(Students – Print out article or put in Notability/Paperport etc…
1. read article **3. Underline writer's thesis**
2. reread article **4. Respond to the issues in this article in an essay – turn into Turnitin.com by**
(think about – does it bother you that your online activity is closely tracked? Why? Why not?
How might the information that is collected about you be abused?)

Digital shadow: How companies track you online

Who's following your every move on the web, asks Alexis Madrigal, and what do they want from you?
PUBLISHED APRIL 13, 2012, AT 11:02 AM

THIS MORNING, IF you opened your browser and went to *NYTimes.com*, an amazing thing happened in the milliseconds between your click and when the news about North Korea or James Murdoch appeared on your screen. Data from this single visit was sent to 10 different companies, including Microsoft and Google subsidiaries, a gaggle of traffic-logging sites, and other, smaller ad firms. Nearly instantaneously, these companies can log your visit, place ads tailored for your eyes specifically, and add to the ever-growing online file about you.

There's nothing necessarily sinister about this subterranean data exchange: This is, after all, the advertising ecosystem that supports free online content. All the data lets advertisers fine-tune their ads, and the rest of the information logging lets them measure how well things are actually working. And I do not mean to pick on *The New York Times*. While visiting *The Huffington Post* or *The Atlantic* or *Business Insider*, the same process happens to a greater or lesser degree. Every move you make on the Internet is worth some tiny amount to someone, and a panoply of companies want to make sure that no step along your Internet journey goes unmonetized.

Even if you're generally familiar with the idea of data collection for targeted advertising, the number and variety of these data collectors will probably astonish you. Allow me to introduce the list of companies that tracked my movements on the Internet in one recent 36-hour period of standard Web surfing: Acerno. Adara Media. Adblade. Adbrite. ADC Onion. Adchemy. ADiFY. AdMeld. Adtech. Aggregate Knowledge. AlmondNet. Aperture. AppNexus. Atlas. Audience Science.

And that's just the A's. My complete list includes 105 companies, and there are dozens more than that in existence. You, too, could compile your own list using Mozilla's tool Collusion, which records the companies that are capturing data about you or, more precisely, your digital self.

While the big names — Google, Microsoft, Facebook, Yahoo, etc. — show up in this catalog, the bulk of it is composed of smaller data and advertising businesses that form a shadow Web of companies that want to help show you advertising that you're more likely to click on and products that you're more likely to purchase.

To be clear, these companies gather data without attaching it to your name; they use that data to show you ads you're statistically more likely to click. That's the game, and there is substantial money in it.

AS USERS, WE move through our Internet experiences unaware of the churning subterranean machines powering our Web pages with their cookies and pixel trackers, their tracking code and databases. We shop for wedding caterers and suddenly see ring ads appear on random Web pages we're visiting. We sometimes think the ads following us around the Internet are "creepy." We sometimes feel watched. Does it matter? We don't really know what to think.

The issues the industry raises did not exist when Ronald Reagan was president and were only in nascent form when the Twin Towers fell. These are phenomena of our time, and while there are many antecedent

forms of advertising, never before in the history of human existence has so much data been gathered about so many people for the sole purpose of selling them ads.

"The best minds of my generation are thinking about how to make people click ads," my old friend and early Facebook employee Jeff Hammerbacher once said. "That sucks," he added. But increasingly I think these issues — how we move "freely" online or, more properly, how we pay one way or another — are actually the leading edge of a much bigger discussion about the relationship between our digital and physical selves. I don't mean theoretically or psychologically. I mean that the norms established to improve how often people click ads may end up determining *who you are* when viewed by a bank or a romantic partner or a retailer that sells shoes.

Already, the websites you visit reshape themselves before you like a carnivorous school of fish, and this is only the beginning. Right now, a huge chunk of what you've ever looked at on the Internet is sitting in databases all across the world. The line separating all that it might say about you, good or bad, is as thin as the letters of your name. If and when that wall breaks down, the numbers may overwhelm the name. The unconsciously created profile may mean more than the examined self I've sought to build.

Most privacy debates have been couched in the technical. We read about how Google bypassed Safari's privacy settings, whatever those were. Or we read the details about how Facebook tracks you with those friendly Like buttons. Behind the details, however, are a tangle of philosophical issues that are at the heart of the struggle between privacy advocates and online advertising companies: What is anonymity? What is identity? How similar are humans and machines? This essay is an attempt to think through those questions.

The bad news is that people haven't taken control of the data that's being collected and traded about them. The good news is that — in a quite literal sense — simply thinking differently about this advertising business can change the way that it works. After all, if you take these companies at their word, they exist to serve users as much as to serve their clients.

AT THE HEART of the problem is that we increasingly live two lives: a physical one, in which your name, social security number, passport number, and driver's license are your main identity markers, and one digital, in which you have dozens of identity markers, which are known to you and me as cookies. These markers allow data gatherers to keep tabs on you without your name. Those cookie numbers, which are known only to the entities that assigned them to you, are persistent markers of who you are, but they remain unattached to your physical identity through your name. There is a (thin) wall between the self that buys health insurance and the self that searches for health-related information online.

For real-time advertising bidding, in which audiences are being served ads that were purchased milliseconds *after* users arrive at a Web page, ad services "match cookies," so that both sides know who a user is. While that information may not be stored by both companies, i.e., it's not added to a user's persistent file, it means that the walls between online data selves are falling away quickly. Everyone can know who you are, even if they call you by a different number.

Further, many companies are just out there collecting data to sell to other companies. Anyone can combine multiple databases together into a fully fleshed-out digital portrait. As a *Wall Street Journal* investigation put it, data companies are "transforming the Internet into a place where people are becoming anonymous in name only."

If a company can follow your behavior in the digital environment — an environment that potentially includes your mobile phone and television set — its claim that you are "anonymous" is meaningless. That

is particularly true when firms intermittently add offline information such as shopping patterns and the value of your house to their online data and then simply strip the name and address to make it "anonymous." It matters little if your name is John Smith, Yesh Mispar, or 3211466. The persistence of information about you will lead firms to act based on what they know, share, and care about you, whether you know it is happening or not.

Militating against this collapse of privacy is a protection embedded in the very nature of the online advertising system. No person could ever actually look over the world's Web tracks. It would be too expensive, and even if you had all the human laborers in the world, they couldn't do the math fast enough to constantly recalculate Web surfers' value to advertisers. So, machines are the ones that do all of the work.

When new technologies come up against our expectations of privacy, I think it's helpful to make a real-world analogy. But we just do not have an adequate understanding of anonymity in a world where machines can parse all of our behavior without human oversight. Most obviously, with the machine, you have more privacy than if a person were watching your clickstreams, picking up collateral knowledge. A human could easily apply analytical reasoning skills to figure out who you were. And any human could use this data for unauthorized purposes. With our data-driven advertising world, we are relying on machines' current dumbness and inability to "know too much."

This is a double-edged sword. The current levels of machine intelligence insulate us from privacy catastrophe, so we let data be collected about us. But we know that this data is not going away, and yet machine intelligence is growing rapidly. The results of this process are ineluctable. Left to their own devices, ad-tracking firms will eventually be able to connect your various data selves. And then they will break down the name wall, if they are allowed to.

THE ADVERTISING LOBBY is explicitly opposed to setting browser defaults for higher levels of "Do Not Track" privacy. If it is successful, there will be nothing to protect the vast majority of Internet users from unwittingly giving away vast amounts of data about who they are.

On the other hand, these are the tools that allow websites to eke out a tiny bit more money than they otherwise would. I am all too aware of how difficult it is for media businesses to survive in this new environment. Sure, we could all throw up paywalls and try to make a lot more money from a lot fewer readers. But that would destroy what makes the Web the unique resource in human history that it is.

I wish there were more obvious villains in this story. The saving grace may end up being that as companies go to more obtrusive and higher production-value ads, targeting may become ineffective. Avi Goldfarb of the Rotman School of Management and Catherine Tucker of MIT's Sloan School found last year that the big, obtrusive ads that marketers love do not work better with targeting but *worse*. And lo and behold, the "failure appears to be related to privacy concerns: The negative effect of combining targeting with obtrusiveness is strongest for people who refuse to give their income and for categories where privacy matters most," they wrote in a 2011 *Marketing Science* journal paper.

Perhaps, in the end, there are natural limits to what data targeting can do for advertisers, and when we look back in 10 years at why data collection practices changed, it will not be because of regulation or self-regulation or a user uprising. No, it will be because the best ads could not be targeted. It will be because the whole idea did not work and the best minds of the next generation will turn their attention to something else.