Using Big Data as a window into consumers' psychology

Sandra C Matz[1] and Oded Netzer[2]

The rise of 'Big Data' had a big impact on marketing research and practice. In this article, we first highlight sources of useful consumer information that are now available at large scale and very little or no cost. We subsequently discuss how this information – with the help of new analytical techniques – can be translated into valuable insights on consumers' psychological states and traits that can, in turn, be used to inform marketing strategy. Finally, we discuss opportunities and challenges related to the use of Big Data as a window into consumers' psychology, and provide recommendations for how to implement related technologies in a way that benefits both businesses and consumers.

Addresses
[1] University of Cambridge, Department of Psychology, Downing Site, Cambridge CB2 3EB, United Kingdom
[2] Columbia Business School, Columbia University, New York, United States

Corresponding author: Matz, Sandra C (sm917@cam.ac.uk)

The availability of data at large volume, variety, velocity and veracity, often termed as 'Big Data', had a big impact on marketing research [1] and practice [2]. The wealth of personal information available about consumers online makes it possible to understand and cater to the individual needs of consumers better than ever before. Whether it is their Spotify playlists, Facebook profile, Google search queries, or mobile location, the digital footprints consumers leave with every step they take in the digital environment create extensive records of their personal habits and preferences. By tapping into this rich pool of consumer data, businesses can enhance consumers' experience by better matching the marketing offering to consumers' preferences and do so at the appropriate moment.

Applications of Big Data in marketing have largely focused on (a) assessing customers' preferences [*e.g.,*3], (b) predicting what customers are most likely to buy next [*e.g.,*4–6], (c) improving targeted advertising [*e.g.,*7,8], (d) understanding brand perceptions [*e.g.,*9,10], and (e) describing the competitive landscape [*e.g.,*11]. See Wedel and Kannan [1] for a review. However, investigations of how Big Data can help inform some of the more psychological aspects of consumer behavior that is aimed at understanding – rather than merely predicting – consumer attitudes and emotions has thus far only received scant attention. Davenport *et al.* [12] (2001, p. 63) note that holding vast amounts of customer data might help businesses to 'know more about their customers' but does not necessarily allow them to 'know the customers themselves'. The focus of this paper is to highlight the existing work and discuss the potential of using Big Data as a means to better understand consumers' stable psychological traits as well as more malleable psychological states.

New sources of consumer information

Traditional approaches to gathering 'human-centric' consumer information include extensive customer surveys, focus groups, interviews, observation studies and limited scope secondary data such as scanner panel data [1]. For example, as part of the Nordstrom's Personal Touch program, personal shoppers recorded detailed information on customers likes and dislikes, their lifestyle and tastes through telephone and face-to-face conversations as well as observations made in the store [12]. While the outlined approaches can generate valuable customer knowledge, they are not only expensive and time-consuming – and therefore difficult to scale – but also prone to numerous well established response biases [13]. For example, even the most motivated customer will find it difficult to accurately recall the purchases they made over the past four weeks or the exact feeling they experienced when purchasing a specific product.

Thanks to technological advances in the collection, storage and analysis of large amounts of data, businesses can now gain valid insights on millions of consumers by looking at the digital records that are passively collected as consumers go about their daily lives. In fact, observing the behavior of a consumer in a traditional retail store is very similar to analyzing the journey of a customer who is browsing a company's online store (*e.g.*, one can examine the characteristics of products the user has looked at and/or bought, measure the time they took to make a decision, or implement mouse-tracking technologies to study the decision process). Similarly, customer forums, product

reviews and posts in social media make it possible to observe large and natural 'focus groups' at very little to no cost [11].

The sources of information businesses can tap into to learn more about their consumers are almost limitless, and it would go beyond the scope of this paper to discuss all of them in detail (for an overview see Wedel and Kannan [1••], Fig. 2). Among the most vital ones are historical purchasing data, credit card records, search queries, browsing histories, blog posts, social media profiles, and smartphone sensor data (e.g., GPS location). Importantly, it is often possible to combine the information extracted from different sources to form a more holistic picture of a consumer's daily habits and preferences. By integrating information obtained from a consumer's social media profile, their phone logs and sensor data as well as their credit card spending, for example, one can get a fairly accurate picture of *what* a consumer has done *when* and with *whom*.

These new sources of data not only come from various sources, but they also come in multiple formats. While traditional data have been primarily structured in a numeric format, social media data, are primarily unstructured including, text, images, audio and video. Accordingly, different analytical approaches are needed to convert such data into knowledge and insights.

Turning Big Data into human-centric customer knowledge

The task of turning vast amounts of – often unstructured – data into insightful consumer knowledge is not easy and often requires the application of analytical techniques that are outside of the standard methodological tool box of consumer behavior researchers [14]. However, recent years have seen the rise of so-called computational social science research, a discipline aimed at applying methodologies from the computer sciences to questions asked by social scientists [15]. While the range of possible applications of such methodologies to social science questions is bounded only by the creativity and imagination of the researcher, here we focus on two types of insights that have recently attracted a considerable amount of attention among researchers and practitioners alike: the prediction of (1) relatively stable *psychological traits* that help explain consumers' general tendency to think, feel and behave in a certain way, and (2) malleable *psychological states* that express consumers' attitudes and emotions in-the-moment and help to put their behavior in context.

Predicting consumers' psychological traits

The investigation of stable psychological traits such as personality, regulatory focus, or need for cognition, has a long-standing tradition in consumer behavior research [16]. One of the most consistent findings suggests that consumers show more positive cognitive, emotional and behavioral responses to products, brands or marketing messages that match their own psychological traits [e. g.,17–20]. For example, an extroverted and open-minded consumer might experience more positive emotions and report a higher intention towards a retail brand that specializes in flashy and unusual clothes, or that uses extroverted and creative language to advertise their products (e.g., 'Stand out from the crowd and feel unique with our latest spring collection'). Businesses have long used such insights for branding and advertising purposes [e. g.,21].

However, because unlike demographics and past purchases, latent psychological traits cannot be observed directly, the opportunities to target consumers and personalize advertising based on psychological traits have been limited. If a mobile phone provider, for instance, decided to create a strong extroverted brand, it was very difficult to focus its advertising efforts on extroverted consumers short of choosing media channels (e.g., TV shows) that are predicted based on questionnaires or managerial judgement to have a larger proportion of extroverts. Instead, the branded marketing message had been primarily focused on mass marketing, broadcasting to large and heterogeneous audiences, thereby limiting its effectiveness.

In the age of Big Data, however, psychological traits – including personality, IQ and political orientation – can be accurately predicted from consumers' digital footprints. Researchers have demonstrated the ability to accurately infer personal traits from (a) personal websites [22], (b) Facebook or Twitter profiles [23,24••,25], (c) blogs [26], and (d) language use [27•,28••,29,30]. This digital form of psychometric assessment promises to be a game changer in the application and empirical evaluation of psychographic marketing. In an early pioneering study, for example, Hauser and colleagues inferred cognitive styles (e.g., analytic vs. emotional) from clickstream data and showed that matching a website's 'look and feel' to consumers' dominant motivational orientation can increase sales by up to 20% [7]. Similarly, Matz and colleagues showed that inferring the personality of Facebook users from their Likes, and matching the content of real advertising campaigns (products and marketing messages) to their dominant personality traits can significantly increase click-through and conversion rates [31••]. As the digital assessment of psychological traits becomes more widespread and readily available (e.g., LIWC for computerized text analysis; ApplyMagicSauce and StatSocial for personality predictions), consumer behavior scholars will be able to build on this early research and test the effectiveness of psychographic targeting in different domains (e.g., retail, charitable giving, political campaigning) and channels (e.g., social media, email, in-store), using different psychological traits (e.g., personality, cognitive style, motivational orientations), and different