Big Data: Goldmine or Minefield?

Joni Jackson
Chicago State University

Big data is the massive amounts of data produced by us and about us. When companies use big data to reach us, we may feel a strong emotional bond with the company, similar to the bonds of a close friendship. Or, we may be left feeling creepy. An exploratory analysis was conducted to examine people's relationships with and feelings toward companies and their use of big data. This paper presents the results of this exploratory analysis and provides a number of suggestions about how companies might create strong bonds with people, without making them feel creepy.

INTRODUCTION

There has been explosive growth in the amount of data and information available to us and about us; this vast amount of information is "big data," a term coined to describe the "proliferation of data and our ability to make productive use of it" (Akimoff, 2013; Desouza & Smith, 2014). When one thinks about the amount of data and information created and collected by us and about us, consider this: "A child born in 2012 will leave a data footprint detailed enough to assemble a day-by-day, minute-by-minute, account of his or her entire life, online and offline, from birth until death" [italics added] (Sullivan, 2012).

This detailed account of an individual's life helps companies not only to understand and meet people's needs but also to build stronger bonds with them, bonds that resemble an intimate friendship. Amazon.com, one of the world's largest online retailers and a company that has consistently achieved top online customer satisfaction scores (Rueter, 2014), illustrates the successful use of big data to build strong customer relationships. While Amazon's customer satisfaction ratings declined slightly at the beginning of 2015, it still delivered high levels of customer satisfaction. Amazon has done so through its pioneering use of data to create customized personalization (e.g., "Inspired by Your Shopping Trends," "Similar Others Have Purchased"). These personalized recommendations are very much like word-of-mouth from a friend, or what Jeff Bezos likened to a "'return to yesterday' when the owner of the corner store knew you and your likes and preferences" (Ferranti, 2000).

Knowing people's likes and preferences has enabled companies like Amazon to create bonds with people that are more like friendships than "superficial transactional relationships;" these friendships can lead to trust and loyalty (Childress, 2014). Amazon's success with customized personalization highlights the strength of this approach; their application and use of big data allows them to give people exactly what they want. This represents a goldmine for people as well as companies like Amazon.

Similarly, Target Corporation used big data to get to know its customers. In 2002 a team of statisticians at Target was asked, "If we wanted to figure out if a customer is pregnant, even if she didn't want us to know (italics added), can you do that?" Yes, they could, and they did (Duhigg, 2012). By combining shopper data with third party data and using predictive analytics, Target was able to create

detailed profiles of customers that enabled them to determine if a woman was pregnant and, if she was pregnant, the stage of her pregnancy. Their use of big data allowed them to tailor offerings to meet a woman's need for pregnancy-related products. Target captured the type of information that friends have about one another, information that might lead to feelings of friendship, trust and loyalty (Childress, 2014), similar to Amazon's customized personalization approach. This creates a goldmine for people.

Yet, Target's use of intimate information made people feel "creepy." One might liken their experience to one in which a complete stranger approaches and addresses you by a favorite nickname that only family and close friends use - far too familiar for someone with whom you do not share an intimate bond - and one that might leave you feeling creepy. Target's use of big data allowed them to achieve a level of intimacy that enabled them to hone in on people's immediate need (for pregnancy-related products). However, Target's intimacy was far too familiar for someone with whom people did not share a close bond. As a result, people were left feeling creepy. Thus while big data creates a potential goldmine of friendships and strong emotional bonds with companies, big data may also create a potential minefield of frenemies who look and act like friends but leave people feeling 'creepy' because they know these people are not their friends.

While companies collect data about people, people also share data about themselves. Along with the explosion in big data from company efforts, we also see an explosion in data that people willingly sharing about themselves (which contributes to big data). There is also a proliferation of tools that allow people to share intimate details about themselves with friends and strangers. Commenting on this trend, Carr (2015) likened people today to "mini-media companies" who listen to "the holy music of the self." Indeed, 2014 was declared "the year of the selfie" by Twitter ((Ng, 2014), a testament to our love affairs with "our self" and our proclivity to post minute-by-minute accounts of our lives and further evidence of our comfort with and desire to share every moment of our lives with others. This represents a goldmine for people as it allows them to create bonds with other people and companies by sharing.

Yet while we see an increase in people sharing data about themselves, often intimate and detailed, people still report concerns about their privacy, particularly about the amount and type of information that is being collected. The term "privacy paradox" was coined to reflect this disconnect, where people reported concerns about disclosure of personal information but continued to engage in behaviors that led to greater disclosure of personal information (Barnes, 2006). This sharing, despite concerns with privacy, may not reflect a lack of concern about privacy but may instead reflect the fact that people are resigned to the reality that the cost of access to online information and services is their data and information. This represents a potential minefield for people.

Further, while people are aware of and contribute to data collection and aggregation, it is not altogether clear the extent to which people are fully aware of the data collection efforts (Baylon, 2014). As the means of data collection and analysis become more sophisticated, this becomes an important question to explore. As companies become more sophisticated, the data collection techniques become far less transparent (to the extent that data collection can be considered transparent today). While many of the benefits of data collection are clear - more information, more sophisticated insights, more intimate understandings that foster long-term, mutually beneficial and "friendly" relationships - the rise of big data leads to challenges for people, challenges that may be magnified as people become less aware of the extent of the collection and use of their data. These challenges represent a potential minefield for people.

METHOD

A small exploratory study was conducted among a convenience sample of eight women and three men, ranging in age from twenty-one to fifty-six. Participants were asked about their understanding of big data and how they felt about companies using big data to get to know them. (Note: Respondents were not asked about specific companies; the companies mentioned in the results are those the respondents offered as examples.)

FINDINGS

The Concept of Big Data

Overall most respondents were familiar with the concept of "big data," either because of recent news stories or by inferring its meaning ("lots of data"). Most also indicated that they were aware that companies conducted some type of data monitoring. While some never thought about data "being used against them personally," others felt less comfortable with "marketers using all of these little tricks" to get to know them. Some indicated that the notion of big data sounded "ominous." One respondent indicated that big data made her feel like she was being "surveilled in the furtherance of consumerism."

How Big Data Is Collected

How companies gathered people's personal information emerged as an issue. Respondents initially indicated that they felt companies should ask permission to access and use their information rather than use "sneaky ways of gathering information." Others noted that companies do ask permission (terms and agreements), but that people do not read ("I think I read [the terms and agreements] once ... long ago"). For one respondent, there is little real choice because "you don't get to opt out of accepting terms and conditions if you want to use the product or service."

Several respondents indicated that people tend to "over-share" and thus have agreed to give companies access to their personal information. "People put a lot of information out there" about themselves, so "a lot is our fault." One respondent indicated that he would like to be able to control the message, but the "new hot thing is for everybody to put information out there.... to be so open to everyone is not good." All of which contributes to big data. There was a sense among some that "we have all signed our lives away," a sense that people are losing or have lost control over their own information.

Getting to Know You - Insights From Big Data

It seems to be important that companies make decisions about what types of relationships they want to have with their customers. One respondent described feeling "genial distrust and skepticism" when asked about companies' efforts to get to know them. Respondents want to feel like they have a choice, "like in a relationship - you want to be a part of what's going on." Another was quite adamant, stating "It's about agency over my own life." While another respondent stated that, "We live in a society where everybody lives out loud." On this basis, one respondent indicated that companies are not invading our privacy but that companies do need to understand "boundaries" because there are no clear rules anymore about invasion of space or privacy.

When asked how a company gets to know you if they don't already know you, respondents offered several examples. Amazon gets to know you by creating a shared space; it is more acceptable to get intimate details about people if you have a relationship in this shared space (this comment related to transparency in collection and use of data).

One respondent stated, "I use my Discover Card to get cash rewards at Amazon. That's a good business relationship and a good idea. Quid pro quo. This was an intelligent move. And I am fine with that. In terms of tracking my behavior -- what makes my hackles rise is the notion of any type of friendship. But I don't have a problem with someone taking data and coming to intelligent conclusions to make goods and services that match my needs. But do you have to track to get this kind of data? Why not just ask people?"

Companies and Brands as Friends

The idea that companies might aggregate information to become friends had positive as well as negative connotations. Regarding the customized personalization pioneered by Amazon, one respondent indicated that, "while I don't love that feature, it can be useful at times." Another was adamant - she did not want companies to use algorithms to peer into her life - "I know what books I want to read." Others neither liked nor disliked the idea of companies using data to get to know them; they said they would use what they found useful and ignore what was not want useful. While some respondents indicated that they

were not friends with brands or companies, nor did they have (or want to have) a friendship with brands or companies, others described their relationships with companies as a friendship ("I have been dating Costco for a long time").

What Friendship Means

When asked about what friendship meant to them, several common themes emerged. Friendships were characterized as long term, trusting, honest and non-judgmental. "You don't have to explain yourself." The basis of the friendship is a true understanding of an individual's values. Favorite brands or companies become favorites when they become identified with values important to the individual ("like patriotism because we do feel emotions;" "mom and apple pie;" "Skippy Peanut Butter, really American"). Friendships are also open and transparent. Did they want companies to be their friends?

Favorite Companies and Favorite Brands as Friends

Favorite brands meet people's needs "on multiple levels;" it is a "feel good experience" interacting with favorite brands or companies. But it was also clear that for others, the feeling was that companies are not persons and "we are not friends." Favorite brands are indeed like best friends for some however. Here's what some respondents said.

> "They know my name and even when I switch it up [change drink order], they know ... my family doesn't even know me like that."
> 1. "They care enough about me not to let me get this drink with extra whip cream."
> 2. "It's about me when I walk through the door, I am not one of those million faceless people, like family and good friends."
> 3. They "learn my habits ... make me feel like they are paying attention ... that I am important enough that they take the time to get to know me ...I am wanted and acknowledged."

Clearly, while people may harbor some concerns about companies' collection and use of data, many acknowledge that it helps companies get to know them. Can companies build relationships with people using big data? Of course, however, companies also must address the challenges in the capture of big data. Clearly the benefit of big data is that it leads to insights that allow companies to get to know their customers. Yet, the challenge is doing so in a way that minimizes people's sense of loss of control.

The Collection and Use of People's Personal Data

Respondents report some level of discomfort with the collection and use of their personal data, which may stem from a sense of a loss of control.

> Companies do what they do "because they can."
> I have "a visceral reaction to being manipulated."
> I would like to see companies use "creative transparency," to make their actions more clear and to "make it easier for people to opt in or opt out."

Van Otterlo (2014) posed a provocative question in his recent article, Automated Experimentation in Walden 3.0, when he asked, "what happens to us, as a society, if ... informed citizens are predictively biased by technology capable of profiling and experimentation? Technology can be used to shape decisions and behavior by manipulation, not force." One of the participants expressed a similar concern when she stated, "They are shaping our preferences and we think it is all natural, but it is not."

DISCUSSION

Big data is big business. There is value in the snippets of data that we share about ourselves as well as the data that is unearthed about us. Using big data, companies can know anything about any of us, with or without our knowledge. While companies clearly benefit from the use of our data, so do we. Targeted,

personalized offerings result from the insights companies glean from all of this data and provide us with a better overall experience. We have access to massive amounts of information, free of charge (Cumbley & Church, 2013; Esposti, 2014). Intimate relationships stemming from "big data" foster closer relationships with companies that may create long-term, brand loyal relationships with their consumers.

But there is a flip side to the coin. While there are clear benefits, there are also costs. One cost is our loss of privacy, or as one respondent noted, loss of "agency over our own lives." This may be one of the many reasons why Target customers felt "creepy" after receiving targeted mailings. Respondents with whom I spoke also pointed out that people are responsible for some of the information that is "out there." So why do people feel so uncomfortable? Are people more comfortable if it is a favorite brand? It seems that they may be. Even those respondents who were adamant about not sharing details of their lives with companies were more relaxed in their body language when they spoke of their favorite brands (they laughed and smiled more). What is the difference?

Among respondents in this exploratory study, trust and honesty emerged as key themes in defining "friendships." Control, while potentially not a factor in friendship, is very likely a moderating factor when people are faced with "marketers using all those little tricks." If big data undermines people's ability to control their data (Newman, 2014), does it decrease their trust in companies and brands? Do people trust brands that are more transparent in their data collection efforts, and does trust follow from greater feelings of control?

Several respondents made reference to an idea that one called a "shared space" as a way for companies to establish closer relationships with their customers. One described "shared space" as sitting next to the same stranger in the same coffee shop at the same time every day. You have never been introduced, so you do not know one another, but you do "know" one another (because you know you share this same space and thus similar habits). You probably share more in common with this stranger than with another stranger who does not frequent this coffee shop regularly. Thus, if he were on the phone and "the stranger in the coffee shop" commented on something overheard in his private phone conversation, he would be less uncomfortable than if a "stranger on the bus" commented on something overheard in his private phone conversation. He likened this to companies who share space with their consumers versus those companies who do not.

So do our favorite companies or brands ("friends") raise fewer privacy concerns, despite the fact that they collect "big data," because they share space with us and are thus less likely to raise "our hackles"? This raises the question, what actions can companies take to be more transparent, friendly and trustworthy - to create shared space? As the field of big data changes, new developments unfold rapidly. Do these new approaches rely on more transparent means of data collection, means that might allay people's concerns about privacy and led to greater friendships and more trusting relationships? Some of the new developments in data collection and surveillance seem to suggest less transparent means of data collection. Let's examine some of these more recent developments.

Recent Developments in Data Collection
Facial Recognition
Facial recognition technology is employed by a number of companies; this software lets companies collect or infer people's age, gender and ethnicity. The Luce X2 Touch TV is a new vending machine that not only greets people by name, but also helps people make better snack choices because it only allows people to buy snacks from a pre-approved list. The machine can also access information about people's age and medical history (Chumley, 2014). Almax SpA also uses facial recognition software in their EyeSee Mannequins that allows them to track people's movements and to identify demographics (e.g., race, gender, age). Future plans include audio recording devices that will capture snippets of people's conversations (Esposti, 2014; Roberts, 2012).

New Patents/Patents Pending
In 2014, Apple submitted a patent application for technology that would allow them to infer people's moods by comparing people's baseline mood data to current mood data" (Greenzeiger, M.F., Phulari, R.,

& Sanghavi, M., 2014). Verizon recently applied for a patent for a DVR set-top box that will monitor people (and pet) behaviors in the home - "eating, exercising, laughing, reading, sleeping, talking, singing, humming, cleaning, playing a musical instrument, ... or engaging in any other physical activity ..." (Esposti, 2014). Amazon received a patent for "anticipatory shipping;" they can ship products that you have not (yet) ordered, but probably would based on your previous orders (Bensinger, 2014). Disney filed a patent to track guests at their theme parks by monitoring their shoes. According to Disney, this would allow them to create a customized experience for guests; "... a Mickey Mouse park employee could call a child by name ...by using the child's shoes." Moreover, Disney indicated that shoe-tracking would be "less-invasive than other options (e.g., Disney's MagicBands)" (Sabri, 2016).

Cognitive Computing

Cognitive computing is a fast-growing area that involves smart algorithms and self-learning systems. Google Assistant, Microsoft's Cortana, Amazon's Alexa, Facebook's M, and Apple's Siri are examples of smart personal assistants that rely on cognitive computing. These apps can predict people's behavior by learning their behavior from their "interests, schedule, family, friends, work, troubling cat-video obsession and everything you've ever bought or wanted to buy" (Beer, 2012; Nield, 2014); and, not only do these algorithms learn people's behavior, they seek to be proactive by listening to people's interactions and telling people what they might be interested in (Kendrick, 2016; Shahani, 2015).

Environmental Data Monitors

The "Array of Things" project was recently launched in Chicago; this project creates one of the first "permanent data collection infrastructures" in the United States. Data trackers monitor environmental conditions (e.g., temperature, humidity, sound, etc.). The trackers can also monitor people and their movements - although this is not part of the project's current plans (Heinzmann, 2014; Moser, 2014). In Toronto, there is a similar data tracker, Turnstyle, that monitors smartphone signals in downtown Toronto. Turnstyle gathers data and information, then sells it to businesses who use this information to create offers for their customers. Most smartphone users are not aware of this data collection" (Crawford, 2014).

CONCLUSION

These developments in data collection are more sophisticated and thus may allow the capture of data that is not subject to socially desirable posturing, which may increase the accuracy of data. But these approaches are becoming less transparent (i.e., people may be more unlikely to be aware of the data collection efforts). Thus, what are the implications, if any, for the relationships between people and the companies who seek to friend them? Does transparency matter in an environment where people understand that some choices are limited (i.e., the cost of access to many services is provision of one's data). Or do companies who successfully build long-term friendships with people enhance these friendships by being more transparent in their collection and use of data? Are people less likely to feel controlled, manipulated or exploited if they are aware of the data collection? Finally, would this explain the difference between people feeling creepy (i.e., Target's case) versus people feeling like we have "harken[ed] back to yesterday" (i.e., Amazon). Companies may benefit by making their data collection activities more explicit or transparent, and perhaps more importantly, understandable. Although people often explicitly (and sometimes unwittingly) agree to share their data, the bonds of friendship may be strengthened when people understand and agree to the use of their data.

REFERENCES

Akimoff, T. (2103). The rising tide of data journalism. WBEZ Radio. Retrieved from http://www.wbez.org/rising-tide-data-journalism-104947

Barnes, S. B. (2006). A privacy paradox: Social networking in the United States. First Monday, 11 (9), n. pag. doi: http://dx.doi.org/10.5210/fm.v11i9.1394

Baylon, C. (2014). Who owns the Internet? Big brother is cashing in on you. World Today, 70 (2), 18-20.

Beer, J. (2012). Google is getting creepy. Canadian Business, 85 (14), 28-31.

Bensinger, G. (2014). Amazon wants to ship your package before you buy it. The Wall Street Journal. Retrieved from http://on.wsj.com/1hyXtoI

Carr, D. (2015). Selfies on a stick, and the social-content challenge for the media. The New York Times. Retrieved from http://www.nytimes.com/2015/01/05/business/media/selfies-on-a-stick-and-the-social-content-challenge-for-the-media.html

Childress, S. (2014). The future of digital marketing is you. Frontline. Retrieved from http://www.pbs.org/wgbh/frontline/article/the-future-of-digital-marketing-is-you/

Chumley, C.K. (2014). Unveiled: Facial recognition vending machines to help snackers quit junk-food frenzy. The Washington Times. Retrieved from http://www.washingtontimes.com/news/2014/dec/9/facial-recognition-technology-put-on-vending-machi/

Crawford, K. (2014). When big data marketing becomes stalking. Scientific American, 310 (4), 14-14. Retrieved from http://www.scientificamerican.com/article/when-big-data-marketing-becomes-stalking1/

Cumbley, R. & Church, P. (2013). Is "big data" creepy? Computer Law & Security Review: The International Journal Of Technology Law And Practice, 29, 601-609. doi:10.1016/j.clsr.2013.07.007

Desouza, K.C. & Smith, K.L. (2014). Big data for social innovation. Stanford Social Innovation Review, 12 (3), 38-43. Retrieved from http://www.ssireview.org/articles/entry/big_data_for_social_innovation

Duhigg, C. (2012). How companies learn your secrets. The New York Times. Retrieved from http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all&_r=0

Esposti, S.D. (2014). When big data meets dataveillance: The hidden side of analytics. Surveillance & Society 12 (2), 209-225. Retrieved from http://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/analytics/analytic

Ferranti, M. (2000). Bezos: Net personalization is key. Computerworld. Retrieved from http://www.computerworld.com/article/2596081/e-commerce/bezos--net-personalization-is-key.html

Greenzeiger, M.F., Phulari, R. & Sanghavi, M.K. (2014). Inferring user mood based on user and group characteristic data. US Patent & Trademark Office. Retrieved from http://patft.uspto.gov/netacgi/nph-Parser?Sect2=PTO1&Sect2=HITOFF&p=1&u=/netahtml/PTO/search-bool.html&r=1&f=G&l=50&d=PALL&RefSrch=yes&Query=PN/8965828

Heinzmann, D. (2104). New sensors will scoop up "big data" on Chicago. Chicago Tribune. Retrieved from http://articles.chicagotribune.com/2014-06-20/news/ct-big-data-chicago-20140621_1_cell-phone-data-big-data-sensors

Kendrick, J. (2016). How to use Amazon Echo and why you should get one. ZDNet. Retrieved from http://www.zdnet.com/article/how-to-use-the-amazon-echo-and-why-you-should-get-one/

Moser, W. (2014). What's up with the data sensors scattered around Hyde Park? Chicago Magazine. Retrieved from http://www.chicagomag.com/Chicago-Magazine/November-2014/Whats-Up-With-the-Data-Sensors-Scattered-Around-Hyde-Park/

Newman, N. (2104). How big data enables economic harm to consumers, especially to low-income and other vulnerable sectors of the population. Journal of Internet Law, 18 (6), 11-23.

Ng, N. (2014). Twitter declares 2014 year of the selfie. CNN. Retrieved from (http://www.cnn.com/2014/12/12/tech/twitter-selfie-trend/

Nield, D. (2014). Personal digital assistants, rise up! Fortune. Retrieved from http://fortune.com/2014/05/06/personal-digital-assistants-rise-up/

Roberts, A. (2012). Bionic mannequins spy on shoppers to boost luxury sales. Bloomberg News. Retrieved from http://www.bloomberg.com/news/articles/2012-11-19/bionic-mannequins-spy-on-shoppers-to-boost-luxury-sales

Sabri, R. (2016). Why the Magic Kingdom wants to track your feet. Christian Science Monitor. n. pag. Retrieved from http://www.csmonitor.com/Technology/2016/0802/Why-the-Magic-Kingdom-wants-to-track-your-feet

Rueter, T. (2014). Amazon's prices lead to more customer dissatisfaction. Internet Retailer. Retrieved from https://www.internetretailer.com/2014/12/31/amazons-prices-lead-more-customer-dissatisfaction

Shahani, A. (2015). How Personal Should A Personal Assistant Get? Google And Apple Disagree. NPR AllTech Considered. Tech, Culture and Connection. Retrieved from http://www.npr.org/sections/alltechconsidered/2015/07/03/419540367/how-personal-should-a-personal-assistant-get-google-and-apple-disagree

Sullivan, M. (2012). Data Snatchers! The booming market for your online identity. PC World. Retrieved from http://www.pcworld.com/article/258034/data_snatchers_the_booming_market_for_your_online_identity.html

van Otterlo, M. (2014). Automated experimentation in Walden 3.0: The next step in profiling, predicting, control and surveillance. Surveillance & Society, 12 (2), 255-272. Retrieved from http://ojs.library.queensu.ca/index.php/surveillance-and-society/article/view/walden3