike the London Wi-Fi contract, contained no unconscionable Herod clause but instead disclosed a purpose that any reasonable user would expect to find there: implicitly consenting to a company's broad end does not entail consenting to any means a company might use to achieve that end. For example, one reasonable user expectation might be that Facebook would take steps to learn more about, and seek to prevent, cyber-bullying on its platform. One means of achieving that end would be for Facebook researchers to "friend" users through dummy accounts, bully them, randomize half to a treatment condition in which they are given access to a novel means of blocking or responding to such abuse, and compare their outcomes to those who lacked access. Surely, implicitly consenting to Facebook's broad end of keeping the platform safe does not entail consenting to this means of pursuing that end.

If the first premise in the critique of the Facebook experiment were true—if the means that the Facebook-Cornell researchers employed to achieve their reasonable end of ensuring that users are psychologically safe was to intentionally psychologically harm subjects—then implicitly consenting to Facebook's end by giving click-through agreement and using the product cannot plausibly be said to entail implicit consent to the means. No reasonable user expects that being intentionally psychologically harmed by the host company is the going price for access to a free social media platform.

As I argue in the next section, the first premise is not true: the experiment cannot fairly be characterized as a knowing imposition of psychological harm on subjects by researchers. Still, the means that Facebook used to pursue the reasonable end of ensuring user safety and enjoyment—filtering out News Feeds items on the basis of the emotional valence of the words they contain and randomizing users—clearly was not expected by most users, whether or not those

means were innocuous. Although A/B testing of websites is ubiquitous,<sup>78</sup> the reaction to the Facebook experiments made clear that most of the public had been unaware of this phenomenon. Similarly, although News Feed has always been curated by an algorithm, many users apparently did not realize that, and fewer still likely anticipated that posts might ever be filtered out on the basis of content (as opposed to user engagement metrics), much less on the basis of positive and negative content.

If we are to conclude that users gave meaningful consent to unread terms at all, it must be because we believe that they had actual or constructive notice of the content of those terms and, having clicked through them and proceeded to use the service, tacitly agreed to them. But it is difficult to believe that reasonable users did or should have known about A/B testing of content-based News Feed algorithms when so many users were clearly surprised by some or all of these elements. And so it seems implausible that the vast majority of users had actual or constructive notice that in signing up for and using a Facebook account, they were agreeing to participate in something like the emotional contagion experiment.

*B. Why Consent Is Not Always an Ethical Requirement of Human Subjects Research*

But of course not all activities that affect others require consent, as a matter of either law or ethics. The requirement to obtain subjects' voluntary, informed consent is the default rule in both law and ethics. Yet neither the federal regulations governing much (but not all)<sup>79</sup> human subjects research—the Common Rule—nor the ethical principles on which the Common Rule is based require informed consent for all human subjects research.<sup>80</sup> That is because

---

78. The odds that a Facebook user has been a subject in *some* experiment on the platform are “100%,” according to Facebook data scientists, and at any given time, the average user is a subject in approximately 10 experiments. *The Trust Engineers*, RADIOLAB (Feb. 9, 2015), <http://www.radiolab.org/story/trust-engineers/> (relevant portion may be accessed at 15:30-15:38).

79. For discussion of the history and scope of these regulations, see Michelle N. Meyer, *Regulating the Production of Knowledge: Research Risk-Benefit Analysis and the Heterogeneity Problem*, 65 ADMIN. L. REV. 237, 243–50 (2013).

80. Even some prominent bioethicists appear to have forgotten this. See, e.g., Arthur Caplan & Charles Seife, *Facebook Experiment Used Silicon Valley Trickery*, NBCNEWS.COM, (June 30, 2014), <http://www.nbcnews.com/health/mental-health/opinion-facebook-experiment-used-silicon-valley-trickery-n144386> (“[T]he experiment should never have been performed. It is a violation of the rights of research subjects . . . The question of whether or not an experiment is ethical hinges upon the question of ‘informed consent.’”); Robert Klitzman, *Did Facebook’s Experiment Violate Ethics?*, CNN (July 2, 2014), <http://www.cnn.com/2014/07/02/opinion/klitzman-facebook-experiment/> (“According to these regulations [the Common Rule], all research must respect the rights

research ethics is informed not only by the principle of respect for persons' autonomy, but also by the principles of beneficence and justice. Balancing these principles yields sensible exceptions to the requirement of informed consent.

The Common Rule was the culmination of a process that began with congressional hearings following public outcry over a series of research scandals. After those hearings, Congress passed the National Research Act of 1974.<sup>81</sup> The Act, *inter alia*, established the ad hoc National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research to "identify the basic ethical principles which should underlie the conduct of" human subjects research, and to recommend regulations that embody these principles.<sup>82</sup>

The principles the Commission took to underlie the ethical conduct of human subjects research are most famously laid out in the *Belmont Report*.<sup>83</sup> That report articulates a rule that researchers obtain subjects' informed consent, and explains that this rule stems from the principle of respect for persons, including treating autonomous individuals as agents.<sup>84</sup> But the report is also careful to note that informed consent can only be a default rule: "In most cases of research involving human subjects, respect for persons demands that subjects enter into the research voluntarily and with adequate information. In some situations, however, application of the

---

of individual research subjects, and scientific investigators must therefore explain to participants the purposes of the study, describe the procedures (and which of these are experimental) and 'any reasonably foreseeable risks or discomforts.' Facebook followed none of these mandates."); Robert Klitzman, *Why Facebook Should Follow Ethical Standards—Like Everybody Else*, HUFFINGTON POST (July 7, 2014), [http://www.huffingtonpost.com/robert-klitzman-md/why-facebook-needs-to-fol\\_b\\_5557862.html](http://www.huffingtonpost.com/robert-klitzman-md/why-facebook-needs-to-fol_b_5557862.html) (claiming that Facebook social scientists violated their disciplines' codes of professional ethical standards "that include stipulations that researchers obtain appropriate informed consent" and citing Am. Psych. Assoc., *Ethical Principles of Psychologists and Code of Ethical Conduct*, Standard 8 (2010), available at <http://www.apa.org/ethics/code/index.aspx?item=11>, which in fact permits "dispensing with informed consent for research" under certain circumstances, including when the Common Rule allows this, *id.* § 8.05(2), as well as deception).

81. See National Research Act of 1974, Pub. L. No. 93-348, 88 Stat. 342 (1974).

82. *Id.* at § 202(a)(1)(A).

83. See U.S. DEPT. OF HEALTH & HUMAN SERVS., NAT'L COMM. FOR THE PROT. OF HUM. SUBJECTS OF BIOMED. & BEHAV. RES., THE BELMONT REPORT: ETHICAL PRINCIPLES AND GUIDELINES FOR THE PROT. OF HUM. SUBJECTS OF RES. (1978), <http://www.hhs.gov/ohrp/humansubjects/guidance/belmont.html> [hereinafter BELMONT REPORT]; see also U.S. DEPT. OF HEALTH & HUM. SERVS., NAT'L COMM. FOR THE PROT. OF HUM. SUBJECTS OF BIOMED. & BEHAV. RES., *Report & Recommendations: Institutional Review Boards* (1978) (translating the Belmont principles into proposed regulations governing IRBs that the Common Rule essentially codified).

84. BELMONT REPORT, *supra* note 83. Respect for persons also requires that the welfare of those who are not autonomous be protected. *Id.*

principle is not obvious.”<sup>85</sup>

The most common reason for departing from the default rule of informed consent is when respect for persons’ autonomy conflicts with a second principle, beneficence: “Persons are treated in an ethical manner not only by respecting their decisions and protecting them from harm, but also by making efforts to secure their well-being.”<sup>86</sup> Although human subjects research is often and lamentably viewed as offering prospective subject volunteers only risks and costs, such that beneficence might be thought to automatically weigh against research participation, the *Belmont Report* correctly observes to the contrary that beneficence “often occupies a well-defined *justifying* role in many areas of research involving human subjects.”<sup>87</sup> This is because beneficence requires both “that we protect against risk of harm to subjects and also that we be concerned about the loss of the substantial benefits that might be gained from research,”<sup>88</sup> as many groups of would-be research subjects, most notably gay men<sup>89</sup> and pregnant women,<sup>90</sup> have forcefully argued. Of particular relevance to the Facebook case, the report notes that “[r]esearch . . . makes it possible to avoid the harm that may result from the application of previously accepted routine practices that on closer investigation turn out to be dangerous.”<sup>91</sup>

Hence, the rule of informed consent is more appropriately phrased: Don’t “withhold information necessary to make a considered judgment, *when there are no compelling reasons to do so.*”<sup>92</sup> The *Belmont Report* suggests that one compelling reason to withhold information from subjects is “where informing subjects of some pertinent aspect of the research is likely to impair the validity of the research.”<sup>93</sup> The report further specifies the conditions under which withholding information from subjects can be ethical, and these were codified in the Common Rule as follows:

An IRB may approve a consent procedure which does not include,

---

85. *Id.*

86. *Id.*

87. *Id.* (emphasis added).

88. *Id.*

89. For instance, within two years of its formation in 1987, the community activist group AIDS Coalition to Unleash Power (ACT-UP) successfully effected changes to FDA policy, accelerating the process for testing and approving drugs. Oliver Morton, *Achievements in Research*, in *AIDS IN THE WORLD* 229, 289 (Jonathan M. Mann et al. eds., 1992).

90. See R. Alta Charo, *Protecting Us to Death: Women, Pregnancy, and Clinical Research Trials*, 38 ST. LOUIS U. L.J. 135 (1993).

91. BELMONT REPORT, *supra* note 83.

92. *Id.* (emphasis added).

93. *Id.*

or which alters, some or all of the elements of informed consent set forth in this section, or waive the requirements to obtain informed consent provided the IRB finds and documents that:

1. The research involves no more than minimal risk to the subjects;
2. The waiver or alteration will not adversely affect the rights and welfare of the subjects;
3. The research could not practicably be carried out without the waiver or alteration; and
4. Whenever appropriate, the subjects will be provided with additional pertinent information after participation.<sup>94</sup>

The regulations, and the *Belmont Report* on which they are based, thus recognize that autonomy is not the only value worth preserving. We should and do care about human welfare as well. When the best evidence available at the time of a proposed study suggests that any additional risk it imposes on subjects is minimal, and when the study could not otherwise practicably be done, it may qualify for an alteration or even waiver of the informed consent that riskier studies require.

This is not a regulatory loophole, but a reasonable, indeed crucial, ethical principle that seeks to balance liberty and welfare rather than fetishizing one over the other. Some of our most profound and practically relevant insights into human behavior could not have been realized without deviating from the default rule of fully informed consent—whether through non-consensual research (in which subjects give no consent at all), deceptive research (in which subjects consent to participate in a study that researchers falsely characterize in potentially material ways), or incompletely informed research (in which subjects agree to participate without being informed of information potentially material to their decision whether to participate). Important lines of research that have depended on departures from fully informed consent include: research on the bystander effect, which attempted to determine why so many people failed to act while Kitty Genovese was being murdered;<sup>95</sup> the effects of social pressure on the distortion

---

94. 45 C.F.R. § 46.116(d) (2015).

95. See, e.g., John M. Darley & Binn Latané, *Bystander Intervention in Emergencies: Diffusion of Responsibility*, 8 J. PERSONALITY & SOC. PSYCH. 377 (1968), (describing bystander effect research in which subjects were falsely told they were participating in a study of personal problems faced by college students, then subjects' reactions were

and modification of judgments;<sup>96</sup> implicit racial and other forms of bias;<sup>97</sup> false memories,<sup>98</sup> change blindness,<sup>99</sup> and inattentional

---

observed and recorded after a confederate “subject” pretended to suffer an anxiety-produced seizure).

96. See, e.g., Solomon E. Asch, *Opinions and Social Pressure*, 193 SCI. AM. 31 (1955) (describing social pressure research in which subjects were told they were participating in a study of perception and would be tested alongside several other subjects, all of whom were in fact confederates of the researchers; when asked to compare the lengths of two lines, confederates gave obviously incorrect answers to test whether this would affect subjects’ responses).

97. See, e.g., REDZO MUJIC & PAUL FRIJTERS, STILL NOT ALLOWED ON THE BUS: IT MATTERS IF YOU’RE BLACK OR WHITE! (2014), available at [http://islandia.law.yale.edu/ayres/mujcic\\_frijters\\_busDec2014.pdf](http://islandia.law.yale.edu/ayres/mujcic_frijters_busDec2014.pdf). This was racial bias research, conducted by Australian economists, in which trained confederates (“testers,” as the researchers refer to them) boarded a city bus, pretended to discover that their fare card was empty, and requested a free ride to their destination from the driver. Both black and white drivers were significantly more likely to offer a free ride when the confederate was white rather than black. Such testing should be familiar to historians of antidiscrimination law; they have long been used by government agencies and private groups to ferret out violators of antidiscrimination laws. *But see also* Nikole Hannah-Jones, *No Sting: Feds Won’t Go Undercover to Prove Housing Discrimination*, PRO PUBLICA (Dec. 20, 2012), <http://www.propublica.org/article/no-sting-feds-wont-go-undercover-to-prove-housing-discrimination> (reporting and lamenting recent decline in systematic testing by U.S. Department of Housing and Urban Development).

The click-through online consent form for participation in at least one version of the Implicit Association Test (IAT) taken by millions of subjects describes the IAT as a test that will “examine your ideas, beliefs, and opinions about different topics. You will answer some questions and take an IAT in which you will sort words into categories as quickly as possible.” *Consent Agreement: Implicit Social Cognition on the Internet*, PROJECT IMPLICIT, <https://implicit.harvard.edu/implicit/>, retrieved Feb. 15, 2015 (copy on file with author; access to Consent Agreement requires email registration). In that form, there is no mention of the specific purpose of the study, which is to investigate whether the subject implicitly holds highly stigmatized racist, sexist and similar associations (e.g., that the subject implicitly associates African-Americans with “rotten” or women with “irrational”). The consent form states that there are “no anticipated risks to you from participation,” *id.*, although it seems likely that the study poses a risk to subjects of receiving upsetting (and easily misunderstood) information about themselves. Indeed, UnderstandingPrejudice.org, “a web site for students, teachers, and others interested in the causes and consequences of prejudice,” recommends that teachers assign students to take the IAT, but warns: “Because the IAT may reveal information that students do not want to know about themselves, instructors should offer an alternative assignment for students who would rather not take the IAT.” See *The Implicit Association Test (Gender Version)*, UNDERSTANDING PREJUDICE <http://www.understandingprejudice.org/teach/assign/iatgend.htm> (last visited May 2 2015). However, telling subjects in advance that researchers are testing them for racist or sexist implicit associations may bias the results by affecting which people agree to participate or, perhaps, by altering subjects’ test performance.

98. See, e.g., Julia Shaw & Stephen Porter, *Constructing Rich False Memories of Committing Crime*, PSYCHOL. SCI. (2015), available at [https://people.ok.ubc.ca/stporter/Welcom\\_files/Psychological%20Science-2015-Shaw-0956797614562862.pdf](https://people.ok.ubc.ca/stporter/Welcom_files/Psychological%20Science-2015-Shaw-0956797614562862.pdf) (published online ahead of print) (describing false memory research in which, during a series of deceptions over several weeks, researchers succeeded in convincing seventy percent of subjects that, in their youth, they had committed a crime—assault, assault with a weapon, or theft—that resulted in police



blindness<sup>100</sup>—all of which have important implications for, *inter alia*, eyewitness testimony, false confessions, and other aspects of criminal law; and the most effective ways to encourage people to save for retirement,<sup>101</sup> conserve energy,<sup>102</sup> reduce littering,<sup>103</sup> and

contact).

99. See, e.g., Daniel J. Simons & Daniel T. Levin, *Failure to Detect Changes to People During a Real-World Interaction*, 5 PSYCHONOMIC BULLETIN & REV. 644 (1998), available at <https://www.msu.edu/course/psy/802/altmann/802/Ch2-4a-SimonsLevin98.pdf>. This was change blindness research in which one investigator approached pedestrians on a college campus to ask for directions. Confederates then passed between the investigator and the unwitting subjects carrying a door, during which a second investigator took the place of the initial investigator. The second investigator continued the conversation with the subject. Subjects were then asked, “Did you notice that I’m not the same person who approached you to ask for directions?” and were told about the purpose of the study.

100. See, e.g., Christopher F. Chabris, et al., *You Do Not Talk About Fight Club If You Do Not Notice Fight Club: Inattentional Blindness for a Simulated Real-World Assault*, 2 J. PERCEPTION 150 (2011), available at <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3485775/>. The inattentional blindness study was inspired by the 1995 case of Boston police officer Kenny Conley, who, while chasing a suspect, ran past an alley where an African-American undercover officer was being beaten by other officers who failed to recognize him as a fellow officer. In the subsequent federal investigation, Conley testified that he had not seen the beating. As jurors later explained, they found Conley’s explanation implausible and believed that he was covering for his fellow officers. He was convicted of perjury and obstruction of justice and sentenced to thirty-four months in jail. To test the hypothesis that, while Conley’s attention was focused on the suspect he was chasing, he was “blind” to the surprising event he otherwise would have easily seen, investigators recreated the scene. They recruited subjects to follow a jogger through a college campus, counting the number of times the jogger tapped his or her head. Confederates staged a fight alongside the jogging path, and subjects were later asked whether they had seen anything unusual and debriefed about the purpose of the study, prior disclosure of which would have rendered the study impossible to conduct. It seems unlikely that the information that researchers withheld from subjects would have been material to their decision about whether to participate, although it is conceivable that some subjects might have declined to contribute to researchers’ attempt to find an alternative explanation for what many in the public saw as a clear case of lying and racial cronyism.

101. See, e.g., JOHN BESHEARS ET AL., THE EFFECT OF PROVIDING PEER INFORMATION ON RETIREMENT SAVINGS DECISIONS (National Bureau of Economic Research Working Paper 17345, 2011), available at <http://www.nber.org/papers/w17345>. This was social norms marketing research used to investigate the effect of a peer information intervention on retirement savings choices. Subjects were employees of a “large manufacturing firm and its retirement savings plan administrator.” Subjects, who never learned that they were part of an experiment, were randomized to one of three conditions: receiving information about the savings behavior of coworkers in their five-year age bracket (e.g., employees at the firm between the ages of 20 and 24); receiving similar information about coworkers in their ten-year age bracket; and receiving a mailing that contained no peer information (the control group).

102. See, e.g., Noah J. Goldstein, Raymond R. Reno & Carl A. Kallgren, *A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels*, 35 J. CONSUMER RES. 472 (2008), available at <http://www.jstor.org/stable/pdf/10.1086/586910.pdf?acceptTC=true&jpdConfirm=true>.

103. See, e.g., Robert B. Cialdini, Raymond R. Reno & Carl A. Kallgren, *A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places*,

engage in myriad other individually and socially valuable behaviors.

*C. Was Informed Consent an Ethical Requirement of the Facebook Experiment?*

1. Infeasibility

Like much behavioral research, if the Facebook experiment had been conducted with users who had consented to participate after being fully informed about what behaviors the researchers were looking for and why and how they intended to elicit those responses, the results would have been badly biased, perhaps to the point of being deemed useless by the scientific community. Subjects' behavior would almost certainly have been altered by this knowledge (and in any case, this biasing effect could not be ruled out without comparing the results of two studies—one with fully informed consent and one, like the actual experiment, without it).

Moreover, although requiring informed consent always poses some risk of rendering the results less generalizable through selection bias, in this case it is especially likely that users who opted into the fully disclosed study would have been different from those who opted out, in ways that would have mattered for producing results that generalize to Facebook's 1.35 billion users.<sup>104</sup>

2. Incompletely Informed Consent

Professor James Grimmelman recently argued that although the Facebook experiment was “eminently eligible for an alteration” (though not a complete waiver) of informed consent, researchers should not “get out of informed consent altogether . . . At the very least, debriefing is completely appropriate.”<sup>105</sup> Whether the

---

58 J. PERSONALITY & SOC. PSYCH. 1015 (1990), *available at* [http://www-personal.umich.edu/~prestos/Downloads/DC/pdfs/Krupka\\_Oct13\\_Cialdinietal1990.pdf](http://www-personal.umich.edu/~prestos/Downloads/DC/pdfs/Krupka_Oct13_Cialdinietal1990.pdf).

104. For instance, the Facebook users who are skittish about participating in a research study may be the same users whose mental health is affected by exposure to emotional content on Facebook.

105. James Grimmelman, Professor of Law, Francis King Carey School of Law, Privacy Conf.: Panel Three—Ethical Standards for Human Subjects Research at Silicon Flatirons (Dec. 4, 2014), *available at* <https://www.youtube.com/watch?v=5-WDw8SZuTk>. In arguing that the Facebook study was “imminently eligible” for an alteration of informed consent, Professor Grimmelman suggests that obtaining “opt-in consent” from nearly 700,000 users would not have been “feasible.” It is important to recognize that obtaining meaningful informed consent online is not impossible, or even especially difficult, for those who are serious about it. Browse wrap and shrink wrap agreements are rarely meaningful, and click wrap agreements are not much better, since even though users are required to click to indicate their agreement to terms, they almost always do so without actually reading those terms. But if one actually *wants* people to

Facebook experiment could have been conducted with incompletely informed consent (rather than no informed consent at all) without significantly biasing the results is a closer question, scientifically speaking. Users might have been invited to participate in a study and told that the algorithm that curates their News Feed would be adjusted (in unspecified ways) for one week. Users who opt in might still be materially different from those who opt out, and those who agree to participate might behave differently in response to the intervention than they would have had they not been primed by the consent process to pay attention to the content of their News Feed. But the worries about biased results would be substantially muted compared to a version of the emotional contagion experiment where subjects were told exactly what researchers would do and why.

Yet, even if incompletely informed consent enabled the research to proceed without sacrificing scientific validity, and whatever its ethical merits in other cases, in this instance, it would not have cured the alleged ethical defect of depriving users of information they would have deemed material to making a considered judgment about whether or not to participate. The information that would need to be withheld is precisely the information that many members

---

read—and understand—what they are agreeing to, technology can actually work to one's advantage. Audio-visual modules can present material information in an accessible way and test individuals' understanding of that information before they are permitted to proceed. The Harvard Medical School-based Personal Genome Project (in which subjects agree to have their whole genome sequenced and published on the Internet for all to see, along with as much personal health and other phenotype information as they are willing to provide) uses such a process, Jeantine E. Lunshof et al., *From Genetic Privacy to Open Consent*, 9 NATURE REVIEWS GENETICS 406, 411 (2008), as do some medical research iPhone apps built on Apple's ResearchKit, such as the mPower app to study Parkinson's Disease. *Mobile Parkinson's Disease Study: How This Study Works*, MPOWER <http://parkinsonmpower.org/> (last visited May 5, 2015). I can think of no reason why a similar process in principle could not be scaled to 700,000 (or more) subjects. Facebook users would simply refresh their News Feed to find a pop-up window explaining the study and they would not be allowed to proceed to their Feed until they indicated either their dissent or their assent, perhaps after having successfully demonstrated that they read and understood the study disclosures. Indeed, such online consent modules make the consent process much more scalable than, say, having a Principal Investigator's research assistant serially "consent" each subject who enters the lab. If there is a practical problem with this approach, it is that many users, in their rush to access their News Feeds, are likely to refuse to participate out of hand rather than take the time to learn about the study, although Facebook could address this problem by requiring users to go through the consent modules before deciding whether or not to participate. This process also could not be used for each of Facebook's numerous experiments, or users would spend all their time going through consent modules and abandon the platform. But it is a plausible model on which Facebook could occasionally rely for important but complex and/or potentially risky or otherwise controversial studies, where selection bias is not a significant concern. For many of Facebook's other A/B studies, one-time blanket consent may suffice.

of the public found alarming (reasonably or not)<sup>106</sup> once they learned about the study—namely, that researchers had removed posts from News Feeds based on the emotional valence of the words they contain in order to study the effect on users' moods.<sup>107</sup> True, users would have known *that* they were being invited into a study—and they would have had the opportunity to decline that invitation. But without hopelessly biasing the results, those who agreed to participate could not have been told precisely the information that, it now appears, many would have deemed determinative to their participation decision.

### 3. Debriefing

The case for debriefing as somehow mitigating subjects' deprivation of autonomy is even weaker.<sup>108</sup> Like Professor Grimmelman, I, too, have suggested that the researchers should have debriefed subjects.<sup>109</sup> But it is a category mistake to view debriefing as even watered down informed consent or as primarily serving the same purposes that informed consent serves.

Obtaining consent is a means of respecting persons' autonomy by enabling them to make a considered judgment according to their own preferences and in light of all available material information

---

106. Researchers (like doctors) cannot possibly disclose *all* information pertaining to a proposed study (or treatment). Perhaps whether particular information ought to be disclosed to prospective subjects should depend on a "reasonable subject standard," according to which information must be disclosed if a reasonable subject would find it material to her decision whether to participate. Under that standard, one could argue in this case that a *reasonable* subject—one properly educated about what the study entailed and not led astray by alarmist commentary likening the experiment to a "pharmaceutical firm . . . randomly, secretly sneak[ing] an experimental drug . . . into the drinks of hundreds of thousands of people, just to see what happens" or offering unsupported speculation about the possibility that the experiment caused people to develop depression or heart failure or commit suicide, Lanier, *supra* note 69—would not have deemed the purpose of the study and the mechanism by which it was pursued to be material to their decision because the intervention fell within the normal range of their Facebook experience. *See infra* text accompanying notes 124–31.

107. *See* Kramer et al., *supra* note 12.

108. The Facebook-Cornell researchers did not debrief subjects. To this day, Facebook users do not know whether they were part of this particular study or not. They can, however, stop wondering whether they have ever been part of *some* Facebook experiment—or, indeed, whether they are participating in any experiments right now. *See supra* note 78.

109. *See* Michelle N. Meyer, *supra* note 65; Michelle N. Meyer, *Misjudgments Will Drive Social Trials Underground*, NATURE (July 16, 2014), <http://www.nature.com/news/misjudgments-will-drive-social-trials-underground-1.15553>.

about what will and will not happen to them. Debriefing—in which subjects are told after the fact either that they were, unbeknownst to them, subjects in a study or that certain aspects of the study were not disclosed or were not as they were led to believe—cannot serve these notice and assent purposes. Occasionally, IRBs require that subjects who are debriefed be given the opportunity to withdraw their data from the study. That gives subjects control over whether their data will be used going forward, but it cannot un-ring the bell of having had their data non-consensually collected (through observation or intervention) in the first place. In the Facebook case, the alleged risk would have materialized during the week-long intervention as a result of increased exposure to positive or negative words. Neither debriefing nor the opportunity to withdraw their data from the analysis of the effects of that intervention can erase any harm that occurred or restore subjects' opportunity to exercise their autonomy by agreeing or refusing to assume the risk of that harm.

This is not to say that debriefing is of no ethical moment, even in this case. Debriefing subjects can serve other facets of the principle of respect for persons *besides* respect for their autonomy. Debriefing demonstrates researchers' awareness that subjects are moral agents with their own preferences and projects that are deserving of others' respect and that the deceptive or non-consensual aspect of the study constituted a departure from that ideal.

Debriefing can also serve the principle of beneficence. Subjects who have been deceived or studied without their knowledge or consent sometimes experience a variety of negative emotions, such as anger, resentment, embarrassment, and self-doubt. These feelings may be exacerbated when subjects are not properly debriefed and instead learn about their participation by happenstance, such as through word of mouth (as when subjects are recruited from a common community, such as a college campus) or media accounts (as in this case). Word of mouth and media are often highly imperfect sources of information compared to researchers who conducted the study, and inaccurate information about the study and its implications can exacerbate subjects' distress.<sup>110</sup> Proper

---

110. For instance, subjects who take the IAT and are told that they strongly implicitly associate African-American children with negative words like "rotten" (an actual example from one IAT test) may, without proper debriefing, incorrectly believe that they are unique in harboring implicit associations and that these implicit associations elicited in the online lab have necessarily manifested themselves in the individual's real-world discriminatory actions towards African-Americans, women, or other groups. Similarly, subjects in cognitive illusion, memory distortion, conformity or obedience studies may feel "stupid" or gullible for having "fallen for" these deceptions if

debriefing serves the principle of beneficence by preventing these magnified harms.

In the Facebook case, for instance, media coverage of the study was often inaccurate—virtually always in the direction of painting a more alarming picture than what actually occurred.<sup>111</sup> As one psychologist who heads an anxiety disorders and social phobia lab at a major university has suggested, “it might be that the biggest risk was finding out the study happened and the amount of upset *that* caused.”<sup>112</sup> But this is less an argument against proper debriefing by researchers than an argument against sensationalistic debriefing by media. Had subjects been told calmly and accurately about what was done and why, and about how tiny the effect sizes were, reaction might have been quite different. Still, in some cases, debriefing subjects may do more harm to subjects than good, and this must be taken into account in determining whether or not debriefing is, as the Common Rule puts it, “appropriate.”<sup>113</sup>

#### 4. Minimal Risk Analysis

In the Facebook experiment, then, neither incompletely

---

researchers fail to explain how ubiquitous these behaviors are.

111. See, e.g., Caplan & Seife, *supra* note 80 (noting that the experiment “does not come close to meeting the conditions when research involves personal information,” thereby erroneously implying that researchers had access to individual users’ posts or other data); *id.* (speculating that “eliminating good news from a user’s newsfeed might send [a user] into a rage,” despite the fact that Facebook has always filtered out the majority of all News Feed-eligible posts for all users, with no evidence that doing so has ever sent anyone into a rage); Klitzman, *Why Facebook Should Follow Ethical Standards—Like Everybody Else*, *supra* note 80 (suggesting the Facebook experiment may have caused “[a]ltered mood [which] can in turn affect drug use, weight and appetite, school and work performance, and suicidal thoughts”); Lanier, *supra* note 69 (likening the Facebook experiment to a “pharmaceutical firm . . . randomly, secretly sneak[ing] an experimental drug . . . into the drinks of hundreds of thousands of people, just to see what happens”); Tweet of Lauren Weinstein, TWITTER, (June 28, 2014, 5:55 PM), <https://twitter.com/laurenweinstein/status/483051171255312384> (“I wonder if Facebook KILLED anyone with their emotion manipulation stunt. At their scale and with depressed people out there, it’s possible.”).

112. Thomas Rodebaugh & Michael Leary, *More Than Meets the IRB Podcast, Social Media Research Ethics: Considering the Facebook ‘Emotional Contagion’ Study*, WASH. UNIV. SCH. OF MED. IN ST. LOUIS (Sep. 2014), <http://digitalcommons.wustl.edu/hrpopods/8>, at 15:10–15:27.

113. 45 C.F.R. § 46.116(d)(4) (2009). See also OFF. FOR HUM. RES. PROT., IRB GUIDEBOOK, Ch. III, § B (1993), available at [http://www.hhs.gov/ohrp/archive/irb/irb\\_guidebook.htm](http://www.hhs.gov/ohrp/archive/irb/irb_guidebook.htm) [hereinafter IRB GUIDEBOOK] (“It is clear that debriefing is appropriate when it contributes to the subject’s welfare (i.e., when it corrects painful or stressful misperceptions, or when it reduces pain, stress, or anxiety concerning the subject’s performance). There is greater uncertainty over whether it is appropriate to debrief subjects when such a debriefing could itself produce pain, stress, or anxiety (i.e., IRBs must be concerned with cases where debriefing subjects might harm them but failure to debrief subjects would wrong them).”).

informed consent nor debriefing would have allowed us to escape the conclusion that subjects' autonomy was traded off against other values. If we wish to defend the position that the Facebook experiment and studies like it are ethically acceptable—with or without debriefing or incompletely informed consent—we will have to defend the lack of any meaningful consent. In this section, I bite that bullet.

As we have seen, under the Common Rule, the ethical acceptability of deviating from the standard of informed consent requires that any risks that subjects do not knowingly assume must be no more than “minimal.”<sup>114</sup> Risks must be judged based on the evidence available at the time—that is, they must be reasonably foreseeable.<sup>115</sup> And the relevant risks are the *incremental* risks of research—that is, those risks above and beyond what subjects would have been exposed to anyway by an ongoing practice onto which the research may piggyback or, as here, which the research may seek to investigate.<sup>116</sup> Nor are we to consider any long-term harms that may result from the way the research results are used.<sup>117</sup> Thus, the

---

114. See also IRB GUIDEBOOK, *supra* note 113, Ch. III, § B (“According to the regulations, research should not be permitted at all if the risk to subjects is more than minimal and the subjects are not being informed of things they would consider material to a decision to participate.”).

115. As part of its default rule requiring informed consent, the Common Rule directs IRBs to require the disclosure to subjects of “any reasonably foreseeable risks or discomforts.” 45 C.F.R. § 46.116(a)(2). Presumably, when determining whether research risks are more than minimal, the inquiry is similarly limited to consideration of those risks that are reasonably foreseeable at the time. BELMONT REPORT, *supra* note 83, Part C, § 2 (“It should . . . be determined whether an investigator’s estimates of the probability of harm or benefits are reasonable, as judged by known facts or other available studies.” (emphasis added)).

116. See 45 C.F.R. §46.111(a)(2) (“In evaluating risks and benefits, the IRB should consider only those risks and benefits that may result from the research (as distinguished from risks and benefits of therapies subjects would receive even if not participating in the research).”). See also IRB GUIDEBOOK, *supra* note 113, Ch. III, § A, (defining research risk as the “probability of harm or injury (physical, psychological, social, or economic) occurring as a result of participation in a research study” (emphasis added)); *id.* (“The IRB must: (1) identify the risks associated with the research, as distinguished from the risks of therapies the subjects would receive even if not participating in research . . .”). In other words, the study must be a but-for cause of research risks. So, for instance, in assessing the risks of a study that calls for patients already providing blood for diagnostic purposes to provide one additional vial for research purposes, the research risks would be limited to the risks, if any, of providing an extra vial of blood (say, a slightly increased risk of light-headedness), and not the risks that the patient-*qua*-patient has already assumed (such as infection at the needle site and the pain of the needle prick). In the case of Facebook, the company’s usual News Feed may impose various risks on users (such as reduced likelihood of seeing a post the user deems important, unhealthy social comparison from being exposed to friends’ “humblebrags”, or to toxic political debates), but in assessing the ethics of the study, we are interested only in any additional risks that the study alone imposed.

117. 45 C.F.R. § 46.111(a)(2) (“The IRB should not consider possible long-range

pertinent question is: Based on the evidence available at the time, what reasonably foreseeable risks, if any, did the Facebook study impose on subjects above and beyond the risks to which subjects were already exposed as regular Facebook users, and were those risks more than minimal?

By initiating or accepting “friend requests,” subjects consented to be exposed to whatever content those friends might produce, no matter how upsetting (short of something abusive that violates Facebook’s terms of service).<sup>118</sup> Researchers did not fabricate and plant especially emotional posts that subjects were not meant to see and might not otherwise have seen. Still, the experimental algorithm resulted in News Feeds that were more densely packed with positive or negative items than they otherwise would have been that week.<sup>119</sup>

---

effects of applying knowledge gained in the research (for example, the possible effects of the research on public policy) as among those research risks that fall within the purview of its responsibility.”). See also IRB GUIDEBOOK, *supra* note 113, at Ch. III, §A (“A . . . potential risk to subjects is the possible long-range effect of applying the knowledge gained through research. For example, information gained about associative memory may enable advertising companies to develop new techniques for encouraging arguably harmful consumer behaviors; associations between race or gender and intelligence may have profound effects on public policy. The regulations specifically provide, however, that IRBs should not consider such effects ‘as among those research risks that fall within the purview of its responsibility’” (citing §46.111(a)(2))). Unlike the rule that limits IRBs’ risk analysis to those risks caused by the research, this regulatory provision is likely better explained by the implausibility of an IRB accurately predicting long-term negative effects of research (and their lack of expertise in making what are effectively broad policy decisions) rather than a lack of ethical relevance, as long-term risks to subjects and third parties surely matter ethically.

118. A user can, of course, unfriend or unfollow friends whose posts produce unpleasant emotional reactions. But posts produced by those individuals are no longer eligible for inclusion in the user’s News Feed under either practice or experimental conditions. It is also true that users can abstain from Facebook on days that seem especially likely to produce upsetting content (say, during the 2014 Ferguson unrest). But as the examples from my own recent feed show, bad (and good) Facebook days are not always or even often predictable, and so users have little control over the emotional valence of their News Feed on any viewing.

119. Although the PNAS paper’s methods section is less than crystal clear, my understanding is that the experiments proceeded as follows. Recall that the average Facebook user is eligible to see approximately 1500 items in her News Feed in any viewing, and that Facebook’s practice algorithm prioritizes approximately 300 items for viewing. The researchers first applied this practice algorithm to subjects’ eligible posts, resulting in Feeds in which, on average across all four conditions, 22.4 percent of posts contained one or more negative words and 46.8 percent of posts contained one or more positive words, Kramer et al., *supra* note 12, at 8789. Note that these two categories likely overlap; that is, a post might contain both positive and negative words, and thus be coded as both positive and negative. This distribution of positive and negative posts describes the News Feed to which the average subject would have been exposed, had she not been enrolled in the study. When the researchers then applied the experimental algorithm, they removed between ten and ninety percent of positive, negative, or randomly selected posts. Subjects in the reduced positivity condition, for instance, would therefore necessarily have seen a News Feed that was more densely packed with



*If* exposure to positive and/or negative text is risky and *if* the size of that risk increases with the number of emotional words to which users are exposed, *then* the experiments involved some incremental risk.

Consider the first of these two conditions: Did these additional words pose a reasonably foreseeable risk to subjects? The relevant expert community—social scientists studying the effects of Facebook and other social media on users’ emotional experiences—was in equipoise over whether exposure to positive or negative words is psychologically harmful, with some studies suggesting the former and others the latter. Still other studies suggested (and continue to suggest) that neither positive nor negative words are harmful, *per se*; their riskiness depends on whether the user engages with the posts containing those words or simply passively receives them. Finally, the (largely unavoidable) methodological weaknesses of most of these studies—small sample sizes and observational methods that did not permit the drawing of causal inferences—raised doubts about the validity of *any* of these studies; their results may be little more than noise. Thus, the best available evidence at the time of the study suggested genuine uncertainty not only about whether the fewer-positive-words group or the fewer-negative-words group would be at increased risk, but whether *either* group would be.

As for the second condition, even assuming that exposure to positive or negative words is psychologically risky, it is unclear how—or even whether—that risk increases with an increase in the number of such words one sees.<sup>120</sup> Prior studies found only that longer or more frequent Facebook sessions correlated with negative affect; they did not investigate the relationship between negative affect and the number of negative or positive words to which users were exposed.

Under these circumstances, even if it is not quite right to say that users’ increased exposure to positive or negative words carried no incremental risk,<sup>121</sup> it is unfair to describe the study as one in

---

negative and neutral posts.

120. The psychological risk of exposure to positive or negative words may be positively correlated with the number of such words one is exposed to (either in absolute terms or as a relative percentage of the total text to which one is exposed). But that correlation may not be linear. For example, a sample increase in exposure may confer no additional risk, while a larger—say, three-fold—increase in exposure may correspond to a two, three, or even four-fold increase in risk.

121. What counts as a “reasonably foreseeable” research risk is a matter of considerable dispute currently. See OFF. FOR HUM. RES. PROT., DRAFT GUIDANCE ON DISCLOSING REASONABLY FORESEEABLE RISKS IN RES. EVALUATING STANDARDS OF CARE (Oct. 20, 2014), <http://www.hhs.gov/ohrp/newsroom/rfc/comstdofcare.html>.

which researchers “intentionally made people sad.” Even if these three researchers personally believed one hypothesis to be more likely than others (presumably, the emotional contagion hypothesis), and regardless of how intuitively obvious the study’s results may seem to some today,<sup>122</sup> researchers simply could not have known in advance that exposure to negative text causes (or even correlates with) negative affect in viewers.<sup>123</sup> Indeed, most studies of Facebook had predicted precisely the opposite result.

Even assuming that the study did impose additional, reasonably foreseeable risks on subjects in one or more research arms, it is unlikely that these risks were more than minimal. The Common Rule defines minimal risk to mean “that the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests.”<sup>124</sup> Some have said that tying the determination of minimal risk to the current level of riskiness in our daily lives amounts to an argument that, “Because everybody does X, X is ethical.”<sup>125</sup> But this provision of the Common Rule is better understood as a sensible refusal to engage in research exceptionalism by holding knowledge-producing activities to a higher standard than other activities that impose similar kinds and degrees of risk, simply because they are designed to contribute to generalizable knowledge.

Let us turn now to the incremental risks that the Facebook experiment imposed on users and how they compare to the risks of

---

122. See, e.g., David Gorski, *Did Facebook and PNAS Violate Human Research Protections in an Unethical Experiment?*, SCIENCE-BASED MEDICINE (June 30, 2014), <https://www.sciencebasedmedicine.org/did-facebook-and-pnas-violate-human-research-protections-in-an-unethical-experiment/> (characterizing the results as “[n]ot surprising[]”).

123. See Emily L. Evans & Alex John London, *Equipose and the Criteria for Reasonable Action*, 34 J. L. MED. & ETHICS 441, 444–45 (2006) (distinguishing “conflict” equipose, in which individual experts have opposing beliefs about the relative merits of two or more interventions, from “agnosticism” equipose, in which most experts are uncertain about the relative merits of the interventions, and arguing that both forms of equipose are appropriate preconditions for an ethical trial in which subjects are randomly assigned to one of these interventions).

124. 45 C.F.R. § 46.102(i). IRBs notoriously find this definition difficult to apply. See Meyer, *supra* note 79, at 261-63.

125. See, e.g., Richard Chirgwin, *Trick-Cyclists Defend Facebook Emoto-Furling Experiment*, REGISTER (July 2, 2014), [http://www.theregister.co.uk/2014/07/02/psych\\_researchers\\_link\\_arms\\_with\\_facebook/](http://www.theregister.co.uk/2014/07/02/psych_researchers_link_arms_with_facebook/); see also Zeynep Tufekci, Assistant Professor, iSchool and Department of Sociology, University of North Carolina–Chapel Hill, Privacy Conf.: Welcome & Panel One—A/B Testing and Manipulation Online: Should We Care? At Silicon Flatirons (Dec. 4, 2014), available at <https://www.youtube.com/watch?v=E55alZr716c>.

our daily lives. As I trust a cursory skim of the reader's own News Feed will confirm, posts frequently contain words of positive and negative emotional valence, some days more so than others. Although the experimental News Feeds were somewhat more concentrated with positivity or negativity than they would have been under the practice algorithm, these levels likely still fell well within the range of positivity and negativity that the practice algorithm produces in users' News Feeds over time.

For instance, as I write this paragraph, my own News Feed contains, among many other many other posts that have affected my emotions, the following: a post to an alumni group from a college classmate (unknown to me) expressing thanks for the support she received during her six-year-old nephew's cancer treatment and announcing that he had died (which caused an imagined scene of my own young son's death to form, unbidden); a friend posting poignant, bittersweet reminiscences about a longtime friend of hers (also unknown to me) who had just died of another form of cancer, leaving behind children and many friends; a friend of another friend who had been part of the spiritual conversion of Kelly Renee Gissendaner, scheduled to be executed several hours later for the contract killing of her husband, desperately pleading for people to sign a clemency petition to "spare the life of [her] friend"; collective mourning over Leonard Nimoy's death from lung disease; and a political argument on a friend's page on which several other friends commented (causing it to reappear at the top of my feed several times), which reached peak toxicity somewhere around the thirtieth comment.

Today happens to be a negative day on (my) Facebook, filled with news of lives that cannot be saved and disagreements that seemingly cannot be resolved. Other days, by contrast, have brought a seemingly endless stream of posts by friends gleefully announcing that they had been notified of manuscript acceptances or conference invitations (that I did not receive) or posting pictures of exotic vacations (that I am not taking); exposure to that sort of news may have its own negative effects. The Facebook experiment simply created conditions—a somewhat more positive or negative news week than these particular users would otherwise have experienced that week—that are almost certain to fall within the normal range of their—and our—Facebook experience.<sup>126</sup>

---

126. Some rare users may have judiciously limited their friends to those who post only emotionally neutral content. In those cases, however, the algorithm's de-prioritization of positive content would not have resulted in any increase in negative content.

By creating those conditions at the same time for a large number of subjects—a remarkable sample size for a social psychology study but, notably, only a tiny fraction of the 1.35 billion users, including the subjects, who are routinely exposed to these exact risks—researchers were able to control conditions and begin to draw causal inferences about the effects of News Feed on everyone.

Moreover, even the most dedicated Facebook users make time to engage in other activities, and many of these deliberately expose sometimes-unwitting individuals to similar emotional risks. Examples include reading the newspaper or watching the news, watching comedies or dramas, using Twitter or other social media,<sup>127</sup> being subject to television, radio, or print “fear appeals,”<sup>128</sup> reading blog comments,<sup>129</sup> and talking to their fellow human beings. Again,

---

127. The varying average happiness or sadness of Twitter on any given day can be learned (alas, only after the fact) by consulting The Hedonometer, a publicly available tool created by researchers who measure Twitter happiness from the “happiness score” given to the words used in a daily random sample of 50 million tweets. See UVM Computational Story Lab, The MITRE Team, *About*, HEDONOMETER, <http://hedonometer.org/about.html> (last visited Mar. 16, 2015). As of April 2013, the saddest day (on Twitter, at least) in the previous five years was the Friday Boston Marathon bombing. Stephanie Pappas, *The Saddest Day in 5 Years Is...*, LIVESCI. (Apr. 30, 2013), <http://www.livescience.com/29160-saddest-day-twitter-happiness.html>. In 2009, Facebook itself developed a Gross National Happiness index which uses the same software as the emotional contagion experiment to code anonymous status updates as containing negative or positive words and thereby tracks country-wide average happiness from day to day. See Matt Hicks, *How Happy Are We?* (March 24, 2010), <https://www.facebook.com/notes/facebook/how-happy-are-we/150162112130>.

128. Fear appeals are messages intended to shape the recipient’s behavior precisely by making her feel a negative emotion (usually fear, but also sadness or distress). Familiar examples include “scared straight” programs for youth warning of the dangers of alcohol, smoking, and drugs; appeals by international charities to donate money to victims of poverty, disease, or natural disasters, which leverage the human bias toward individual over statistical lives by showing images of attractive but forlorn individuals (often, children); and ASPCA donation appeals, including an animal cruelty appeal featuring singer-songwriter Sarah McLaughlin and graphic images of abused cats and dogs and an animal rescue appeal currently running featuring images of shivering dogs “who are clinging to life” as the Dickensian Christmas carol “In the Bleak Midwinter” plays in the background. For a collection of emotionally provocative “social issue” print advertisements, several of which would require a trigger warning on many of today’s college campuses (consider this yours), see *60 Powerful Social Issue Ads That’ll Make You Stop And Think*, DIGITAL SYNOPSIS, <http://digitalsynopsis.com/inspiration/60-public-service-announcements-social-issue-ads/> (last visited Mar. 17, 2015). Ironically, out of concern for subjects’ welfare, IRBs reportedly often make it impossible to study the effects of appeals that carry the same intensity of negative emotional stimulus as real-world appeals to which people are exposed routinely, and on a mass scale, with unknown consequences. See Michelle N. Meyer, *How an IRB Could Have Legitimately Approved the Facebook Experiment —and Why that May Be a Good Thing*, THE FACULTY LOUNGE (June 29, 2014), <http://www.thefacultylounge.org/2014/06/how-an-irb-could-have-legitimately-approved-the-facebook-experimentand-why-that-may-be-a-good-thing.html>.

129. The comments that appear beneath *60 Powerful*, *supra* note 128, are a good

the degree of positivity or negativity to which subjects were exposed in the Facebook experiment is almost certainly no more than minimally greater (if that) than what the average Facebook user routinely experiences outside of Facebook.

Even Professor Grimmelmann, the Facebook experiment's most vociferous and prolific critic,<sup>130</sup> in recently conceding that the experiment was "imminently eligible" for an alteration of informed consent under the Common Rule,<sup>131</sup> necessarily also conceded that the study posed no more than minimal risk to subjects. But readers who remain unconvinced are invited to imagine a version of the experiment they *would* deem to pose no more than minimal incremental risk—say, filtering out between one and nine percent of positive and negative posts, an intervention one-tenth the size of the actual intervention. For the point of this article is less to defend a specific study (a ship which, after all, has very much already sailed) than to urge that we view a certain category of field experiments from a different perspective. It is to that different perspective, of the role of experimentation in responsible innovation, that I now turn.

---

example both of the toxicity of many online comments sections and of the myriad negative emotions that emotionally provocative words and images can (perhaps) cause.

130. See James Grimmelmann, *Illegal, Immoral, and Mood-Altering: How Facebook and OkCupid Broke the Law When They Experimented on Users*, MEDIUM, Sept. 23, 2014, <https://medium.com/@JamesGrimmelmann/illegal-unethical-and-mood-altering-8b93af772688>; James Grimmelmann & Leslie Meltzer Henry, Letter to Proc. of the Nat'l Acad. of Sci. (July 17, 2014), *available at* <http://james.grimmelmann.net/files/legal/facebook/PNAS.pdf> (calling on PNAS to retract the Facebook-Cornell paper); James Grimmelmann & Leslie Meltzer Henry, Letter to the Off. for Hum. Res. Prot. (July 17, 2014), *available at* <http://james.grimmelmann.net/files/legal/facebook/OHRP.pdf> (requesting that OHRP investigate the Cornell University IRB for its role in the experiment); James Grimmelmann & Leslie Meltzer Henry, Letter to the Fed. Trade Comm'n (July 17, 2014), <http://james.grimmelmann.net/files/legal/facebook/FTC.pdf> (requesting that the Federal Trade Commission investigate "whether Facebook has engaged in unfair and deceptive trade practices by conducting unethical research on its users"); James Grimmelmann & Leslie Meltzer Henry, Letter to Facebook (July 24, 2014), *available at* <http://james.grimmelmann.net/files/legal/facebook/Facebook.pdf> (requesting, under Maryland's human subjects research law, that the company "make available for our inspection the final minutes of all meetings of Facebook's institutional review board"); James Grimmelmann & Leslie Meltzer Henry, Letter to OkCupid (July 30, 2014), *available at* <http://james.grimmelmann.net/files/legal/facebook/OkCupid.pdf> (same); James Grimmelmann & Leslie Meltzer Henry, Letter to Maryland Attorney General Douglas F. Gansler (Sept. 23, 2014), *available at* <http://james.grimmelmann.net/files/legal/facebook/MDAG.pdf> (requesting that the Attorney General of Maryland "seek an injunction requiring Facebook and OkCupid to refrain from human subject research on Maryland residents until they obtain full Common Rule informed consent from users and approval of each research protocol from a Common Rule-compliant IRB").

131. See Grimmelmann, *supra* note 103.

### III. FRAME TWO: RESPONSIBLE INNOVATION

“The world is just the A of the A/B test.”<sup>132</sup>

#### A. *The Ethical Relevance of the Distribution of Research Risks and Benefits*

IRBs are generally charged with ensuring “equitable” subject selection,<sup>133</sup> in part, in order to ensure that “[t]hose who accept the risks or burdens of being research subjects should be the ones who share in its benefits whenever possible.”<sup>134</sup> One feature of the Facebook experiment that has been almost completely ignored in the considerable debate about it is that the subjects who bore the risks of the study were randomly selected from among those most likely to benefit from its results.

This cannot be said of all corporate (or academic) experiments. For example, a different scenario would have been raised had Facebook data scientists teamed up with academics to study inattentive blindness in users by floating a woman in a gorilla suit across users’ Facebook page and observing status updates for signs of which users did and did not see it.<sup>135</sup> We all stand to benefit from better understanding of the nature and limits of our attention,<sup>136</sup> but Facebook users are no more likely than others to so benefit.

Field experiments designed to quantify the effects of an existing or proposed practice, by contrast, are not orthogonal to the practitioner-subject relationship. Facebook’s “emotional contagion” experiment was an attempt to determine which (if either) kind of posts regularly seen by 1.35 billion users—positive or negative—expose users to psychological risk. As the next section discusses, OkCupid’s similarly infamous experiment sought to ensure that what it touts as its competitive advantage—its matching algorithm—in fact does what it tells users it does: accurately predict compatibility. Neither experiment could have been conducted with fully informed consent without rendering the results all but meaningless. The

---

132. Duncan Watts, Fireside Panel: Experimentation and Ethical Practice, The Conference on Digital Experimentation at MIT (Oct. 10, 2014) (with Sinan Aral (moderator), Esther Dyson, Leslie Meltzer Henry, Michelle Meyer, and Jonathan Zittrain), available at <http://mitsloan.mit.edu/ide/code/>.

133. 45 C.F.R. § 46.111(a)(3) (2009).

134. IRB GUIDEBOOK, *supra* note 113, Ch. III, § B. See also BELMONT REPORT, *supra* note 83, Part B, § 3 (“justice demands . . . that [publicly funded] research should not unduly involve persons from groups unlikely to be among the beneficiaries of subsequent applications of the research”).

135. See Daniel J. Simons & Christopher F. Chabris, *Gorillas in Our Midst: Sustained Inattentive Blindness for Dynamic Events*, 28 PERCEPTION 1059 (1999).

136. See notes 98–100, *supra*.

results of these corporate experiments tell us something important about the respective safety and efficacy of these services—knowledge that can inform improvements in those services by the companies or (if the results are made public, and not hidden away for fear of the bad publicity they will invite) at least better-informed decisions by users themselves about whether and how to use those services. This essentially perfect fit between those who bear the risks of a study and those who are expected to benefit from it provides an important additional reason why a low-risk field experiment may be ethically justified despite the lack of consent.

Many have assumed, too quickly, that the only purpose of the Facebook experiment was to produce generalizable knowledge,<sup>137</sup> or to quiet criticism that exposure to friends' unrealistically happy lives saddens users.<sup>138</sup> Certainly, the data contributed to the broad scientific understanding of emotional contagion and social comparison, and on that score, subjects are no more likely to benefit than anyone else. But the data also contributed to a better understanding of how these phenomena do and do not apply to News Feed, in particular, and that is ethically relevant.<sup>139</sup>

---

137. See, e.g., Hill, *supra* note 68 (referring to users as “guinea pigs made to have a crappy day for science”); Kashmir Hill, *OkCupid Lied to Users About Their Compatibility as an Experiment*, FORBES (July 28, 2014), <http://www.forbes.com/sites/kashmirhill/2014/07/28/okcupid-experiment-compatibility-deception/> (“Facebook wanted users to have crappy days for science; OkCupid hoped they’d have crappy dates for science. What else are companies doing to us for the sake of experimentation?”); Cat Zakrzewski, *Why OKCupid’s Experiments Aren’t the Same as Facebook’s*, TECHCRUNCH (July 30, 2014), <http://techcrunch.com/2014/07/30/why-okcupids-experiments-arent-the-same-as-facebooks/> (“Unlike OKCupid, Facebook didn’t alter the user’s experience simply to improve the algorithm for a business purpose. In this study, the company essentially conducted a psychological experiment that many consider unethical.”); Penny, *supra* note 67 (“Facebook can manipulate the emotions of hundreds of thousands of people just to see what happens.”); Tim Carmody, *Why Don’t OKCupid’s Experiments Bother Us Like Facebook’s Did?*, KOTTKE (July 28, 2014), <http://kottke.org/14/07/why-dont-okcupids-experiments-bother-us> (“Facebook seemed to be testing user’s emotional expressions partly to solve a scholarly dispute and partly just to see if they could.”); Dylan Matthews, *Did OkCupid Send a Bunch of Incompatible People on Dates on Purpose?*, VOX (July 28, 2014), <http://www.vox.com/2014/7/28/5944865/okcupid-experiment-facebook-matches-ethics-moral-research> (“[T]he company had attempted to alter the emotional content of hundreds of thousands of people’s news feeds, just to see how they’d react.”).

138. See, e.g., James Grimmelmann, *Reboot*, LABORATORIUM (2D SER.) (Jan. 1, 2015), <http://2d.laboratorium.net/post/106852882870/reboot> (“Facebook’s . . . defense[] may have rung false with some observers [due to] a suspicion that the purported public benefit was really a smokescreen for corporate self-interest.”); Caplan & Seife, *supra* note 80 (“[C]ompanies are trying actively to manipulate you for their own interests. Even, apparently, if it harms you.”); Penny, *supra* note 67 (characterizing the Facebook experiment as “making tens of thousands of people sad for [the company’s] personal gain” and “to prove a point”).

139. One may object that there is no guarantee that Facebook, or any other

### B. *Mini Case Study in the A/B Illusion: OkCupid's Matching Algorithm Experiment*

In part, the instinct to assume that the emotional contagion experiment served either society at large or Facebook itself, but not users, may be due to a naïve assumption that companies like Facebook (and “practitioners” in many other fields, including medicine) somehow already know what does and does not work in their practice.

The following example is illustrative. Shortly after news of the Facebook experiment broke, the CEO of OkCupid, Christian Rudder, took to the company’s blog to declare that that it, too, experiments on its users.<sup>140</sup> OkCupid is an online dating platform that

---

corporate or other practitioner who conducts experiments will actually use the data to inform data-driven practice that benefits users. That’s true; there is no such guarantee. But, for what it is worth, IRBs generally do not incorporate into their ethical analysis of proposed research speculation about how the results will and will not be used. Indeed, the Common Rule specifically directs them not to do so. *See* note 117, *supra*. Moreover, even if a company fails to incorporate what it learns into its practice, publishing the results may empower users (with the help of media and academics) to make better decisions. In the case of Facebook, users will not be able to alter the algorithm any more than Facebook allows them to do, of course, but in light of the emotional contagion results, they may make different decisions about how, how often, or when they use Facebook, or they may choose to leave the platform entirely.

Ironically, it was the act of publishing the results in an academic journal (and characterizing them there in terms of broad theories of social psychology) that most clearly suggested that even if the experiment served internal quality assurance (QA)/quality improvement (QI) purposes, it also constituted “research,” as defined by the Common Rule. Because this study was not federally funded and because Cornell did not contract with OHRP to subject all university research to IRB review regardless of funding (so-called “checking the box” in its Federalwide Assurance, or FWA), the Common Rule would not have applied to this study in any event (except to whatever extent Cornell may, as a matter of private policy and employment contract, have adopted a policy requiring its affiliates to submit all research to IRB review). Moreover, as I have written elsewhere, *see* Meyer, *supra* note 65, my own interpretation of OHRP guidance as it applies to this case is that even if Cornell had “checked the box” on its OHRP FWA, the Cornell affiliates’ particular contributions to this study were such that Cornell was not “engaged in research” and hence the study did not require IRB review. But had Facebook not framed and published the results in a way that contributed to generalizable knowledge (but just filed the results in a desk drawer somewhere in Menlo Park), the experiment could have proceeded without IRB review without running afoul of the Common Rule even if the Cornell affiliates had participated in all aspects of the research. Thus, an otherwise identical activity may be subject to extensive *ex ante* regulation if it seems designed to contribute to generalizable knowledge (in which case it is “research”) but may escape that regulation altogether if steps are taken to avoid learning anything that might be useful to too many other people. As should be clear, the *ethical* relevance of this legally salient distinction is dubious. Indeed, if anything, ethics would have the distinction cut in the other direction, towards placing *less* regulatory burden on an activity with potential public benefit than on the same activity that cannot constitute a public good because it is proprietary.

140. Christian Rudder, *We Experiment On Human Beings!*, OKTRENDS (July 28th, 2014), <http://blog.okcupid.com/index.php/we-experiment-on-human-beings/>.



distinguishes itself from its competition on the basis of its matching algorithm:

We use math to get you dates. It's extremely accurate, as long as (a) you're honest, and (b) you know what you want . . . Most other matching sites are just glorified personals services. Their 'matching' systems are nonexistent or overly subjective.<sup>141</sup>

When a user views the profile of a prospective love interest, OkCupid tells him the probability that he and the potential amour will be a "match," displayed as a percentage from zero percent (horrible match) to 100 percent (perfect match). Users are encouraged to initiate contact with "good matches" through the platform's messaging system.

In his blog post, Rudder described one experiment in particular, "Experiment 3: The Power of Suggestion," that rankled critics.<sup>142</sup> In it, the company displayed to users different compatibility probabilities than its algorithm had computed.<sup>143</sup> Although some pairs of users computed to be a thirty percent match were told that they were a thirty percent match, others were told that they were a sixty or ninety percent match.<sup>144</sup> Although some pairs of users computed to be a sixty percent match were told that they were a sixty percent match, others were told that they were a thirty or ninety percent match.<sup>145</sup> And although some pairs of users computed to be a ninety percent match were told that they were a ninety percent match, others were told that they were a thirty or sixty percent match.<sup>146</sup>

"Not surprisingly," writes Rudder, "the users sent more first messages when we said they were compatible. After all, that's what the site teaches you to do."<sup>147</sup> What the company really wanted to

---

141. *About OkCupid*, OKCUPID, <http://www.okcupid.com/about> (last visited Feb. 28, 2014). OkCupid claims to have a patent pending on its matching algorithm, but they provide some explanation of what goes into the algorithm. *See Match Percentages*, OKCUPID, <http://www.okcupid.com/help/match-percentages> (last visited Mar. 17, 2015).

142. Rudder, *supra* note 140.

143. *Id.*

144. *Id.*

145. *Id.*

146. *Id.*

147. *Id.* The experiment "was 'short' and involved fewer than 1,000 users." *See Hill, OkCupid Lied to Users About Their Compatibility as an Experiment*, *supra* note 137. In an interesting twist on not-quite-debriefing, "a few days after the experiment was over," subjects received the following email: "Dear [nameA], Because of a diagnostic test, your match percentage with [nameB] was misstated as [%]. It is actually [%]. We wanted to let you know! Best, OkCupid." *Id.* Rudder explained: "Because 'experiment' has become such an emotionally loaded word, we used the more neutral phrase 'diagnostic test,' which we felt had the same meaning." *Id.*

know was, once one member of a pair reached out to the other, whether the algorithm accurately predicted whether they would *keep* talking—or whether the power of suggestion (simply being told by the algorithm that you *should* hit it off) was a better predictor. As Rudder put it, “maybe our matching algorithm was just garbage and it’s only the power of suggestion that brings people together.”<sup>148</sup>

Before turning to the results, consider how news of the experiment was received by academics, lawyers, and industry alike. Professor Grimmelmann characterizes it as follows:

OkCupid set up some of its users with deliberately bad matches, a move that seems to make a mockery of its claims to help users find love. In response, Christian Rudder argued that OkCupid’s mismatching experiment did indeed benefit users, but indirectly rather than directly, by validating the matching algorithm. Thus it helped users in general even if some particular users were mismatched . . . Rudder missed the point that the moral interests of individual users and the moral interests of users in general are not the same kinds of interests.<sup>149</sup>

One law partner and marketing and media law specialist opined that OkCupid’s experiment may have run afoul of the Federal Trade Commission (FTC) Act’s prohibition on unfair or deceptive acts or practices:<sup>150</sup> “When you’re matching people up with individuals who are not good matches, that would certainly be deceptive.”<sup>151</sup> An Executive Vice President of Operations at a mobile app development company was apoplectic: “OkCupid simply lied, falsifying their results and intentionally mismatching people. This manipulation invariably lead to countless terrible dates, wasted money, increased frustration and quite likely questions as to why these users could not find the love they were seeking.”<sup>152</sup> He called the experiment “staggeringly arrogant,” “an abuse of [the company’s] customer

---

148. Rudder, *supra* note 140. The experiment measured “bringing people together” only in the short-term sense of whether pairs who made initial contact ended up having “a real conversation,” which OkCupid deems to have happened after a pair exchanges four messages in its system. *Id.*

149. Grimmelmann, *supra* note 138.

150. Federal Trade Commission Act, 15 U.S.C. § 45(a)(1).

151. Pace Lattin, *You Won’t Believe What OKCupid Did* (July 31, 2014), <http://performinsider.com/2014/07/you-wont-believe-what-okcupid-did/> (quoting Jesse Brody, partner in the Advertising, Marketing & Media practice of Los Angeles law firm Manatt Phelps & Phillips).

152. Joseph Farrell, *Why OKCupid’s ‘Experiments’ Were Worse Than Facebook’s*, HUFFINGTON POST (Aug. 6, 2014), [http://www.huffingtonpost.com/joseph-farrell/why-okcupids-experiments-\\_b\\_5655217.html](http://www.huffingtonpost.com/joseph-farrell/why-okcupids-experiments-_b_5655217.html).

base,” and a “breach of the principles of corporate citizenship.”<sup>153</sup>

All of these comments claim that actual and dignitary harms flowed from telling pairs of users that their compatibility percentage was something other than the percentage the algorithm computed for them. But that conclusion only follows—and characterizing the experiment as displaying “deliberately bad” or “not good” matches only makes sense—if OkCupid already knew, prior to the experiment, that its computed compatibility probabilities were accurate. That premise is false: according to Rudder, OkCupid had no (non-correlational) evidence that whatever personality traits or other criteria comprise its algorithm cause compatibility, such that the algorithm accurately predicts compatibility.<sup>154</sup> (Rather than accusing OkCupid’s experiment of constituting an unfair or deceptive act, a stronger claim might be that OkCupid’s advertising of its *practice* violates the FTC Act’s ban on false, misleading, and unsubstantiated representations.<sup>155</sup>)

The same fallacy—what I will call the A/B illusion<sup>156</sup>—leads critics to describe the Facebook experiment as one in which the company “actively change[d] [its] customers’ moods to the negative.”<sup>157</sup> But when Facebook created News Feed six years earlier, it “manipulated” the information users posted by aggregating it and placing a select proportion of it front and center in every user’s Facebook home page. That manipulation was sure to change users’ moods *somehow*, just as newspaper editors can be certain that their decision in which stories to publish will change readers’ moods. The question the experiment sought to begin to answer was exactly how much, and in what direction, News Feed affects moods. It may be thought that newspapers that adhere to their own algorithm of “it bleeds, it leads” already know exactly what effect that policy will have on readers’ moods. That is in fact unclear—perhaps reading about others’ misfortune produces an uplifting feeling of *schadenfreude* in some readers. But whatever the case with newspapers, it is clear that the effects of News Feed on mood were

---

153. *Id.*

154. *See infra* text accompanying note 159.

155. FTC Act § 5(a)(1), 15 U.S.C. § 45(a)(1); FTC Act § 12(a), 15 U.S.C. § 52(a). *Cf.* POM Wonderful LLC v. F.T.C., No. 13-1060, \_\_F.3d\_\_ (D.C. Cir. Jan. 30, 2015) (upholding FTC order that company’s claims that a product helps treat or prevent a disease must be supported by a randomized, controlled trial and rejecting company’s argument that such RCTs were infeasible, in part, due to ethical concerns).

156. I use “illusion” metaphorically. But it may be that the tendency to entertain the A/B illusion is related to status quo bias or similar cognitive biases.

157. Chris Matyszczyk, *How Facebook Conducts Experiments on Your Emotions*, CNET (June 29, 2014), <http://www.cnet.com/news/how-facebook-conducts-experiments-on-your-emotions/>.

not well understood.

When a practice is implemented across the board, we tend to assume that it has value—that it “works”—even if it has never been compared to alternatives to see whether it works as well as those alternatives, or at all. Attempts to establish safety and efficacy through A/B or similar testing are then seen as depriving some people (those who receive B) of the standard practice. Those under the spell of the A/B illusion—as we all are at some time or another—view the salient moment of moral agency as the moment when an experiment to compare practices A and B was commenced, when it should more properly be recognized as the moment when practice A was unilaterally and uniformly implemented without evidence of its safety or effectiveness (i.e., without ever comparing it experimentally to B, or C, or anything else).

In a blog post presumably meant to debunk the A/B illusion by communicating how little practitioners typically know about the effects of their products and services, and the importance of experimentation to quality assurance and quality improvement, Rudder wound up confusing the issue himself by describing OkCupid’s algorithmically computed compatibility percentages as “actual” “good” and “bad” matches.<sup>158</sup> This characterization is highly misleading: it implies not only the true but uninteresting claim that two people were “actually” *deemed by the algorithm* to be compatible, but also the interesting but unsupported claim that they “actually” *are in reality* compatible. Determining whether the second claim is true was the reason for conducting the experiment in the first place:

The ultimate question at OkCupid is, does this thing even work? By all our internal measures, the “match percentage” we calculate for users is very good at predicting relationships. It correlates with message success, conversation length, whether people actually exchange contact information, and so on. But in the back of our minds, there’s always been the possibility: maybe it works just because we tell people it does.<sup>159</sup>

In other words, the correlation that OkCupid has observed between the criteria that comprise its proprietary algorithm and the rate of four-message “conversations” may be the result of reverse

---

158. He writes, for instance, “we told people who were actually good for each other, that they were bad, and watched what happened.” Rudder, *supra* note 140. Before testing the algorithm, he could not have known that people it deems “good” for each other in fact were.

159. *Id.*

causation. Normally, businesses are criticized for making unsupported claims when marketing their goods and services. And it is rare for businesses to show awareness of, let alone explicitly test for as sophisticated a possibility as reverse causation. Here, ironically, it was precisely OkCupid's attempt to determine whether its chief competitive marketing claim—its promise of being unique in “us[ing] math to get you dates”<sup>160</sup>—was empty or not that landed the company in trouble.

Some will respond that in these studies, as in many post-marketing (or so-called Phase IV) drug trials,<sup>161</sup> the results are more a means of advertising the product than a sincere attempt to scientifically determine its safety or efficacy.<sup>162</sup> Yet, as even a cursory review of the Facebook and OkCupid experiments reveals, neither one fully exonerated the safety or efficacy, respectively, of the company's product.<sup>163</sup> After the Facebook experiment, claims that positive posts harm users' mental health through social comparison were cast in doubt, and instead positive posts now seem more likely to confer mental health benefits through emotional contagion. But *negative* posts remain risky through the very same emotional contagion mechanism.

The OkCupid experiment similarly yielded both good and bad news for the company's existing practice. Averaged across all three levels of displayed compatibility (i.e., regardless of what they were told about their match), computed thirty percent matches were less likely to converse (14.3) than were computed ninety percent matches (17.6 percent).<sup>164</sup> That's evidence that OkCupid's matching algorithm “works”—that is, that it has an effect, that the algorithm itself (and not just the power of its suggestion) predicts some variation in the probability that subjects would converse. But the

---

160. About OkCupid, <https://www.okcupid.com/about>.

161. See Alex John London, Jonathan Kimmelman & Benjamin Carlisle, *Rethinking Research Ethics: The Case of Postmarketing Trials*, 336 *Sci.* 544 (2012).

162. See, e.g., Grimmelman, *supra* note 138 (For “some observers,” “Facebook's and OkCupid's defenses may have rung false . . . [due to] a suspicion that the purported public benefit was really a smokescreen for corporate self-interest.”).

163. Even where the results of corporate research appear self-serving, so long as those results are published, critics can point out methodological and other flaws and publication may at least put users on notice that there was a potential problem in need of addressing. Similarly, if the results suggest that a current practice is problematic but the company fails to take steps to alter it, publication at least provides users with the information they need to make better decisions, including pressuring the company for change or ceasing to use the company's product or service. Those actions, in turn, may drive competitors to offer safer or more effective alternatives.

164. Rudder, *supra* note 140. Rudder inexplicably makes none of this explicit, and these calculations are my own. The average of the first row in the final figure of Rudder's blog post is 14.3 percent. The average of the third row is 17.6 percent.

experiment showed that the power of suggestion also has an effect on whether people converse. Averaged across all three levels of computed compatibility (i.e., regardless of how compatible OkCupid thinks they are), users who were told that they were ninety percent compatible were more likely to converse (17.6 percent) than were those who were told that they were only thirty percent compatible (thirteen percent).<sup>165</sup> As Rudder concludes: “does the mere suggestion cause people to actually like each other? As far as we can measure, yes, it does.”<sup>166</sup> Indeed, this second, placebo-like effect is a bit larger (i.e., explains more of the variation in whether people conversed) than the effect of the algorithm (4.6 percent versus 3.3 percent). Rudder sums up the take-away message for users as follows: “OkCupid definitely works, but that’s not the whole story. And if you have to choose only one or the other, the mere myth of compatibility works just as well as the truth.”<sup>167</sup>

That Facebook’s experiment did not perfectly confirm the safety of its News Feed algorithm, and that OkCupid’s experiment did not perfectly confirm the efficacy of its matching algorithm, are not surprising results. Much practice—including, alarmingly enough, the practices of medicine, public health, and healthcare delivery—is driven more by habit, tradition, hunch, bias, and accidents of geography than by rigorous evidence. With relatively rare exceptions, such as marketing a novel drug, which legally requires experimental “A/B” testing in a small group of people before it may be marketed at scale, practitioners are free to implement whatever products, services, and policies they please, free from the burden of demonstrating that these are safe and effective.<sup>168</sup>

Some will respond that, aside from those same few exceptions, it is not necessary to obtain such evidence, because we already know from observing these practices currently in place—or from common sense—what their effects are.<sup>169</sup> There is some truth to this,<sup>170</sup> but

---

165. *Id.* The average of the third column in the final figure of Rudder’s blog post is 17.6 percent. The average of the first column is thirteen percent.

166. *Id.*

167. *Id.*

168. Of course, virtually all human activity is subject to some form of *ex post* regulation, if nothing else, through the possibility of common law tort and contract claims. And some innovations are subject to less burdensome *ex ante* regulations, such as pre-registration requirements.

169. *See, e.g.,* Penny, *supra* note 67 (“The findings of the study—that people are influenced by the emotions of others online as they are offline—surprised precisely nobody.”).

170. Observations do not allow us to infer causation, but they are far from meaningless evidence. This is especially so where we have observed the apparent effects of a practice on a large scale and for a long time.

not nearly so much as many believe. Often, we have very strong intuitions that an experiment is unnecessary because the outcome is inevitable: *Of course* we harbor no implicit biases towards disfavored groups, especially if we are members of that group ourselves.<sup>171</sup> The mere posting of a short checklist of already-standard procedures could *never* save thousands of lives and millions of dollars.<sup>172</sup> No one who carefully studies the face of her rapist could ever misidentify him later.<sup>173</sup> How could free legal aid by Ivy League law students *not* help indigent clients<sup>174</sup> and microloans *not* help the global poor?<sup>175</sup> *Of course* we would notice<sup>176</sup>—and take simple and safe steps to stop<sup>177</sup>—a crime happening nearby. And so on.

Such intuitions often turn out to be wrong, sometimes dangerously so. Practices that seem intuitively certain to be beneficial (or harmful) have been shown to be a mixed bag, at best.<sup>178</sup> Some simple potential practices that seem unlikely to make

---

171. See Darley Latané, *supra* note 95.

172. See *infra* text accompanying note 179.

173. See JENNIFER THOMPSON-CANNINO & RONALD COTTON WITH ERIN TORNEO, PICKING COTTON (2010) (memoir jointly written by a woman who carefully studied the face of her rapist during the attack and later was certain that she had correctly identified him and the man she incorrectly identified who served over a decade in prison before being exonerated by DNA evidence).

174. See *infra* note 178.

175. “Thirty years into the movement, it might seem strange that researchers are still asking whether microfinance reduces poverty. In fact, by the standards used to judge whether drugs are safe and effective in the bloodstreams of people, the safety and effectiveness of microfinance injected into the fabric of villages and barrios remains unproven.” David Roodman, *What Do We Really Know About Microfinance Impact?*, MICROFINANCE GATEWAY (Aug. 2009), <http://www.microfinancegateway.org/library/what-do-we-really-know-about-microfinance%20%92-impact>. After emerging as a major trend in the 1970s, the first of several RCTs to actually determine the effectiveness of microfinance was conducted in 2006, with mixed results, at best. See, e.g., Abhijit Bannerjee et al., *The Miracle of Microfinance? Evidence from a Randomized Evaluation*, 7 AM. ECON. J. 22 (2015), available at <https://www.aeaweb.org/articles.php?doi=10.1257/app.20130533>.

176. See Shaw and Porter, *supra* note 98.

177. See BELMONT REPORT, *supra* note 83.

178. For instance, job training programs have been found to reduce the earnings of participants; offering wage subsidies to employers to incentivize them to hire welfare recipients has been found instead to reduce their hiring rates; and offers of legal representation by Harvard Law School students have been found to produce no increased probability of prevailing but a delay in receiving benefits. D. James Greiner & Cassandra Wolos Pattanayak, *Randomized Evaluation in Legal Assistance: What Difference Does Representation (Offer and Actual Use) Make?*, 121 YALE L.J. 2118, 2424 (2012) (reporting results of an RCT finding that “an offer of [Harvard Legal Aid Bureau] representation had no statistically significant effect on the probability that a claimant would prevail, but that the offer did delay the adjudicatory process” and calling for further RCTs of legal services). See also Jeffrey Selbin et al., *Service Delivery, Resource Allocation, and Access to Justice: Greiner and Pattanayak and the Research Imperative*, 122 YALE L.J. ONLINE 45, 53-54 (2012) (noting that “[l]egal services programs and law school

much difference at all can turn out to save millions of dollars and thousands of lives.<sup>179</sup> And even when we feel confident (rightly or not) that an intervention is safe and effective, we often have no evidence of how it compares to alternatives, some of which may be equally beneficial but less expensive, or have fewer (or different) negative side effects, or are otherwise preferable to some or all users. Hence, rigorously studying the effects of an innovative practice on a small scale before implementing it more widely limits the extent of any negative impact of such practices.<sup>180</sup> It is highly unlikely that the benefits of a rule requiring premarket or pre-implementation testing of every new product, service, and policy would outweigh its costs. But retrospective testing of already-accepted practices may yield substantial welfare gains by revealing ineffective or inefficient practices. And field testing of existing practices, whether or not they were subject to premarket testing,<sup>181</sup>

---

clinics have served tens of millions of low-income clients since the 1960s, yet we lack basic information, let alone rigorous empirical data, about the impact of our work,” and that “few legal services programs or law school clinics conduct formal quality control or evaluate service outcomes, and fewer still have opened their practices to external scrutiny”).

179. For instance, every year, catheter-related bloodstream infections affect an estimated 80,000 patients in intensive care units (ICUs), costing 2.3 billion dollars and resulting in 28,000 deaths. A team of Johns Hopkins researchers, led by Peter Pronovost and funded predominantly by the Agency for Healthcare Research and Quality (AHRQ), posted in Michigan hospital ICUs a simple checklist reminder of five procedures previously shown by the CDC to have the greatest effect in reducing the rate of catheter infections: hand washing, using full-barrier infection precautions during catheter insertion, cleaning the patient’s skin with chlorhexidine, avoiding when possible the femoral site for line placement, and timely removing of unnecessary catheters. Posting this reminder about what clinicians should already have been doing, educating physicians about practices to control infection, and discussing this at daily rounds resulted in a sustained reduction in the rate of infections up to sixty-six percent. Peter Pronovost et al., *An Intervention to Decrease Catheter-Related Bloodstream Infections in the ICU*, 355 N. ENG. J. MED. No. 26 2725, 2726 (2006).

180. The sample size of the emotional contagion experiments was simultaneously massive and miniscule: nearly 700,000 subjects is huge for a social science experiment, but comprises a mere 0.04 percent of all Facebook users. This was one of the lessons of the early 1960s, when some 12,000 babies were born with severe deformities after their mothers took a drug, thalidomide, to control sleep and nausea during pregnancy. Canada and more than twenty countries in Europe and Africa had approved the drug for such use, but Frances Oldham Kelsey, who had recently joined the U.S. FDA as one of a handful of physicians reviewing drug approval applications, refused to approve it until additional clinical trials determined the drug’s effects. Although relatively few U.S. women were affected, the global tragedy led to U.S. Senate hearings and, in 1962, the Kefauver Amendments to the Food, Drug and Cosmetic Act, under which drug manufacturers were required to prove the effectiveness of their products by testing them in a few people before marketing them at scale.

181. Both the U.S. FDA and the European Medicines Agency sometimes require, as a condition of initial or continuing market approval, that drug and device makers commit to post-marketing clinical trials (sometimes called Phase IV trials) or post-marketing



can help close the effectiveness-efficacy gap.<sup>182</sup> As Rudder sums up these observations rather more pithily:

OkCupid doesn't really know what it's doing. Neither does any other website. It's not like people have been building these things for very long, or you can go look up a blueprint or something. Most ideas are bad. Even good ideas could be better. Experiments are how you sort all this out.<sup>183</sup>

*C. How the Research/Practice Distinction Fosters the A/B Illusion (And Overprotects Subjects and Underprotects Users)*

Unfortunately, the Common Rule fosters the A/B illusion by sharply distinguishing—without sound conceptual or normative reasons<sup>184</sup>—research and practice, and subjecting the former to

---

surveillance (involving data mining of electronic health records and similar activities) of these products. They do so both in order to close the efficacy-effectiveness gap and out of concern that some serious but rare side effects may not show up in small, brief premarketing clinical trials. The Food and Drug Administration Modernization Act of 1997 (FDAMA) amended the Food, Drug and Cosmetic Act by adding a new section to provide additional authority for these activities. Food and Drug Administration Modernization Act of 1997 § 130, 21 U.S.C. §356b (1997).

182. In controlling for potential confounds and biases, RCTs provide the most reliable method of deducing causation as opposed to mere correlation. But RCTs' high internal validity comes at the expense of limited external validity—the narrow range of real-world situations for which those claims are valid. The results of research conducted under highly controlled conditions using highly selected subjects may not extend to the real world of practice, where practitioners, patients/clients, and the environment rarely conform to RCT conditions. For instance, neither practitioners nor patients/clients are blind to the intervention being used. Patients/clients are not always compliant as are subjects in highly controlled lab settings. Nor are practitioners always as effective at implementing an intervention as are researchers adhering to a carefully detailed protocol, unsaddled with competing duties and distractions. Subjects are often “treatment naïve,” and have “pure” forms of the condition under investigation. But real patients/clients have almost always tried alternative interventions in the past and may even employ them concurrently with the new intervention, and they often have various co-morbidities. Because research subjects can never fully reflect the full range of genetic and other heterogeneity of patients, interventions often turn out, once introduced into practice, to have different effects on particular populations. In addition, latent harmful (or beneficial) effects of interventions for even populations reflected in the underlying study may only emerge once the intervention has been introduced into practice for some time. The emerging consensus in the health field with respect to this “effectiveness-efficacy gap” is not that evidence-based practice should be abandoned, but that research will need to be thoroughly and permanently integrated into the practice setting. *See generally* THE LEARNING HEALTHCARE SYSTEM: WORKSHOP SUMMARY (IOM ROUNDTABLE ON EVIDENCE-BASED MEDICINE) (LeighAnne Olsen et al eds., National Academies Press 2007), available at [http://www.nap.edu/download.php?record\\_id=11903](http://www.nap.edu/download.php?record_id=11903).

183. Rudder, *supra* note 140.

184. *See generally* Nancy E. Kass, et al., *The Research-Treatment Distinction: A Problematic Approach for Determining Which Activities Should Have Ethical Oversight*, HASTINGS CENTER REP. SPECIAL REP., 2013, S4-S15; T.L. Beauchamp, *Viewpoint: Why Our*

much stricter regulation than the latter. Indeed, defining research so that it could receive extra regulation was explicitly among the tasks assigned to the National Commission by Congress in the National Research Act.<sup>185</sup>

Although the Common Rule has come to apply to research of all kinds, it was prompted by scandals in biomedical research and developed by commissioners whose expertise was in medicine, bioscience, and biomedical ethics. Hence, the “practice” against which “research” was defined was *medical* practice. The working assumption of the Commission was that in medical practice, clinicians pursue the best interests of their patients; any risks or costs that patients bear are at least imposed in an attempt to benefit them. When an activity (“research”) is designed to contribute to generalizable knowledge, by contrast, subjects bear risks in an attempt to benefit future patients or society at large; any immediate benefit to subjects themselves is unlikely and, in any event, a happy accident. The *Belmont Report* distinguishes research and practice as follows:

[T]he term ‘practice’ refers to interventions that are designed solely to enhance the well-being of an individual patient or client and that have a reasonable expectation of success. The purpose of medical or behavioral practice is to provide diagnosis, preventive treatment or therapy to particular individuals. By contrast, the term ‘research’ designates an activity designed to test an hypothesis, permit conclusions to be drawn, and thereby to develop or contribute to generalizable knowledge (expressed, for example, in theories, principles, and statements of relationships).<sup>186</sup>

However, as some prominent bioethicists have recently noted—including, remarkably, the staff philosopher on the National Commission who was largely responsible for drafting the *Belmont Report*,<sup>187</sup> the research/practice distinction and the regulatory infrastructure that was erected on top of it increasingly constitute obstacles to learning healthcare systems and evidence-based

---

*conceptions of research and practice may not serve the best interest of patients and subjects*, 269 J. INTERNAL MED. 383 (2011).

185. See Tom L. Beauchamp & Yashar Saghai, *The Historical Foundations of the Research-Practice Distinction in Bioethics*, 33 THEORETICAL MED. & BIOETHICS 45, 46 (2012).

186. BELMONT REPORT, *supra* note 83, at Part A.

187. See Tom L. Beauchamp, *The Distinction Between Research and Practice*, YOUTUBE (May 8, 2011), <https://www.youtube.com/watch?v=qPQ2HE2CfCA> (video of talk given on May 3, 2011 upon receipt of The Hastings Center’s 2010 Henry Knowles Beecher Award).

medicine—and, less frequently noticed, to other forms of data-guided practice outside of medicine.<sup>188</sup>

Consider the following example. Every year, catheter-related bloodstream infections affect an estimated 80,000 patients in intensive care units (ICUs), costing 2.3 billion dollars and resulting in 28,000 deaths. In Michigan hospital ICUs, researchers installed a simple checklist of five evidence-based procedures deemed by the CDC to have the greatest effect in reducing the rate of catheter infections. Simply posting this reminder about what clinicians should already have been doing resulted in a sustained reduction in the rate of infections up to sixty-six percent, surprising even the researchers.<sup>189</sup>

Despite saving over 1,500 lives and nearly \$200 million in its first eighteen months,<sup>190</sup> the federal agency in charge of overseeing human subjects research—the Office for Human Research Protections (OHRP)—halted the study in its tracks, forbidding further data collection. The agency determined that even though the project was quality improvement (QI), it was also research and, as a result, required the written informed consent of every subject—including every patient and every *provider* (the risk to these subjects apparently being discovery of professional incompetence)—as well as IRB approval from each of the sixty-seven participating hospitals. This is a costly and lengthy process that would likely have significantly delayed the project and, thus, its life- and cost-saving results.<sup>191</sup>

In a remarkable response to the considerable protest that ensued, the then-Director of OHRP made clear that it was precisely the project's evidence-based nature that subjected it to these potentially crippling regulatory hurdles:

[T]he regulations do not apply when institutions are only

---

188. See Michelle N. Meyer, *From Evidence-Based Medicine to Evidence-Based Practice*, HASTINGS CENTER REP., March–April 2013, at 11. Such a healthcare system aims at continuous “learning” as the lessons from research and each care experience are systematically captured, assessed, and translated into reliable care. See generally, INST. OF MED., *BEST CARE AT LOWER COST: THE PATH TO CONTINUOUSLY LEARNING HEALTH CARE IN AMERICA* (Mark Smith et al. eds., 2013).

189. Peter Pronovost et al., *An Intervention to Decrease Catheter-Related Bloodstream Infections in the ICU*, 355 N. ENG. J. MED. No. 26 2725, 2731 (2006).

190. Atul Gawande, Op-Ed., *A Lifesaving Checklist*, N.Y. TIMES, Dec. 30, 2007, § 4, at 8.

191. *Statement Regarding The New York Times Op-Ed Entitled “A Lifesaving Checklist”*, U.S. DEP’T OF HEALTH & HUMAN SERVS., OFFICE FOR HUMAN RESEARCH PROT., (Jan. 15, 2008), <http://archive.hhs.gov/ohrp/news/recentnews.html> (emphasis added). Obtaining approval from even a single IRB can be time-consuming, but doing so from 67 separate IRBs, none of which communicate with one another and all of whom must agree on every aspect of the protocol (to preserve scientific validity) can preclude research.

*implementing* practices to improve the quality of care. At the same time, if institutions are planning research activities examining the *effectiveness* of interventions to improve the quality of care, then the regulatory protections are important to protect the rights and welfare of human research subjects . . .<sup>192</sup>

In other words, had these hospital administrations simply exercised their power by implementing their preferred practice (the checklist or something else) and then hoped for the best, that *practice implementation* would be subject to no prior review and require no one's informed consent. Placing a checklist in ICUs would no longer constitute an "experimental intervention," because no one would be bothering to try to determine the effects of this innovation. Implementing a checklist would simply be a policy choice, one of many affecting the welfare of others that those in power are privileged to make every day. Only because the actors instead laudably attempted to scientifically determine the *consequences for patients* of this practice did their actions invite aggressive regulatory scrutiny.<sup>193</sup> That is a remarkable, and regrettable, state of affairs.

#### D. *A Brief Note on the Problem of Unethical Underlying Practices*

Professor Grimmelmann recently explained his motivations for criticizing the Facebook experiment;<sup>194</sup> "It wasn't that I objected to being experimented on. Instead, knowing about Facebook's

---

192. Michael A. Carome, Office for Human Research Protections, *OHRP Statement Regarding The New York Times Op-Ed Entitled "A Lifesaving Checklist"*, U.S. DEPT. OF HEALTH AND HUM. SERVS., (Jan. 15, 2008), <http://archive.hhs.gov/ohrp/news/recentnews.html> (emphasis added).

193. OHRP eventually reversed course on the Pronovost study (importantly, without revising regulations or guidance), Ivor Pritchard, "OHRP Concludes Case Regarding Johns Hopkins University Research on Hospital Infections," *Off. for Hum. Research Protections News* (Feb. 15, 2008), <http://archive.hhs.gov/ohrp/news/recentnews.html>, but only after criticism reached an apex, including an unusual public shaming by Atul Gawande in the pages of the *New York Times*, Gawande, *supra* note 176, and a mocking editorial by the *Times* editors, *Pointy-Headed Regulation*, N.Y. TIMES (Jan. 27, 2008); *see also* Richard H. Savel, Evan B. Goldstein & Michael A. Gropper, *Critical Care Checklists, the Keystone Project, and the Office for Human Research Protections: A Case for Streamlining the Approval Process in Quality-Improvement Research*, 37 CRITICAL CARE MED. 725 (2009). The agency's new reasoning was remarkable: it concluded that, by that time, the research had so successfully proven the efficacy of the checklist that it amounted to standard-of-care rather than an experimental intervention, and so was no longer subject to the Common Rule. Of course, that reasoning would not have applied to earlier stages of the project, without which the effectiveness of the checklist could not have been demonstrated and incorporated into practice. OHRP did, however, note that even in its earlier, research stages, the project likely would have qualified for expedited IRB review and a waiver of informed consent.

194. *See* Grimmelmann, *supra* note 105.

manipulations in the name of science made me uncomfortably conscious of its other manipulations in the name of increasing my ‘engagement’ and selling me things.”<sup>195</sup>

Although I do not have space to defend this view here, I am skeptical that users given free access to a platform are entitled to be insulated from advertisements that enable that platform’s very existence.<sup>196</sup> But let us stipulate that legitimate concerns with the News Feed algorithm exist, as they surely do.<sup>197</sup> Even so, we generally ought not seek to change an underlying practice by attacking attempts to rigorously study its effects.

A similar strategy was suggested in a remarkable recent essay, *The CIA Didn’t Just Torture, It Experimented on Human Beings*, in which the author argues that those who wish to hold U.S. public officials accountable for torture should re-characterize those deeds as “human experimentation” in order to garner broader public condemnation:

As Americans from the Beltway to the heartland debate—again—the legality and efficacy of “enhanced interrogation,” we are reminded that “torture” has lost its stigma as morally reprehensible and criminal behavior. . . . Human experimentation, in contrast, has not been politically refashioned into a legitimate or justifiable enterprise. Therefore, it would behoove us to appreciate the fact that the architects and implementers of black-site torments were authorized . . . to experiment on human

---

195. Grimmelmann, *supra* note 138. Professor Grimmelmann, who recently announced that he was taking a “break” from Facebook, explained how Facebook “agitates” him, and his “relief at escaping Facebook,” as follows:

I like reading the news, don’t get me wrong, but Facebook is a singularly infuriating place to encounter it. Someone is always wrong on the Internet; and someone else is always complaining about it on Facebook. I’d log in to Facebook to unwind for a few minutes, and it would wind me up instead. Stepping away was like remembering to breathe—at least until I checked Twitter.

*Id.* There is some irony in this explanation. It is precisely these kinds of emotional effects of News Feed on users—being “wound up,” “infuriated,” and “agitated”—that Facebook’s experiment sought to quantify and, perhaps, redress.

196. Perhaps Facebook ought to be more *transparent* about how it prioritizes News Feed items, but that is a different claim.

197. For example, Facebook drives up to twenty percent of traffic to news sites, some thirty percent of adults in the United States get their news on Facebook, and major media outlets have seen their traffic nosedive following a significant adjustment to the News Feed algorithm. Ravi Somaiya, *How Facebook Is Changing the Way Its Users Consume Journalism*, N.Y. TIMES (Oct. 26, 2014), [http://www.nytimes.com/2014/10/27/business/media/how-facebook-is-changing-the-way-its-users-consume-journalism.html?\\_r=0](http://www.nytimes.com/2014/10/27/business/media/how-facebook-is-changing-the-way-its-users-consume-journalism.html?_r=0). When the News Feed algorithm affects which of their friends’ “news items” users see on Facebook, that, in turn, likely affects which *actual* news items they see elsewhere. That could be concerning.

beings.<sup>198</sup>

The author, an academic who should know better—and perhaps does—offers a glaringly incorrect “textbook definition of human experimentation” as being “subjected to psychological and physical torments, . . . the results . . . methodically documented and analyzed.”<sup>199</sup> But accuracy aside, what is wrong with equating human subjects research with torture, if it helps bring torturers to justice? What is wrong with criticizing Facebook for an experimental algorithm if it helps bring attention to problems in its practice algorithm?

These approaches are ethically suspect because human subjects research is critical to improving human welfare. Human subjects research even saves lives. Not everyone is convinced of the importance or even existence of News Feed’s psychological effects (although those who are not should not be especially exercised about this particular experiment, which involved no other risks), and I am certainly not suggesting that the Facebook experiment in particular saved lives. But once rung, the bells of “human experimentation” as inherently dangerous<sup>200</sup> and informed consent as an absolute requirement cannot easily be un-rung when public support of research one *does* find critical suddenly matters. An underlying practice worthy of criticism should by all means be criticized directly—but not indirectly, by stoking public

---

198. Lisa Hajjar, *The CIA Didn’t Just Torture, It Experimented on Human Beings*, NATION (Jan. 5, 2015), <http://www.thenation.com/article/193185/cia-didnt-just-torture-it-experimented-human-beings>.

199. *Id.* The Common Rule defines “research” as “a systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge.” 45 C.F.R. § 46.102(d) (2015). That’s it. Simply *talking to people* often meets this definition. An “experiment” usually refers to a subset of research involving intervention in a subject’s body (e.g., a blood draw) or environment (e.g., showing them one shade of blue on the Google homepage rather than another) rather than interaction (e.g., interviews or surveys) or observation (e.g., watching or recording private or public behavior). *Contra* Grimmelmann, *supra* note 105, distinctions among intervention, interaction, and observation are not “ethically salient” under the Common Rule. The Common Rule distinguishes these, but to make clear that all three are potential routes to regulation—not to subject some to regulation while dismissing others as intrinsically benign. Studies involving any of the three methods may be subject to full IRB review and fully informed consent and, conversely, studies involving any of the three methods may qualify for expedited IRB review or a waiver or alteration of informed consent, or be exempt from IRB review altogether.

200. Studies have shown that when an identical protocol is described as an “experiment,” it is judged to be more risky than when it is described as a “study.” Stephen John Cico, Eva Vogeley & William J. Doyle, *Informed Consent Language and Parents’ Willingness to Enroll Their Children in Research*, 33 IRB 6 (2011); Elisa J. Gordon, Amy Harris Yamokoski & Eric Kodish, *Children, Research, and Guinea Pigs: Reflections on a Metaphor*, 28 IRB 12 (2006).

misperceptions and “misfearing”<sup>201</sup> of human subjects research and thereby threatening the already fragile ecosystem of knowledge production.

#### CONCLUSION: RESPONSIBLE INNOVATION & A CULTURE OF CONTINUOUS TESTING

What I have called the A/B illusion involves the tendency to focus on the experiment in the foreground rather than the ongoing practice that exists in the background. As a final example of the A/B illusion, imagine that the head of a company is concerned that some of her employees are failing to save enough for retirement. She sends them mailings explaining the benefits and inviting them to automatically enroll in a 401(k) plan, of course, but some employees always fail to sign up. She decides that from now on, when she sends out 401(k) mailings, she will include a statement about how many co-workers within five years of the employee’s age have signed up for automatic enrollment. She hypothesizes that the minority of employees who haven’t enrolled may be influenced to do so by knowledge of the majority’s contrary behavior.

If more employees do indeed enroll following implementation of this new practice, our CEO won’t know for sure whether the uptake in retirement savings was caused by the innovative mailing or by some other factor that occurred during the same time period (perhaps *Nudge*<sup>202</sup> was released during that period and some of her employees read it, or perhaps a tax code change altered the incentives or provided more reminders).

Nor will she know whether the effect would have been even larger had she given employees information about, say, their peers within a ten-year age range. During the next enrollment cycle, she could implement that alternative policy, observe its effects, and attempt to compare them to the prior policies. But again, she will not be certain whether any differences were caused by the different policies or by other factors, observed or not, beyond her control.

Assuming that information about co-workers’ saving habits is sufficiently anonymized, few, if any, are likely to object that the CEO is unethically manipulating her employees’ behavior, abusing her power over them, foisting “experimental” interventions with unknown effects on them, or depriving them of important information about the effects on them of her innovation. Companies

---

201. See Cass R. Sunstein, *Misfearing: A Reply*, 119 HARV. L. REV. 1110 (2006).

202. RICHARD H. THALER & CASS R. SUNSTEIN, *NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS* (2008).

could more definitively determine the effects of these and other practices through randomized, controlled experiments, often in collaboration with academic researchers. And indeed, in the domain of retirement savings, such experiments are conducted.<sup>203</sup> But when they are, subject-employees are not told, even after the fact, often because companies fear (probably correctly) that employees will balk at being “experimented upon,”<sup>204</sup> even though the true experiment that treats people like guinea pigs is an innovation foisted on others whose effects are never investigated.

There is, of course, a large academic literature concerning the ethics of radical innovation in the biosciences (e.g., nanotechnology, fracking, human germline genetic modification), much of it an extended debate over the appropriateness of various versions of the precautionary principle. In the far more common (indeed, daily) case of incremental innovation, which is what Facebook, OkCupid, similar companies, and many other practitioners are usually best described as engaged in, there is an emerging literature on responsible innovation in finance.<sup>205</sup> That literature tends to emphasize the need to anticipate and avoid unintended consequences, to engage

---

203. *See, e.g.*, John Beshears, et al., *The Effect of Providing Peer Information on Retirement Savings Decisions* (Nat'l Bureau of Econ. Research, Working Paper No. 17345, 2011), available at <http://www.nber.org/papers/w17345>. In this experiment, academic researchers, working with a “large manufacturing firm and its retirement savings plan administrator,” used “social norms marketing” to investigate the effect of a peer information intervention on retirement savings choices. Subject-employees were randomized to one of three conditions: receiving information about the savings behavior of coworkers in their five-year age bracket (e.g., employees at the firm between the ages of twenty and twenty-four); receiving similar information about coworkers in their ten-year age bracket; and receiving a mailing that contained no peer information (the control group). As in the Facebook experiment, telling subjects in advance what they planned to do and why would have biased the results by causing different responses to the mailings, and had employees been permitted to opt out, the results would have been of limited generalizability due a selection effect.

204. *See id.* (noting that subject-employees never learned that they were part of an experiment). Companies frequently refuse to allow researchers to debrief employee-subjects, anticipating that employees will object to having been “experimented upon.” Personal communication with anonymous economist (March 3, 2015).

205. *See, e.g.*, Fabian Muniesa & Marc Lenglet, *Responsible Innovation in Finance: Directions and Implications*, in RESPONSIBLE INNOVATION 185, 185 (Richard Owen, John Bessant & Maggy Heintz eds., 2013) (finding it “striking . . . that, despite the spread of metaphors of ‘toxicity’ in accounts of financial products involved in the subprime crisis of the late 2000s (practitioners and commentators alike talk about ‘toxic assets’ or ‘toxic products’), principles of testing and vigilance, such as the ones put forward in the pharmaceutical industry, are still marginal in the financial sector”); Margaret Armstrong et al., *Towards a Practical Approach to Responsible Innovation in Finance: New Product Committees Revisited*, 20 J. FIN. REG. & COMPLIANCE 147 (2012) (analogizing innovation in financial products and activities to pharmacological innovation and proposing that existing bodies known as New Product Committees be adapted to take on the task of ensuring responsible innovation of financial practices from their inception).



stakeholders in deliberation, to be transparent, and so forth. Rarely is a culture of continuous testing for safety and efficacy mentioned in scholarly discussions of responsible incremental innovation, much less incorporated into corporate culture. It should be.

Ethical analysis of field experiments designed to determine the effects of an existing or proposed practice should include a consideration of the risks of that practice and of alternative methods of quantifying those risks. Facebook has been accused of abusing its power over users, of treating users as mere means to Facebook's (or its Cornell collaborators') ends, and of depriving users of information critical to making a considered judgment about participating in an activity that may set back their interests (by failing to inform them of the experiment). But Facebook's primary and perfectly financially tenable alternative to conducting an experiment—to simply change its business practice once and for all, without collecting any data to guide its decision, and dismissing others' probative (but not dispositive) evidence that News Feed was, one way or another, harming users—seems worse on all these scores.

Not only is field testing products, services, and practices usually a more ethical option than declining to do so, online companies are usually the best positioned to do so, and at the least cost. Online companies are often uniquely situated to rigorously evaluate the effects of their innovations through randomized, controlled studies (either independently or in collaboration with third-party researchers), because they alone have access to the algorithms and large numbers of users and data.<sup>206</sup> Moreover, A/B testing is usually quick and inexpensive—in many cases, an RCT literally can be conducted with the flip of a switch—especially if a company has an in-house data science team.

None of this means, of course, that every corporate experiment, even every experiment designed to investigate the safety or efficacy of a company's products, services, or practices, is ethically laudable or even permissible. Some corporate research is unethical, just as some academic research is unethical. There are legal reasons why the federal Common Rule does not generally reach private actors like

---

206. This is why it is disappointing that Facebook responded to public outcry over its "emotional contagion" experiment by suggesting that in the future it will retreat from experimental methods in favor of what are often second-best methods resorted to only when randomized, controlled studies are impossible. Mike Schroepfer, *Research at Facebook* (Oct. 2, 2014), <http://newsroom.fb.com/news/2014/10/research-at-facebook/>. Academics, including those Facebook's statement references in its announcement, often have to resort to non-experimental methods in studying social media because they lack access to corporate data and algorithms.

Facebook, OkCupid, and companies who experiment with different 401(k) enrollment processes. But as an *ethical* matter, whether research is permissible, forbidden, or obligatory cannot plausibly rest on the source of the actor's funding or her status as an affiliate of an educational institution versus some other (or no) institution.

Neither, however, can the answer to this ethical question plausibly depend on whether many people may learn from an activity (as compared to just a few) or whether the activity is systematic and data-driven (versus haphazard and intuition-driven). Laws, regulations, and policies often must draw lines that may be under or over-inclusive and hope thereby to achieve a measure of rough justice. But ethics has the luxury—and burden—of nuance and of attending to case specifics. The bare fact that an activity meets the federal definition of “research”—“a systematic investigation . . . designed to develop or contribute to generalizable knowledge”—tells us nothing about its ethical status. Any assumption that research is intrinsically risky or unethical is empirically unwarranted and dangerous in a world already too governed by myriad important practices that are based on little more than hunch, intuition, or the personal preferences or idiosyncrasies of those in power.

In this article, I have suggested that the Facebook and OkCupid experiments, and many like them, can be seen as a form of responsible innovation. But responsible innovation only begins—it does not end—with a culture of continual testing. The responsible innovator will respond appropriately to the results of A/B testing. This may not require, in every case, that a practice be changed, implemented, or abandoned accordingly. But if the practitioner opts not to change her practice, then sometimes—especially if A/B testing was conducted without subjects' consent, imposed any incremental risk, and yielded important results—ethics will require that she at least transparently communicate the results so that users or consumers can make a more informed decision about whether and how to avail themselves of that practice. Although Facebook ought to have debriefed subjects in the emotional contagion experiment and, ideally, communicated the results of the study directly to all Facebook users, the company did take the important but unusual step of publishing their results. It would be a great shame if the primary outcome of the furor over the emotional contagion experiment were that the potentially illuminating results of A/B testing at Facebook and elsewhere remain secret.<sup>207</sup>

---

207. In researching this article, I corresponded by email several times with Adam Kramer, the Facebook data scientist who is listed as the corresponding author of the

---

emotional contagion experiment paper published in PNAS. I had several questions about the study's methodology, and a few about how it related to Facebook's underlying News Feed practice. Kramer explained, not surprisingly, that his responses to me needed to be vetted by Facebook's legal and public relations departments. The result was that each response took weeks and, in the end, he told me nothing that I didn't already know from reading the PNAS paper. My questions pertained to my minimal risk analysis, and if I were giving him legal counsel (or, for that matter, public relations advice), I, too, probably would have told him to stick to the four corners of the published paper and provide no new, potentially damaging, information. But from any other perspective, this is a suboptimal state of affairs, to say the least.

