Received: 17 November 2020 | Revised: 19 November 2020 | Accepted: 19 November 2020

DOI: 10.1002/arcp.1068

REVIEW

CONSUMER PSYCHOLOGY REVIEW WILEY

# Predicting consumers' choices in the age of the internet, AI, and almost perfect tracking: Some things change, the key challenges do not

David Gal[1] | Itamar Simonson[2]

[1]University of Illinois at Chicago, Chicago, IL, USA

[2]Stanford University, Stanford, IL, USA

**Correspondence**
David Gal, University of Illinois at Chicago, Chicago, IL, USA.
Email: davidgal@uic.edu

**Abstract**

Recent technology advances (e.g., tracking and "AI") have led to claims and concerns regarding the ability of marketers to anticipate and predict consumer preferences with great accuracy. Here, we consider the capabilities of both traditional techniques (e.g., conjoint analysis) and more recent tools (e.g., advanced machine learning methods) for predicting consumer choices. Our main conclusion is that for most of the more interesting consumer decisions, those that are "new" and non-habitual, prediction remains hard. In fact, in many cases, prediction has become harder due to the increasing influence of just-in-time information (user reviews, online recommendations, new options, etc.) at the point of decision that can neither be measured nor anticipated ex ante. Sophisticated methods and "big data" can in certain contexts improve predictions, but usually only slightly, and prediction remains very imprecise—so much so that it is often a waste of effort. We suggest marketers focus less on trying to predict consumer choices with great accuracy and more on how the information environment affects the choice of their products. We also discuss implications for consumers and policymakers.

**KEYWORDS**

AI, Choice Prediction, Conjoint Analysis

## 1 | INTRODUCTION

Marketing scientists have believed for a long time that techniques can be developed that can rather accurately predict consumer preferences and purchase decisions (e.g., Bass, 1993; Converse, 1945). Recent advances in machine learning, big data, and other tools and technologies seem to have bolstered this view, with claims that new methods are allowing, or will soon allow, firms to predict exactly what we want when we want it and even manipulate our behavior. Recent documentaries (e.g., The Great Hack, The Social Dilemma) and numerous other sources have warned society of the dangers of marketers' ability to know exactly what consumers want and what tailored messages can get them to do what these sophisticated data analysts and those who hire them order.

Consider the following recent claims:

> ...even data freely given harbors rich predictive signals... It isn't only what you post online, but whether you use exclamation points or the color saturation of your photos... people have become targets for remote control... "We are learning how to write the music," one scientist said, "and then we let the music make them dance." (Zuboff, 2020)

> [Technology Companies] have a data voodoo doll, which is a complete digital representation of our lives. With it, they can manipulate our behavior. Technology investor Roger McNamee quoted in Johnson (2019)

  

> Cambridge Analytica had the specific intent of skewing voter behavior [through psychographic targeting] in service of a right-wing agenda and putting right-wing politicians in office… it's sobering how much it succeeded at its goals — and how much wreckage was left in its wake. Analytica played a key role in both Brexit and the American presidential election in 2016… (Hasan, 2019)

> Much like in the scene in the science fiction movie Minority Report, where advertising billboards are personalized to the emotional state of the person walking past them, businesses will be able to optimize the advertising a consumer is exposed to in real-time and at a level of detail never before possible. For example, one could use information about a person's momentary heart rate extracted through their headphones to determine which song to play next, extract emotions from a person's facial expression to change the color scheme of a website, or recommend the next tourist attraction in a new city as a function of the person's predicted personality and their current level of physical activity. (Matz & Netzer, 2017)

> Would you like fries with that? McDonald's already knows (Yaffe-Bellany, 2019)

How credible are such statements? More generally, how concerned should we be about the ability of marketers to know exactly what consumers will choose and the messages and tools needed to get them to make specific choices?

To address such questions, it is important to first understand how consumers make choices, particularly in the current information environment in which they have access to an unprecedented amount of information at the time they are making decisions. To begin, therefore, we briefly review past research on how consumers make choices and then discuss how the way they make choices is likely to be affected by the current information environment. Afterward, in light of our understanding of how consumers make choices, we review evidence regarding the accuracy of contemporary methods in predicting consumer choices. We consider both traditional methods (e.g., conjoint) involving measurement of pre-existing (presumably stable) preferences to predict future choices as well as more recent approaches (e.g., predicting future choices from past choices through the use of sophisticated machine learning methods and big data). Following our review, we discuss some tentative conclusions regarding the capabilities and limits of contemporary and currently foreseeable methods for predicting consumer choices. We also discuss implications for marketers, consumers, and policymakers.

## 2 | HOW DO CONSUMERS MAKE CHOICES?

As discussed above, to understand the capabilities of preference prediction tools, it is important to first understand how consumers make choices. The classical economic view of how choices are made, based on rational choice theory, assumes that consumers have stable, well-defined preferences that precisely determine their choices (Von Neumann and Morgenstern 1944). It thus assumes that consumer choices are invariant to how options are described, framed, and preferences are elicited, inter alia. Evidence of even systematic deviations from the theory has often been dismissed as exceptions or anomalies, rather than the rule (Tversky & Thaler, 1990).

While this conception of how choices are made is useful for the tractability of mathematical models of behavior, a great deal of research has shown that it does not describe how consumers make decisions in general (e.g., Bettman et al.., 1998). To be sure, in some cases consumers do have strong, precise, stable preferences for particular products or attributes, and they may habitually buy the same options. For instance, some people prefer to buy a 2% organic milk. Likewise, a few consumers may have a self-imposed rule as to the highest price they are willing to pay for a water bottle, which prevents them from buying water at airports. When preferences for products or attributes are strong, stable, and precise, consumer choices are relatively easy to predict, such as by simply asking consumers about their preferences.

However, most of the choices made by consumers that are not habitual or routine are not the result of precise, stable preferences for those products, but are constructed (or discovered) at the time a decision is being made on the basis of interactions among many individual and situational factors (e.g., Bettman et al., 1998; Lichtenstein & Slovic, 2006). As indicated, preferences have been referred to as "constructed" to reflect the fact that individuals often do not have precise, stable preferences for specific products or attributes but derive them in the context of a particular choice task. It also reflects the fact that individuals may lack insight into their preferences and do not know what they want when they encounter a choice task (e.g., Simonson, 2005). As such, many choices, particularly those that are not regular or that are being made for the first time, require consumers to assess options, make trade-offs, and reflect on the value they might derive from different attributes. Given that such choices are only crystalized near the time a decision is made, they are particularly susceptible to influence, including by sources of information (e.g., product reviews, recommendations), social influence, the manner in which options are presented or framed, and so forth.

To illustrate, a person is unlikely to have a preference stored in memory for a particular model or configuration of toaster or for trading off a particular toaster feature for a given discount. Rather, when buying a toaster, consumers might have a tendency to purchase household items on Amazon and thus might expose themselves only to toasters available on Amazon. They might not have given much thought to purchasing a toaster that can toast four slices at once, but after seeing some of these models, they might appreciate the value of such toasters to their lifestyle and subsequently only consider those models. They might remember that they recently purchased some expensive household items, leading them to prefer to buy a budget toaster, and thus will only consider the lower-priced models. Among the models that meet these criteria, they might pick the first one that Amazon displays that features both a design they like and largely positive reviews. Ultimately, the person's choice will reflect an interaction among multiple individual and situational tendencies, transient mental states (e.g., mood), the available information before and at the time the decision is made, and other factors rather than a pre-existing preference for a particular toaster.

The constructive nature of consumer choice has important implications for choice prediction. Notably, it suggests that any method (like conjoint analysis) that relies on measuring preferences at one point in time to predict future choices, or any method that uses past choices to predict future choices (like common predictive analytic methods), will be limited by the fact that preferences for specific products and attributes tend not to be formed until the time a decision is being made. Moreover, as we discuss next, the limitations of these choice prediction methods are likely to be amplified in the contemporary information environment.

## 3 | THE INFLUENCE OF THE CONTEMPORARY CONSUMER INFORMATION ENVIRONMENT ON CHOICE

As illustrated by the quotations in the introduction to this article, much recent attention has been devoted to how the current information environment, in which consumers are tracked and their choices are analyzed by algorithms, will lead to increases in the accuracy of choice prediction. Comparatively little attention, by contrast, has been devoted to how the unprecedented amount of information now available to consumers at or near the time decisions are made will affect the ability of marketers to predict their choices (for exception, see Simonson & Rosen, 2014).

How might, in fact, the increasing amount of information available to consumers at the time they are making decision influence their choices and thereby marketers' ability to predict them? Given the constructive nature of choice, as described above, it serves to reason that information encountered at the time decisions are being made, and therefore at the time preferences for particular products or attributes are being formed, is likely to have an outsize influence on choice. To elaborate, consumers are increasingly likely to encounter important informational influences at the point of decision. These can take the form of expert and user reviews, detailed information about product features, online recommendations, information about potential options readily available to the consumer that the consumer was not previously aware of, and so forth. For example, while browsing the wine aisle of a grocery store, consumers might solicit recommendations from a mobile app; likewise, while buying a grill, they might learn of a seemingly trivial product feature they otherwise might never have considered through reading online reviews. As another example, in buying a toaster, at the tap of a screen consumers could view, sort, browse, and instantly purchase any of hundreds of models. To the extent consumers' preferences are being constructed at the time the decision is being made, as they are in the case of many non-routine decisions, then consumers will be particularly susceptible to being influenced by such information. That is, information that becomes available closer to the time a decision is being made—that is, at the time preferences are being constructed—is likely to have an outsize influence on choice.

This reality is likely to make choice prediction more difficult. This is because informational influences encountered for the first time when a decision is being made cannot be measured or anticipated ahead of time to predict choice (the basis of conjoint analysis and other preference/utility measurement methods). Likewise, past choices are unlikely to be good predictors of future choices when consumers' decision-making heavily relies on information that they first encounter at the time they are making a choice (and where the information they encounter is likely to vary substantially from decision to decision). Thus, insofar as consumers' decisions are increasingly influenced by inputs encountered for the first time close to when a decision is made, as in the current consumer information environment, the task of choice prediction becomes increasingly challenging.

## 4 | TRADITIONAL UTILITY MEASUREMENT-BASED METHODS: THE CASE OF CONJOINT ANALYSIS

Traditional utility measurement-based methods of choice prediction attempt to measure consumers' utilities for product attributes through surveys or experiments in order to predict their future choices. Conjoint analysis and choice-based conjoint, in particular, are among the more popular of these traditional tools used to measure consumers' preferences and predict their choices. In conjoint analysis, consumers are asked to make a series of choices or provide overall evaluations of different product profiles consisting of different "bundles of attributes" (Green & Rao, 1971; Green & Srinivasan, 1978). These evaluations can take the form of ratings, rankings, or, most commonly, choice ("choice-based conjoint"). For example, a sample of consumers might be asked to make a series of choices among 2–4 laptop options that vary on a number of attributes, such as weight, memory, display resolution, processing power, brand, and price. Based on such choices/evaluations, consumer preferences for options are then decomposed into utilities or "part-worths" for attribute levels. Consumers' preference (or utility) for any combination of attributes—including combinations that were not directly evaluated by the consumers—can then be computed from these part-worths (Ben-Akiva et al., 2019).

Besides being widely used in marketing since the early 1970s, conjoint analysis has also been relied upon outside of marketing. For instance, it has been used to assess preferences in healthcare settings (Bridges et al., 2011), to assess the preferences of voters (Hainmueller et al., 2014), and to measure consumer preferences in order to determine damages in legal disputes (Sidak & Skog, 2015). It is noteworthy that the mere fact that conjoint has been used for decades has often been used as "evidence" that it must be a reliable and externally valid measure of actual consumer choices, despite very few attempts at actual real-world validation (and even those confined to a narrow context, more on this later). For example, Hainmueller et al. (2014) write, "...conjoint analysis is widely used by marketing researchers to measure consumer preferences, forecast demand, and develop products... these methods have been so successful that they are now widely used in business as well"; see also, e.g., Arning, 2017; Itsubo et al., 2004; Kuzmanovic & Martic, 2012; Moise et al., 2018.

Despite the very limited external validation of conjoint's predictive accuracy, it has often been used for two related reasons. First, because respondents in conjoint surveys are tasked with evaluating combinations of attributes together (the term conjoint is derived from "consider jointly") rather than rating or ranking the importance of individual attributes, conjoint analysis—and choice-based conjoint in particular—is assumed to be more realistic in terms of mirroring real choice behavior (Natter & Feurstein, 2002). Second, although consumers often lack insight into how individual attributes or attribute levels affect their choices, observing their evaluations of product profiles involving different combinations of attributes enables a researcher to uncover the value of individual attributes as compared with simply asking consumers about the value of those attributes. For instance, Simon (2018) states, "It doesn't make sense to ask consumers directly for the utility or their WTP, as they aren't able to give a direct and precise estimate. The most important method to quantify utilities and WTP is the conjoint analysis" (p. 53).

As indicated, there is little to no evidence that these assumptions hold or that conjoint is superior to other methods that reveal attribute values by requiring respondents to make trade-offs (e.g., constant-sum; Malhotra, 2010). Moreover, as explained above, the conjoint notion that stable preferences exist and can be used to predict choices and values under different conditions misrepresents the manner in which decisions are made, particularly in the current consumer information environment.

So how useful is conjoint for predicting consumer choices? It is widely recognized that conjoint's validity largely depends on its ability to represent the essential characteristics of choices in reality, including capturing the key decision criteria, values, and information that consumers will consider when making decisions, as well as the manner in which actual decisions are made (Ben-Akiva et al., 2019). To evaluate the ability of conjoint to capture how real-world choices are made, we consider how the manner in which choices are made in conjoint analysis studies compares to how choices are made in the real world. Afterward, we review the empirical evidence regarding the predictive accuracy of conjoint, with a particular emphasis on studies that have compared the accuracy of conjoint to that of simpler methods.

### 4.1 | Can conjoint capture typical real-world choices?

As noted above, methods such as conjoint, that depend on measuring preferences at one point in time to predict future choices, are limited by the reality that consumers' preferences tend not to be fully formed until the time a decision is being made. Nonetheless, we can surmise that the more a conjoint study mimics real options and choices, the more valid its predictions are likely to be. In this regard, the attributes and information (e.g., reviews) included in the study design and their presentation to respondents should reflect the attributes and influences consumers rely on and the manner in which consumers encounter them when making real decisions. Notably, if attributes or other influences that significantly drive purchase are not or cannot be included in the design, such as certain specifications, "coolness," uniqueness, user-friendliness, sales promotions, advertising, social influence, or user ratings, then the validity of the results must be called into question.

Furthermore, it is recognized by conjoint practitioners that misrepresenting reality by omitting important attributes while focusing attention on just a few generates "focalism," which inflates the measured value of the focal attributes (i.e., those that differentiate the presented options in choice-based-conjoint; Tomlin & Zeithammer, 2018). This does not mean that all attributes must be included in the design, but the important dimensions should be included to allow for reliable predictions of consumer preferences. Indeed, Ben-Akiva et al. (2019) note that contrary to the argument of some conjoint practitioners, conjoint can accommodate a large number of attributes (p. 21). Also, of importance, it is recognized that respondents must be familiar with similar products or attributes to those in the conjoint study; otherwise, their responses will be highly unreliable (Ben-Akiva et al., 2019).

While it is true, as described above, that limitations of preference measurement using conjoint and similar methods have been recognized, the severity and breadth of these limitations have often been ignored (perhaps because of a tendency to overemphasize precise, quantitative measures). Indeed, in some of the literature on conjoint the underlying assumption appears to be that the limitations of conjoint mostly arise because the context of conjoint tasks tends to differ from the context in which people make real-world decisions and that by bridging this gap these limitations could be avoided (see Ben-Akiva et al., 2019). However, conjoint and similar attribute utility measurement methods fundamentally depend on the assumption that stable preferences for prespecified attributes determine choices and that these can be measured through asking consumers to explicitly or implicitly compare attributes (e.g., through making choices among different "bundles of attributes"). The methodology is therefore limited by its inability to capture the way preferences are formed—including the inability to match the decision process involved in actual choice and the inability to identify or measure many of the inputs that affect the formation of preferences at the time of decision. In fact, even for simple decisions involving commonplace, familiar product categories, the mismatch between how consumers form preferences when they are being measured in a conjoint study and how consumers make real-world choices is liable to be high.

For example, if consumers tend to choose based on one factor (e.g., the toaster can toast four slices of bagel simultaneously) yet the preference measure leads respondents to consider multiple features and make trade-offs among them, then the resulting predictions are likely to be misleading. Indeed, by having respondents make a series of choices among options that differ on a number of attributes, a typical conjoint design inherently pushes respondents toward a decision strategy focused on making trade-offs among attributes and toward doing so in a relatively consistent manner across choices. That is, because choice-based conjoint typically presents the options using an attribute X option matrix format, it encourages making trade-offs and using compensatory processing. Indeed, it is well established that the manner in which options are presented tends to strongly affect the evaluation strategy (e.g., Bettman & Kakkar, 1977). Thus, for example, even if consumers tend to use a more lexicographic decision strategy in actual choices and rely primarily on the brand (e.g., Bettman et al., 1991), the matrix format encourages a more compensatory rule that distorts the decision process.

Further, because respondents know their preferences are being measured, in mundane and low-involvement categories, they are likely to be more reflective about their choice process than they would be when actually making a choice. At the same time, in high involvement categories, respondents are likely to be less motivated to rigorously evaluate the options and attributes during measurement than during actual purchase. Overall then, measuring preferences through conjoint is likely to involve a very unnatural decision process.

Moreover, the inputs to measured choices using conjoint and actual purchase are likely to be very different. For instance, when purchasing a car, individuals might rely on the recommendations of friends or might simply visit the nearest dealer of a trusted brand, thereby making their considered options much more limited compared to the universe of options they might be exposed to in a conjoint measurement task. As noted, these limitations are likely to be amplified in the current consumer information environment. Simonson and Rosen (2014), for instance, provide the following example:

> Think of a guy named Jim who agreed to participate in [a conjoint] study. He is presented with several product combinations and is asked to make some choices: Do you prefer a Samsung laptop with 2 GB of RAM, 80 GB hard drive and 15.6-inch screen? Or would you rather have an HP with 4 GB of RAM, 60 GB hard drive, and 11.6-inch screen? After many similar questions that require Jim to make such choices, the market research firm uses sophisticated statistical techniques to derive the relative importance of different attributes. This is all very nice. But what happens in reality when Jim is ready to buy his next laptop? He goes on CNET, Amazon, Decide.com, BestBuy.com, gdgt.com, or similar sites to read what others have to say. He's naturally attracted to the laptops with the highest ratings and scores (which are usually the first thing you see on these sites). When he starts reading reviews, he may be sidetracked by a new feature or consideration. A friend on Facebook posts something about her new laptop that takes Jim in yet a different direction.

As noted by Simonson and Rosen, the effects of such just-in-time information on consumer decisions do not only impact the usefulness of conjoint analysis to predict preferences but also other tools involving stated preference measures, such as brand equity measures or measures of consumer willingness to pay for a product.

In sum, the way preferences are elicited in a typical conjoint task is likely to miss many key inputs to choice and to skew the relative importance of the inputs it does identify. While these limitations might be remedied to some degree through improving the design of conjoint studies, they are largely a function of the nature of conjoint tasks which, at a fundamental level, do not replicate the manner in which most choices are made. Moreover, although we have examined these limitations in the context of conjoint analysis, they are not unique to conjoint, but would apply to other methods that aim to measure attribute utilities at one point in time in order to predict future choices.

### 4.2 | (The lack of external) empirical validation of conjoint

Despite its wide popularity, validation of conjoint (e.g., in terms of its predictive accuracy in reality, particularly compared to simpler, more direct methods) is surprisingly limited. In fact, most tests of conjoint simply collect data from two different conjoint tasks and use estimates from the first task to make predictions and measure the accuracy of those predictions in the second task. As Kamakura and Ozer (2000) point out, these are essentially test–retest reliability assessments, not validation tests. Some conjoint users tend to rely on simulations, which cannot be more reliable than the data they rely on.

Some exceptions include tests of the predictive accuracy of conjoint in predicting the job choices of MBA students. Perhaps surprisingly, these tests found that simpler methods yielded slightly higher predictive accuracy than the far more popular, and more complicated, conjoint procedure. For instance, Srinivasan (1988) finds that MBA students' simple ratings of the relative importance of attributes and attribute levels (i.e., a "self-explicated approach") of jobs provided slightly better predictions than conjoint analysis did of their real job choices. Likewise, Srinivasan and Park (1997) found the simple self-explicated approach provided slightly better predictions of MBA students' actual job choices than combining self-explicated ratings with conjoint analysis. Notably, this was in contrast to the performance of the two methods in predicting respondents' choices among hypothetical validation profiles, where the conjoint approach was more accurate than the self-explicated approach (probably because the validation profiles and profiles used to assess respondent preferences as part of the conjoint procedure were very similar and thus were more like test-retest reliability measures, as per above). Srinivasan and Park concluded, *"surprisingly, and contrary to our expectations, the best predictive validity came from the simple self-explicated approach*. [emphasis in original]*"* (p. 290).

Other studies involving actual choices of jobs (Huber et al., 1971) and colleges (Wright & Kriewall, 1980) also showed greater predictive validity for self-explicated approaches over conjoint. Likewise, research has generally not shown an advantage for conjoint over the self-explicated approach in predicting actual market share data (Hensel-Borner & Sattler, 2000).

In terms of overall predictive accuracy, Srinivasan and Park (1997) found that the self-explicated approach in which respondents simply rate the relative importance of different attributes and attribute levels correctly predicted 76% of MBA students' job choices, compared to 72% for their customized conjoint approach. These were both superior to the 36% expected accuracy if choices were to be predicted randomly. In other words, there was value to eliciting respondents' stated preferences in terms of enabling more accurate predictions of their choices relative to chance alone. However, the value appeared to come from the relatively simple task of assessing relative attribute importance rather than from any specific aspects of the conjoint process, such as its relatively complex decompositional approach to identifying part-worths.

Moreover, the value of eliciting stated preferences illustrated in Srinivasan and Park's (1997) study needs to be caveated by a few points. As an initial matter, while predicting 76% of respondents' choices is better than predicting randomly, it does not seem particularly high in an absolute sense. Second, it is unclear how these findings might translate to other contexts. Of note, several of the attributes in Srinivasan and Park's (1997) study depended on respondents' subjective construals, such as whether the job offered opportunity for advancement (rapid versus. moderate), the work environment (desirable versus. reasonable versus. unattractive), and salary (Expected salary + 20% versus. Expected salary + 10% versus. Expected Salary −10% versus. Expected salary −20%). This is obviously very different from typical marketplace behavior in which a marketer must make predictions based on objective attributes rather than on how they are subjectively construed by individual consumers.

Along the same lines, the predictive accuracy of methods that measure respondents' utility for attributes, as in conjoint, is likely to be higher for a category like job choices where preferences are probably better defined and more stable than those for many typical consumer choices such as what laptop or toaster to buy, what financial app to use, or what movie to watch. Indeed, as noted above, the ability to predict choices for typical consumer products through conjoint or other preference measurement tools is likely to be significantly diminished in the contemporary information environment where many important influences on decisions are only encountered at the time of decision.

Finally, while it is useful to understand the degree to which measuring individuals' utility accurately captures the actual preferences of those individuals, the results do not tell us how accurately these measures would predict the choices of a broader population. In this regard, Hainmueller et al. (2015) examined how well conjoint predicted choice (voting preference) in a referendum on awarding Swiss citizenship to various applicants. Their best performing conjoint design predicted that 21% of applicants would be rejected, whereas in actuality 37% were rejected. This seems to reflect low predictive validity.

#### 4.2.1 | Willingness to pay

Besides being used to predict choice shares, conjoint is sometimes applied to estimate consumers' willingness to pay for products. A recent meta-analysis specifically compared the validity of conjoint and similar indirect methods to more direct methods (e.g., simply asking consumers for their WTP) for eliciting consumers' willingness to pay for a product (Schmidt & Bijmolt, 2020). The main finding was that indirect methods such as conjoint overestimate WTP to a significantly greater degree than do more direct methods. Thus, overall, evidence for the advantages of conjoint over simpler methods is lacking.

#### 4.2.2 | Variations in procedure and estimation methods

In addition to examining the predictive validity of the conjoint methodology broadly, other studies have examined differences in the predictive validity of different conjoint estimation strategies (i.e., different methods for calculating the part-worths) or approaches (e.g., rating-based versus. choice-based). Using hypothetical choices of real automobiles, Moore (2004) found that choice-based conjoint performed no better than rating-based conjoint in predicting aggregate market shares. Likewise using studies involving hypothetical choices, Hagerty (1986) and Green (1984) found that more complex models involving interactions among relative preferences for attributes actually made individual-level prediction (i.e., the ability to predict the choices of individuals) worse. Natter and Feurstein (2002), examining 43 stock keeping units (SKU's) of four products, found, using choice-based conjoint, that while accounting for individual heterogeneity in preferences outperformed in hold-out samples in terms of predicting aggregate market share, accounting for individual heterogeneity provided no advantage in predicting real-world market share. These findings suggest that, at least in many cases, increasing sophistication in estimation or procedures provides little added value.

In sum, despite its popularity, validation of conjoint has been rather limited. Moreover, the existing evidence from validation studies suggests that conjoint does not yield predictions of very high accuracy in an absolute sense (e.g., WTP in dollars and cents) and that it generally does not perform better in predicting preferences than simpler, more direct stated preference methods. Likewise, it does not appear that increased sophistication in methods or estimation strategies yields material improvements in prediction. In a nutshell, the key takeaway from conjoint validation studies can be summarized by the statement that the value of conjoint is in crude identification of the relative importance of attributes and attribute levels (those that are easily specified, familiar to consumers and that can be anticipated ex ante).

While we have specifically examined the validity of conjoint due to the popularity of the methodology, it is important to recognize that the limitations we identified for conjoint extend to other methods that involve preference/utility measurement at one point in time to predict future choices. As a result, like conjoint, the most we should expect of such methods is that they will accomplish the relatively easy task of identifying the relative importance of easily specified attributes in a crude way.

## 5 | PREDICTING FUTURE CHOICES FROM PAST CHOICES (AND OTHER BEHAVIOR): HOW MUCH PROGRESS IS BEING MADE?

While conjoint analysis and similar methods aim to predict consumer choices through measuring pre-existing utilities for attribute values, predictive analytics typically refers to the use of statistical methods to predict future behavior, such as choices, from historical behavior (e.g., past choices, browsing history). Simple linear regression is the most basic tool for this purpose though many more sophisticated tools have been developed, including increasingly sophisticated machine learning methods, typically combined with "big data." These methods yield predictions through identifying patterns in data.

As noted in our introduction, claims made around recent "AI" technologies, mostly advanced machine learning methods, is that they will facilitate leaps in the ability to predict consumer preferences, even allowing marketers to make consumers "targets for remote control" (Zuboff, 2020). Machine learning methods can be differentiated from traditional econometric methods in that they involve algorithms that iteratively build models from data (rather than the models being specified ex ante) to maximize the amount of variance explained (Dzyabura & Yoganarasimhan, 2018). The focus of machine learning algorithms is thus to maximize the accuracy of out-of-sample predictions typically without any underlying theory. This is different from traditional econometric methods that are focused on causal inference and identifying the best unbiased estimators.

There is no doubt that advances in machine learning (in concert with big data) are revolutionizing many tasks, such as image recognition, language translation, chess playing, and spam detection. Many of these advances are the result of "deep learning" (or "neural") machine learning methods in which information is processed through a hierarchy of layers (Bengio et al., 2015).

What is the prospect that we might see similar advances from the use of such methods in choice prediction? It is important to recognize that predicting preferences (and human behavior generally) is very different from the tasks at which sophisticated machine learning algorithms excel, such as perception (e.g., image recognition) or judgment (e.g., spam detection), in which outcomes are generally well-defined by the available data (Narayanan, 2019). Unlike the objects of perception and judgment tasks, as discussed above, preferences for specific products and attributes tend to be formed at the point of decision, and they are determined by tendencies and situational influences that cannot be precisely measured or even anticipated ahead of time. Accepting that preferences and choices of particular products or attributes are often determined or crystallized near or at the time of decision, the questions we are interested in here are as follows: how accurately can choices be predicted through predictive analytics and how much value is there to increasing methodological sophistication, such as that represented by sophisticated machine learning methods and big data?

We next review the evidence regarding the strengths and limitations of data collection tools and predictive analytic methods for predicting consumer choices in light of our understanding of how consumers make choices. In doing so, we highlight how the predictive accuracy obtained through more sophisticated methods and big data compares to that obtained through simpler methods and data.

### 5.1 | Tracking and data collection

Among the tools that enable marketers to predict consumers' choices are those that allow marketers to capture data about consumers. Indeed, marketers today can obtain a tremendous amount and variety of information ("big data") about users, including both demographic information and behavioral information, such as highly detailed web-browsing and purchase history (Narayanan, 2018). They can link behavior across websites and applications (e.g., through third party trackers) and tie online to offline behavior (e.g., through data brokers; Narayanan, 2018). Marketers can also track an individual user across devices (de Haan et al., 2018), including through their television and streaming services (Schweidel & Moe, 2016), and they can track users' position (Fong et al., 2015) and trajectory (Ghose et al., 2019) through their mobile devices. By tracking people, such as through their online search and shopping behavior, marketers can identify people's secrets, such as undisclosed sexual orientation (Stephens-Davidowitz, 2017) or pregnancy (Duhigg, 2012).

Despite the availability of such detailed information, as the amount of consumer data available to marketers has proliferated, significant limitations to marketers' ability to collect relevant data on consumers remains. In particular, to-be-constructed preferences and internal psychological states and thoughts do not leave explicit digital traces and remain mostly inaccessible to tracking; likewise, exposure to various persuasive communications and reviews and most interpersonal interactions tend not to leave digital traces. Further, regulations, protocols, and tools allowing consumers to preserve some of their privacy limit tracking capabilities (Pujol et al., 2015). In addition, there are technical challenges associated with tracking people on digital devices, notably the problem of identity fragmentation whereby outcomes and exposures are observed at the level of devices and cookies rather than at the level of individuals (e.g., marketers might track the behavior on a particular streaming service, without being able to connect it to the specific individual in the household that watched each program; Johnson, 2020). Indeed, recent research suggests that user profiles assembled by data brokers tend to suffer from low accuracy (Neumann et al., 2019).

## 5.2 | Predicting choice in response to targeted messages (e.g., ads)

A question of obvious interest is whether collecting extensive data about consumers can allow marketers to accurately predict which individual consumers will be most receptive to pitches for their products? Traditionally, marketers have used data on consumer demographics, geography, and behavior, to target broad consumer segments with products and messages that match their (presumed) needs. To provide a rather obvious example, a marketer might pitch iPhone accessories to consumers that are known to have purchased an iPhone (rather than to consumers known to have purchased an Android phone). In other words, marketers have traditionally used consumer data to segment consumers and thereby increase the likelihood of pitching their products to the consumers most likely to buy them.

A question that arises is whether the more recently introduced data collection and analytic methods substantially enhance the accuracy with which the receptivity of particular consumers to targeted pitches can be predicted? While a degree of improved accuracy is likely in certain contexts, consistent with our above analysis, we believe that the fundamental characteristics of preference construction create a ceiling on improvements in terms of the efficiency of targeted ads and marketing offers.

Of course, the degree to which greater accuracy in predicting who will respond to an offer can be achieved is a function of many factors. To the extent more of our behaviors (purchases, likes, visits, searches, etc.) are tracked today, more relatively easy predictions can be made. For instance, knowing who recently bought prenatal vitamins is likely to increase the accuracy of predicting who will respond to ads for diapers (Duhigg, 2012). Likewise, knowing who previously browsed a product website increases the ability to accurately predict who will buy the product in response to an ad for it (i.e., behavioral re-targeting, Johnson et al., 2017b; Sahni et al., 2019). Such increases in predictive accuracy are not due to any particular methodological sophistication or "big data" per se, but due to the availability of data relevant to predicting the particular decision – similar to how marketers have used consumer purchase data to target offers for decades.

At the same time, there are contexts where contemporary methods and big data can improve the ability to predict what pitches are most likely to appeal to individual consumers. For example, contemporary machine learning methods are likely to be particularly useful for improving prediction when the available data are unstructured and diffuse and where the marketer lacks insight or expertise into relevant variables for targeting (see, e.g., Hu et al., 2019; Rafieian & Yoganarasimhan, 2020). In the latter regard, when the targeted group is overly broad, methods for identifying users similar to those users who clicked on an ad (or who purchased in response to an ad) can help advertisers (and advertising platforms) narrow the target to those within the broader group more likely to click on the ad (or purchase in response to it; Robinson, 1999).

Although improvements in prediction resulting from high levels of consumer data and contemporary methods can be economically meaningful, in most cases, the accuracy of predicting who will respond to a particular ad in an absolute sense still appears quite low. For instance, Johnson et al. (2017) sent display ads for an apparel retailer to 3 million Yahoo! users who were also customers of the retailer. Despite "exceptional consumer-level data," including demographics, ad exposure data, and two years worth of past purchase history, they were only able to predict 10% of the variance in sales data subsequent to ad exposure. They viewed these results as "disappoint[ing]." Likewise, Lewis and Rao (2015) found that knowing who is exposed to an ad (among those targeted by the ad) explained a barely perceptible share of consumer purchase behavior.[1] While Johnson et al. (2017) and Lewis and Rao (2015) used regression in their analyses, there is little evidence to date that more sophisticated methods yield dramatic improvements (Johnson, 2020). In another case that appears to illustrate limitations in the accuracy of choice prediction, P&G recently scaled back targeted advertising on Facebook because narrow targeting was leading them to miss out on many potential purchasers of their products (Terlep & Seetharaman, 2016). That is, they were not able to use microtargeting to pinpoint with sufficient accuracy who was most likely to respond to ads for their products.

The difficulty in predicting choice in response to ads aligns with our understanding of the constructive nature of consumer choices. Namely, choice is influenced at the time of decision both by relatively stable tendencies and by variable, essentially unpredictable factors. To the extent demographic, geographic, or behavioral variables (e.g., having previously purchased an iPhone), can capture strong, stable tendencies to choose particular options, they are likely to be useful predictors. However, to the extent options, information, and other influences at the point of decision are variable and cannot be anticipated, highly accurate prediction at the level of individual consumers will be challenging.

## 5.3 | Predicting preferences with recommendation systems

Besides the question of predicting sales in response to ads, a more general question is whether marketers can predict what items consumers will want and when they will want them. Recommendation engines are designed largely for such a purpose. Common approaches to recommendation involve collaborative filtering and content-based systems (Felfernig et al., 2014). In one variant of collaborative filtering (a "user-user" algorithm), items are recommended to a consumer that similar consumers purchased (or liked), with similarity between consumers typically determined by their having previously purchased (or liked) similar items. In another variant of collaborative filtering (an "item-item" algorithm), items are recommended to a consumer based on similarity to other items the consumer purchased (or liked) previously. Here, similarity between items can be determined based on the number of users that have purchased both items. In content-based systems, user profiles are matched with item profiles, so that users are recommended items similar to those they previously purchased or browsed. For example, users might be recommended songs by artists or from genres that they previously listened to.

Next, we consider the limits of such techniques given what we know about consumer choices and then review the empirical evidence regarding the performance of advanced recommendation algorithms in predicting consumer preferences.

### 5.3.1 | There are inherent limitations to predicting future choices from past choices

As noted, a key collaborative filtering approach is to recommend products to users that have been purchased/liked by similar users. The logic underlying this approach is that if two individuals like some things in common, they are likely to like other things in common. For example, if Person A and Person B share similar preferences across some products, perhaps Person A will also like products Person B has purchased but that Person A has not.

While collaborative filtering can, in some cases, lead to useful recommendations, it also has significant accuracy limitations. One major reason is that it does not account for the constructive nature of preferences. For instance, suppose Person A and Person B are judged to be similar based on their having purchased similar items in the past. As a result, if Person A recently purchased a toaster that Person B has not yet purchased, we can recommend that toaster to Person B. However, this assumes that Person A and Person B both share similar stable preferences for particular toaster models. However, given the constructive nature of consumer choice, it might be that Person A purchased the toaster not because it particularly matched her pre-existing preferences, but because the retail website where she purchased it happened to list it as one of the top results in a search for toasters, and because she was momentarily too busy to perform an exhaustive search. Or perhaps she was influenced to choose it because it was an intermediate option in terms of price and quality among the options she was presented (Simonson, 1989). In such a situation, there is little reason to suspect it would make a particularly useful recommendation for Person B.

Moreover, the notion of providing a precise and accurate recommendation to Person B assumes that Person B has precise and stable preferences to be predicted. However, as noted, the constructive nature of preferences suggests that while people have general preferences and tendencies (e.g., for variety, uniqueness, quality, ease-of-use, a favorite color), they often lack preferences for specific models or tradeoffs among attributes (e.g., for a specific configuration of toaster or for how much of a discount they would require to choose a slightly less attractive toaster). Such specific preferences usually will not exist until they are constructed at the time of decision, and how they are constructed will depend on many factors, including the availability and nature of the recommendations themselves.

Content-based approaches, where users are recommended items similar to those they purchased in the past, also have accuracy limits because of their inability to account for the constructive nature of choice. For example, the fact that a person purchased a particular brand of toaster does not necessarily reflect a preference for that brand, but could reflect the influence of, effectively random, just-in-time information, such as having encountered positive reviews for the particular toaster close to the time of purchase. As such, recommending the user other appliances by the same brand might not result in a particularly useful recommendation for the user.

While so far in this section we have described the limitations of collaborative filtering and content-based approaches, the same principles limit the predictive performance of any other recommendation approach that aims to capture stable preferences through observations of past choices. The constructive nature of choice, for instance, means that past choices are the reflection not only of stable preferences for attributes but of many other, effectively random and unpredictable, influences, including just-in-time information. Likewise, future choices will not be determined by stable preferences that can be predicted ex ante but will be constructed based on multiple factors at the time of decision. As such, we can expect that recommendations based on past choices can typically only offer a relatively crude match to (imprecise and noisy) preferences. This, in turn, suggests limited advantages to methods that aim to achieve high precision through additional sophistication.

### 5.3.2 | The case of the Netflix Prize

What does the empirical evidence show regarding the predictive performance of recommendation algorithms? It can be difficult to assess empirically how accurate recommendation engines are, and particularly how much improvement in predicting preferences and choice has come from newer, more sophisticated algorithms in recommendation engines. However, as an initial look at the evidence, it is useful to consider the Netflix Prize to improve the firm's customer movie recommendations, which was contested by multiple teams for 3 years. In 2009, one team finally won the prize after it achieved a 10% improvement in recommendation performance according to a preselected metric. This solution ended up not even being implemented by Netflix because of its complexity and the engineering resources it would have required relative to the returns (Masnick, 2012).

Moreover, the Netflix Prize, despite presumably examining the value of new methods in predicting preferences, was a poor test of preference prediction since it only predicted how individuals would rate movies based on their past ratings of other movies and the ratings of other users. That is, it did not involve predicting consumer choice (or even how a user would rank their enjoyment of one movie relative to another), making it unclear how it would have translated to such a task. As such, the winning solution might simply have been better at identifying which users were likely to rate every movie very favorably, which were likely to rate every movie very unfavorably, which were likely to rate every movie using extreme ratings, and so forth. In other words, it is unclear how the 10% improvement in predicting customer ratings might have translated, if at all, to improvements in predicting choice or relative satisfaction of watching one movie versus another.

Another limitation was that participants in the challenge had to work with a tightly constrained data set which offered limited information on relevant variables to choice, such as user demographics or whether the movie was recommended to the user; this likely inflated the predictive value of increased sophistication compared to a real-world environment featuring more relevant data (analogous to the focalism that can inflate the importance of attributes in conjoint studies, as described above). Yet another limitation was that unlike in the real world, the data set in the Netflix challenge was static, a known limitation in testing recommendation engines (i.e., results from such "offline" studies often fail to translate to online performance; Beel et al., 2013).

Given such limitations, it remains unclear how much algorithms such as those employed in the Netflix challenge increase the accuracy of preference prediction in the real world over simpler methods. It is notable, however, that even a decade after the Netflix challenge, improvements in the accuracy of Netflix's recommendations are still far from apparent. For instance, one technology writer, writing in Fast Company, recently wrote (Diaz, 2018):

> If you use Netflix, you're probably familiar with the "smart" recommendations of this system. It thinks I should watch its abysmal series Insatiable because I previously watched Ozark. It thinks the movie Inferno is a "98% match" because I watched one scene from Star Wars: Rogue One. This machine guessing game has turned my Netflix home screen–allegedly the site's prime method for content discovery–into a mosaic of titles that I emphatically don't want to watch. Every time I tried to follow any of its obscure suggestions, I find myself turning it off after 20 minutes. Which apparently triggers more recommendations of things that I don't want to watch.

#### 5.3.3 | Empirical validation of "progress" in recommendation engines

Consistent with casual observations such as the one related to Netflix's recommendations above, although many sophisticated algorithmic approaches, particularly those involving "deep learning" (also termed "neural" approaches), have been advanced in recent years as offering improved predictive accuracy for recommendation engines, recent reviews cast doubt on the validity of such claims (Dacrema et al., 2019; Ludewig, 2019). In one recent review, the authors found that out of seven neural recommendation algorithms presented at top level conferences, six were outperformed on the relevant recommendation task by "comparably simple heuristic methods" (Dacrema et al., 2019). The authors write, "Given the increased interest in machine learning in general, the corresponding number of recent research publications, and the success of deep learning techniques in other fields like vision or language processing, one could expect that substantial progress resulted from these works also in the field of recommender systems." However, they conclude that most of the supposed progress might, in fact, have been "phantom progress." They write, "Despite their computational complexity, our analysis showed that several recently proposed neural methods do not even outperform conceptually or computationally simpler, sometimes long-known, algorithms."

Another recent article reviewed progress specific to session-based recommendation engines; that is, engines tuned to make recommendations in a particular interactive online session in the absence of long-term history about the user (Ludewig et al., 2019). Comparing the performance of recent sophisticated neural methods to that of simpler methods, the authors conclude, "In the majority of the cases, and in particular when precision and recall are used, it turned out that simple techniques in most cases outperform recent neural approaches." They further conclude, "progress seems to be still limited… despite the increasing computational complexity of the models."

#### 5.3.4 | The value of data relevance to matching consumer preferences

In contrast to the limited improvements identified in prediction resulting from increased methodological sophistication, some recent work on recommendation engines suggests the possibility of material improvements from incorporating more relevant data (besides past choice or ratings data) for specific recommendation tasks. For example, recent research has shown promise in improving recommendations by incorporating user and product demographic data when recommending products (i.e., recommending products to users that users with similar demographics have purchased/liked; Zhao et al., 2016) and through incorporating location-specific information when recommending restaurants (Bao et al., 2012).

In sum, the empirical evidence suggests that greater sophistication in methods generally has not resulted in obvious improvements in the predictive performance of recommendation engines. This is consistent with our discussion of the limits to preference prediction arising from the constructive nature of consumer preferences. Namely, we surmised that because choices reflect a combination of stable preferences and many other influences, many of which are unpredictable, past choices can usually only be used to predict future choices in a crude fashion. Conversely, and unsurprisingly, it appears that additional information (beyond past choices/ratings) that is relevant to the particular recommendation task at hand has potential to offer meaningful, albeit still limited, performance improvements.

### 5.4 | Predicting choices based on deep insights into consumers' "psyche"

Many of the most frightening claims that have been made regarding marketers' abilities to predict consumer preferences, such as those quoted in the introduction to this article, go well beyond asserting that marketers might be able to target ads or recommend products with precision. They claim that marketers can model our preferences and desires based on gaining deep insights into our psychology from the digital traces we leave behind. For example, Matz and Netzer (2017) describe how a facial expression can be used to infer the person's emotional state, that, in turn, would influence the color scheme of a website (presumably in a way to influence purchase). Likewise, they describe how a person's personality (as inferred through their behavior) could be used to recommend tourist attractions in a new city. In other words, the claims are that marketers will be able to precisely predict and even manipulate our behavior, making us "targets for remote control" (Zuboff, 2020).

Popular media has likewise emphasized the supposedly complete information and thereby control that the new techniques provide unsavory marketers and other researchers, as reflected in the recent Netflix films "The Great Hack" and "The Social Dilemma." Clearly, the notion that evil people do evil things and are very good at that resonates and is consistent with the affinity to conspiracy theories. But a closer examination suggests that the claims are greatly exaggerated and are accepted with little scrutiny.

One of the most prominent claims made about the use of AI technologies to influence people, such as in the Netflix film The Great Hack (2019) and elsewhere, is that the UK firm Cambridge Analytica influenced the outcome of the 2016 US Presidential election as well as the UK's Brexit referendum by targeting ads to people on social media based on deep insights into their individual psychology. The Cambridge Analytica story came to light in 2018 when one of the company co-founders told the *Guardian* (Cadwalladr, 2018):

> We exploited Facebook to harvest millions of people's profiles and built models to exploit what we knew about them and target their inner demons. That was the basis the entire company was built on.

Elsewhere, Cambridge Analytica's CEO Alexander Nix stated[2]:

> we don't need to guess at what creative solution may or may not work. We can use hundreds or thousands of individual data points on our target audiences to understand exactly which messages are going to appeal to which audiences.

Lending support to such claims, shortly before the Cambridge Analytica scandal broke, Matz et al. (2017) claimed to have shown experimentally that it was possible "to influence the behavior of large groups of people by tailoring persuasive appeals to the psychological needs of the target audiences." In particular, through placing ads matched to people's personality profiles to over 3 million users on Facebook, Matz et al. observed an increase in clicks and conversions (i.e., sales) of as much as 50% (for a beauty product) relative to unmatched ads. These findings added credibility to the view that advanced algorithms matched to psychological profiles of the sort employed by Cambridge Analytica could have influenced the US presidential election.

However, there are a number of reasons to be skeptical of the notion that Cambridge Analytica had a significant impact on the 2016 US Presidential election. First, personality measures generally tend to be modest predictors of behavior (Judge et al., 2008; Mischel, 1968; Murphy, 2005). In the realm of voting behavior, for example, personality traits have been found to predict about 5 percent of the variance in individuals' left-right political orientations (Furnham & Fenton-O'Creevy, 2018). Second, personality measures derived from online data, such as through patterns of "likes" or through text analysis, are validated against self-report measures as the gold standard or "ground truth" (Kosinski et al., 2014; Ortigosa et al., 2014) to which they tend to be moderately correlated at best; as a result, to the extent interventions are designed to work for particular personality profiles, personality profiles derived from online data are likely to be less well-matched to the interventions than those derived from surveys.

Third, Matz et al. (2017) obtained click through rates of about 0.3% and conversion (purchase) rates of about 0.01% (390 conversions out of 3.1 million ad impressions). Note that the absolute conversion rate was tiny. Likewise, the absolute increase in conversions through matching ads to personality profiles (about 100 total conversions) was small.[3] Einarsen (2018) has calculated that if Cambridge Analytica had the same conversion rate as Matz et al., they would have been able to convert 600 voters in the 3 states where Trump had the narrowest victory and where the result, if overturned, would have meant victory for Clinton; this would be 100X short of what would have been needed to flip these states.

Finally, a conversion for Cambridge Analytica would have meant changing actual voting behavior, not simply buying a benign beauty product linked to an ad as in Matz et al. (2017). At the time, voters would have been inundated with countless competing political messages diminishing the effect of any particular Facebook ad. In addition, voting behavior is very hard to change. For instance, Kalla and Broockman (2018) conducted a meta-analysis of field experiments and concluded:

> the best estimate of the effects of campaign contact and advertising on Americans' candidates' choices in general elections is zero.

In sum, as illustrated by the case of Cambridge Analytica, the idea that marketers are able to use people's digital traces to gain deep insights into their psychology and thereby to accurately predict and manipulate their behavior as if with a digital "voodoo doll" (Johnson, 2019) is fanciful. To be sure, marketers are able to infer some information about consumers' psychological traits and preferences from online behavior, such as from the text they write (Berger et al., 2020), the images they post (Hartmann et al., 2020; Liu et al., 2020), and the brands they follow or like (Culotta & Cutler, 2016; Hu, Xu, et al., 2019; Kosinski et al., 2013; Schoenmueller et al., 2020), but these are very crude measures. Moreover, for the most part—as in the case of attempting to use personality measures to predict political preferences—these measures are not very relevant for predicting consumer preferences, and therefore don't much increase the absolute accuracy of prediction.

### 5.5 | There is limited evidence that methodological sophistication helps accurately predict social (non-choice) behaviors

So far, the evidence we have reviewed has been related to the use of analytic methods to predict consumer choices. However, other research has examined the predictive performance of analytic methods in predicting other complex human behaviors. Reviewing the evidence from such studies is likely to inform our understanding of the ability of similar methods to predict consumer choices. That is, if sophisticated methods can substantially improve our ability to accurately predict human behavior in some domains, then it would be reasonable to surmise that such methods might similarly improve our ability to accurately predict consumer choices.

Of note, in various contexts, researchers have found that simple models tend to be competitive with machine learning methods in predicting behavior. Jung et al. (2020) found that simple rules predicted a number of behavioral outcomes, such as whether defendants who are released from jail will appear for future court proceedings, nearly as well as machine learning methods. Kizilcec et al. (2020) in a large-scale study involving a quarter million students enrolled in 247 online courses over a period of 2.5 years and using machine learning methods to identify which students would benefit most from which interventions (in terms of influencing them to complete a course) found minimal evidence of advantage to the individualized intervention. In particular, they estimated that the average completion rate using an individualized intervention identified by their methodology was 13.38% versus an average completion rate of 13.08% for an intervention assigned at random.

In a recent mass collaboration (Salganik et al., 2020), 160 teams competed to build predictive models to predict six life outcomes (e.g., children's GPA, being evicted) using data from the Fragile Families and Child Wellbeing Study. The authors concluded, "despite using a rich dataset and applying machine learning methods optimized for prediction, the best predictions were not very accurate and were only slightly better than those from a simple benchmark model. Within each outcome, prediction error was strongly associated with the family being predicted and weakly associated with the technique used to generate the prediction." In a follow-up commentary, Garip (2020) surmised, "the results produced by 160 independent teams using myriad strategies are clearly not an artifact of any one method and suggest that SML [Supervised Machine Learning] tools offer little improvement over standard methods in social science data." Other researchers, attempting to predict recidivism obtained similar results, finding little improvement when comparing a complex machine learning model using 137 input features to a two feature logistic regression (Dressel & Farid, 2018).

In sum, there is limited evidence that complex human behaviors can generally be predicted with much accuracy. Likewise, there is a lack of evidence that more sophisticated methods are dramatically increasing the accuracy with which these behaviors can be predicted. These findings should serve to limit our expectations of the degree to which consumer choices are predictable using current technologies.

## 6 | DISCUSSION

Despite a proliferation of claims regarding marketers' increasing ability to accurately predict and manipulate consumer choices, our review reveals that the key challenges to predicting consumer choices remain largely unchanged. Predictions, to be sure, are sometimes easy, especially for stable, repeat, habitual choices that do not involve any new trade-offs, assessments, introspection on values, social influences, real-time information at the point of decision, and so forth. But for most of the more interesting consumer decisions that are "new," non-habitual, and require new evaluations, new sources of information, and that are susceptible to various effects, predictions remain hard, no matter how sophisticated the methodology is. It appears there are too many moving, unpredictable parts, malleable perceptions and preferences, and idiosyncratic variation in situational and informational influences such that any method that relies on the assumption of stable preferences is likely to yield poor predictions. A high level of methodological sophistication can improve predictive performance in certain contexts, but usually not by much. In fact, as illustrated by the results of the Netflix Prize, the limited gains might not be worth the cost of implementation.

By contrast, we can surmise that data relevant to a particular choice prediction task—more so than methodological sophistication or big data—make a difference in predicting consumer choices. As noted, for example, recent research suggests that adding relevant data beyond consumer choice data—such as location data when making restaurant recommendations—can improve the performance of recommendation engines (e.g. Bao et al., 2012; Zhao et al., 2016). We can likewise surmise that highly relevant data such as political party registration are likely to be a strong predictor of future voting preferences. Conversely, as illustrated by the case of Cambridge Analytica, disparate measures, such as personality features extracted from consumer text or social media likes, are unlikely to be of much value to predicting consumer or voter preferences because they are not directly relevant to the choice being predicted. In other words, obtaining highly relevant variables that capture strong, stable tendencies toward choosing particular options is likely much more important for improving predictive accuracy than collecting vast troves of disparate data and applying sophisticated algorithms to analyze it.

However, although relevant data that capture strong, stable tendencies toward choosing particular options are likely to be useful for prediction, there is a question as to the usefulness of such predictions. When people have strong, stable preferences they are likely to be aware of those preferences and are relatively unlikely to be influenced by the actions of a marketer. That is, in such cases, prediction is less interesting and less useful because the person's behavior will be easy to predict and the person is likely to have sufficient insight into their preference that there will be little ability for marketers to influence the consumer's choice. For example, the fact that a consumer searches for a particular retailer on a search engine is likely highly predictive of her likelihood to browse and buy from that retailer, but the prediction might not be of much value because the consumer is already engaged in seeking out the retailer on her own (see Blake et al., 2015). Likewise, recommendation systems have been criticized for offering "obvious," and therefore not very useful recommendations (Adamapolous, 2013).

Although we have focused here on marketers' ability to predict preferences and choices, our conclusions can also be related to other types of consumer evaluations such as consumer attitudes toward products and brands. Like preferences, research suggests that many consumer attitudes are not stable, but are, to a degree, constructed at the time they are called for (Schwarz, 2006, 2007). As such, measures of attitudes (e.g., ratings) toward products or brands are likely to be noisy measures, and thus noisy predictors of future attitudes and preferences. At the same time, like preferences, some attitudes are particularly strong, stable, and well-defined (Nayakankuppam et al., 2018). However, as with preferences, in such cases, prediction is likely to be less interesting and less useful for marketers because these attitudes will be less susceptible to influence by marketers, at least in the short term.

### 6.1 | Implications for consumers and policymakers

What can consumers and policymakers take away from the limited ability of marketers to predict consumer preferences? The limited ability to predict and even impact individual choices should be somewhat reassuring for consumers and policymakers. At the same time, this reassurance regarding the (in)ability to predict choices should not lead consumers and policymakers to overlook genuine causes for concern about the ways in which marketers and others might manipulate or otherwise harm consumers. Foremost, the increasing ability of marketers to track consumers' behavior means that protection of consumers' privacy remains a concern. Likewise, while there is little evidence that microtargeted messaging is particularly effective in manipulating consumers, the proliferation of misleading information and outright disinformation through both traditional channels and social media channels raises concerns (Aral & Eckles, 2019).

In addition to questions pertaining to privacy and disinformation, the current information environment and new sources of information (e.g., reviews) indicate that consumers and policymakers ought to be educated about new forms of deception, especially those that influence decisions at the time and place of purchase. Such manipulations include, for example, fake reviews (Luca & Zervas, 2016), the provision of incentives for positive reviews (Burtch et al., 2018), hiding negative reviews (Zhuang et al., 2018), and misleadingly presenting paid recommendations as having been generated on the basis of consumer preferences. Thus, the potential for such manipulation, combined with the increasing importance of the consumer information environment to shaping consumer choices, means that consumers and policymakers need to be particularly attentive to attempts to unfairly manipulate the information consumers are exposed to.

### 6.2 | Implications for marketers

What do our conclusions about the difficulty of predicting consumer choices mean for marketers? There is a view that, due to data and technology, marketers are increasingly able to predict consumer choices with accuracy and precision or steer them to whatever choices marketers want them to make. For example, a recent lecture for a Coursera Course on data analytics, suggested that "the future of marketing is business analytics… now that you can measure stuff, marketing really has become a science."[4]

In contrast, the conclusions from our review reinforce the view that marketing remains as much an art as science, whether or not the analyses produce seemingly precise numbers. Marketers, as much as ever, must rely on their creativity, insight and judgment, as well as trial and error, and often some serendipity, to identify and develop truly new products (and messages) that match dormant (or "inherent") consumer preferences (see Simonson, 2008).

Moreover, the outsize influence of information near the point of decision in shaping consumer choices suggests that rather than attempting to measure stable consumer preferences and match their products accordingly, marketers must increasingly focus on how the information environment affects the choice of their products. We might expect, given the proliferation of options consumers now have access to at the tap of a screen, that the importance of being in a consumer's consideration set in advance of a decision to being chosen has diminished, because sets increasingly tend to be constructed closer to the point of decisions. Relatedly, often relied upon measures such as Net Promoter Scores (NPS), which reflect consumer attitudes toward companies or brands, have become much less reliable for predicting consumer choices because consumers are increasingly evaluating individual products on their own merits (Simonson & Rosen, 2014).

As a result of these changes in consumer decision-making brought about by the current consumer information environment, marketers must increasingly focus on having their products considered by consumers when and where it counts. This depends on ensuring that one's products are featured and promoted (e.g., through recommendations) by retail and content distribution partners (e.g., Amazon, Netflix, Yelp) near the point of decision and that product reviews are favorable. The latter—ensuring review favorability—is most consistently achieved through an emphasis on product quality, which thereby becomes increasingly important in the information age (Simonson & Rosen, 2014). Quality, it should be noted, also provides an incentive for retail and content distributors to feature the product because high-quality products are likely to result in increased customer satisfaction, and thereby also increased satisfaction with the distributor that features and recommends the product.

In addition to ensuring that their products are considered when and where it counts, marketers must focus on the just-in-time information and influences that increasingly shape consumers' choices when they are deciding among the options they are considering. This means being cognizant that much of the burden of communication increasingly occurs at the point of decision. As such, on top of creating products that will yield favorable reviews, marketers must carefully customize product imagery and messaging for the platforms and mediums through which consumers encounter their products and clearly articulate and demonstrate their benefits at the point of decision.

In light of the above discussion of the importance of informational influences at the time of decision to consumer choice, an interesting question for marketers and distributors, as well as for consumer researchers, is to understand how the manner in which information is presented to consumers influences their choices. For instance, what are the factors that lead consumers to rely on reviews more versus less? Early research investigating this question suggests that, in some contexts, moderately favorable reviews are perceived as more thoughtful, and therefore are more persuasive, than extremely favorable reviews (Kupor & Tormala, 2018). Other research indicates that reviews written on mobile devices (Grewal & Stephens, 2019), reviews in which the reviewer admits to past purchase mistakes (Reich & Maglio, 2020), and reviews that tell good stories (van Laer et al., 2019) are particularly persuasive. Yet other research suggests that people rely on reviews less for experiences than for material purchases because the former are thought to be more subjective (Dai et al., 2020).

Likewise, although recommendation agents have had much less success in accurately matching consumer preferences than might have been expected when they first emerged about 30 years ago, an interesting question is what will lead to greater consumer receptivity to recommendations? For example, would consumers be more likely to respond to recommendations they believe are customized to them or will they prefer recommendations that they believe are more general? In the domain of politics at least, Hersh and Schaffner (2013) found that voters "rarely prefer targeted pandering to general messages." Other research has found that consumers respond more favorably to user-based frames ("people who liked this product also liked this other product") than to product-based frames ("this product is similar to this other product;" Gai & Klesse, 2019). Related questions include how the number, display, and breadth of recommendations might affect the usefulness of and consumer receptivity to recommendations?

Beyond issues of framing and presentation, some research on recommendation engines, recognizing the problem of offering too obvious recommendations, has considered "beyond accuracy" metrics that are likely to affect consumer receptivity to recommendations, including the degree to which recommendations are perceived as surprising (Zhao & Lee, 2016), novel (Adamopoulous & Tuzhilin, 2015), or arousing of curiosity (Abbas & Niu, 2019). A greater understanding of these and other factors that influence consumer receptivity to recommendations is likely to be of critical importance to marketing products in the current consumer information environment.

## 7 | WHETHER AND WHEN WILL MACHINES TAKE OVER MARKETING?

Up to now, the evidence we reviewed suggests that, contrary to many claims, contemporary methods for predicting consumer choices are generally not very accurate and that there tend to be diminishing returns to increasing methodological sophistication. As such, human creativity, judgment, and insight remain as central to marketing as ever. However, the question arises as to whether machines ever will develop the creativity and insight required to match wits with the most talented human marketers.

Currently, we are still far from a point where machines are able to abstract high-level concepts from data or engage in reasoning and reflection (Dehaene, 2020). Indeed, contemporary machine learning algorithms are only able to learn superficial statistical regularities in data, not to abstract high-level concepts (Jo & Bengio, 2017). Thus, until leaps are made in imbuing algorithms with the ability to abstract and understand higher-level preferences (dispositions, tendencies, goals), to reason, to reflect, to test hypotheses, and to integrate knowledge across domains, we should not expect machines to develop the creativity and judgment of human marketers.

In principle, we might expect machines to eventually reach the capabilities of the human brain, and perhaps even surpass it, in abstracting high-level concepts from data, reflecting, reasoning, hypothesis-testing, and so forth. For example, Kahneman (2017) has stated, "I don't see any reason to set limits on what AI can do." However, even if machines are able to match or surpass the reasoning abilities of humans, it is not clear that machines will be able to fully abstract insights about consumer preferences from data without also experiencing human motivations and emotions themselves. In people, it is thought that "theory of mind," the ability to attribute mental states such as beliefs, intentions, emotions, and goals, both to the self and others, requires introspection and social interaction (Demetriou et al. 2010). That is, to maximally understand, and therefore predict, consumer preferences is likely to require information outside of data on choices and behavior, but also on what it is like to be human.

## ORCID

*David Gal* https://orcid.org/0000-0002-3138-7776

## ENDNOTES

[1] In general, the effect of ads tends to be small and highly variable. A recent meta-analysis shows that display ads work (have a positive effect) in aggregate, but the effects require lots of experiments and/or very large sample sizes to identify (Johnson, Lewis, and Nubbemeyer 2017a). As Garrett Johnson recently wrote us in a personal communication, "Ads are not a form of mind control. Most of the time, they barely work.".

[2] At the 2016 Concordia Annual Summit in New York; https://www.youtube.com/watch?v=n8Dd5aVXLCc&feature=emb_logo

[3] The validity of Matz et al. (2017)'s results have been questioned, with Eckles et al. (2018) arguing that true random assignment was not achieved in their study because of the mechanics of Facebook's ad platform (see also Matz et al.'s 2018 reply).

[4] https://www.coursera.org/learn/wharton-customer-analytics

## ACKNOWLEDGMENTS

The authors are grateful to Yuheng Hu, Garrett Johnson, Derek Rucker, and Zak Tormala for comments.

## REFERENCES

Abbas, F., & Niu, X. (2019). One size does not fit all: Modeling users' personal curiosity in recommender systems. *arXiv Preprint arXiv:1907.00119*.

Adamopoulos, P. (2013 October). Beyond rating prediction accuracy: On new perspectives in recommender systems. In *Proceedings of the 7th ACM conference on Recommender systems* (pp. 459–462).

Adamopoulos, P., & Tuzhilin, A. (2015). On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *5*(4), 1–32. https://doi.org/10.1145/2559952

Aral, S., & Eckles, D. (2019). Protecting elections from social media manipulation. *Science*, *365*(6456), 858–861.

Arning, K. (2017). Conjoint measurement. *The International Encyclopedia of Communication Research Methods*, 1–10.

Bao, J., Zheng, Y., & Mokbel, M. F. (2012 November). Location-based and preference-aware recommendation using sparse geo-social networking data. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems* (pp. 199–208).

Bass, F. M. (1993). The future of research in marketing: Marketing science. *Journal of Marketing Research*, *30*(1), 1–6. https://doi.org/10.1177/002224379303000101

Beel, J., Genzmehr, M., Langer, S., Nürnberger, A., & Gipp, B. (2013). A comparative analysis of offline and online evaluations and discussion of research paper recommender system evaluation. In *Proceedings of the International Workshop on Reproducibility and Replication in Recommender Systems Evaluation* (pp. 7–14).

Ben-Akiva, M., McFadden, D., & Train, K. (2019). Foundations of Stated Preference Elicitation: Consumer Behavior and Choice-based Conjoint Analysis. *Foundations and Trends in Econometrics:*, *10*(1–2), 1–144. https://doi.org/10.1561/0800000036

Bengio, Y., LeCun, Y., & Hinton, G. (2015). Deep Learning. *Nature*, *521*(7553), 436–444.

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of Marketing*, *84*(1), 1–25.

Bettman, J. R., & Kakkar, P. (1977). Effects of information presentation format on consumer information acquisition strategies. *Journal of Consumer Research*, *3*(4), 233–240.

Bettman, J. R., Luce, M. F., & Payne, J. W. (1998). Constructive consumer choice processes. *Journal of Consumer Research*, *25*(3), 187–217.

Bettman, J. R., Payne, J., & Johnson, E. J. (1991). Consumer decision making. *Handbook of Consumer Behaviour*, 50–84.

Blake, T., Nosko, C., Tadelis, S. (2015). Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica*, *83*(1), 155–174.

Bridges, J. F. P., Brett Hauber, A., Marshall, D., Lloyd, A., Prosser, L. A., Regier, D. A., Reed Johnson, F., & Mauskopf, J. (2011). Conjoint Analysis Applications in Health—a Checklist: A Report of the Ispor Good Research Practices for Conjoint Analysis Task Force. *Value in Health*, *14*(4), 403–413. https://doi.org/10.1016/j.jval.2010.11.013

Burtch, G., Hong, Y., Bapna, R., & Griskevicius, V. (2018). Stimulating online reviews by combining financial incentives and social norms. *Management Science*, *64*(5), 2065–2082.

Cadwalladr, C. (2018). 'I made Steve Bannon's psychological warfare tool': Meet the data war whistleblower. *The Guardian*, .

Converse, P. D. (1945). The development of the science of marketing—An exploratory survey. *Journal of Marketing*, *10*(1), 14–23.

Culotta, A., & Cutler, J. (2016). Mining brand perceptions from twitter social networks. *Marketing Science*, *35*(3), 343–362.

Dacrema, M. F., Cremonesi, P., & Jannach, D. (2019). September). Are we really making much progress? A worrying analysis of recent neural recommendation approaches. In *Proceedings of the 13th ACM Conference on Recommender Systems* (pp. 101–109)

Dai, H., Chan, C., & Mogilner, C. (2020). People rely Less on consumer reviews for experiential than material purchases. *Journal of Consumer Research*, *46*(6), 1052–1075. https://doi.org/10.1093/jcr/ucz042

De Haan, E., Kannan, P. K., Verhoef, P. C., & Wiesel, T. (2018). Device switching in online purchasing: Examining the strategic contingencies. *Journal of Marketing*, *82*(5), 1–19. https://doi.org/10.1509/jm.17.0113

Dehaene, S. (2020). *How We Learn: Why Brains Learn Better Than Any Machine... for Now.* : Penguin.

Demetriou et al.,Demetriou Demetriou, A., Mouyi, A., & Spanoudis, G. (2010). The development of mental processing. In *The Handbook of Life-Span Development*. Springer

Diaz, J.. (2018). Netflix's recommendations suck–but it's not too late to fix them, *Fast Company* (8/22).

Dressel, J., & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, *4*(1), eaao5580. https://doi.org/10.1126/sciadv.aao5580

Duhigg, C. (2012). How companies learn your secrets. *The New York Times*, *16*(2), 1–16.

Dzyabura, D., & Yoganarasimhan, H. (2018). *Machine learning and marketing*. Edward Elgar Publishing.

Eckles, D., Gordon, B. R., & Johnson, G. A. (2018). Field studies of psychologically targeted ads face threats to internal validity. *Proceedings* of the National Academy of Sciences, *115*(23), E5254–E5255. https://doi.org/10.1073/pnas.1805363115

Einarsen, M. S. (2018). https://twitter.com/matseinarsen/status/976137451025698821

Felfernig, A., Jeran, M., Ninaus, G., Reinfrank, F., Reiterer, S., & Stettinger, M. (2014). Basic approaches in recommendation systems. *Recommendation Systems in Software Engineering* (pp. 15–37). Springer.

Fong, N. M., Fang, Z., & Luo, X. (2015). Geo-conquesting: Competitive locational targeting of mobile promotions. *Journal of Marketing Research*, *52*(5), 726–735. https://doi.org/10.1509/jmr.14.0229

Furnham, A., & Fenton-O'Creevy, M. (2018). Personality and political orientation. *Personality and Individual Differences*, *129*, 88–91. https://doi.org/10.1016/j.paid.2018.03.020

Gai, P. J., & Klesse, A. K. (2019). Making recommendations more effective through framings: Impacts of user-versus item-based framings on recommendation click-throughs. *Journal of Marketing*, *83*(6), 61–75. https://doi.org/10.1177/0022242919873901

Garip, F. (2020). What failure to predict life outcomes can teach us. *Proceedings of the National Academy of Sciences*, *117*(15), 8234–8235. https://doi.org/10.1073/pnas.2003390117

Ghose, A., Li, B., & Liu, S. (2019). Mobile targeting using customer trajectory patterns. *Management Science*, *65*(11), 5027–5049. https://doi.org/10.1287/mnsc.2018.3188

Green, P. E. (1984). Hybrid models for conjoint analysis: An expository review. *Journal of Marketing Research*, *21*(2), 155–169. https://doi.org/10.1177/002224378402100204

Green, P. E., & Rao, V. R. (1971). Conjoint measurement-for quantifying judgmental data. *Journal of Marketing Research*, *8*(3), 355–363.

Green, P. E., & Srinivasan, V. (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research*, *5*(2), 103–123. https://doi.org/10.1086/208721

Grewal, L., & Stephen, A. T. (2019). In mobile we trust: The effects of mobile versus nonmobile reviews on consumer purchase intentions. *Journal of Marketing Research*, *56*(5), 791–808. https://doi.org/10.1177/0022243719834514

Hagerty, M. R. (1986). The cost of simplifying preference models. *Marketing Science*, *5*(4), 298–319. https://doi.org/10.1287/mksc.5.4.298

Hainmueller, J., Hangartner, D., & Yamamoto, T. (2015). Validating vignette and conjoint survey experiments against real-world behavior. *Proceedings of the National Academy of Sciences*, *112*(8), 2395–2400. https://doi.org/10.1073/pnas.1416587112

Hainmueller, J., Hopkins, D. J., & Yamamoto, T. (2014). Causal inference in conjoint analysis: Understanding multidimensional choices via stated preference experiments. *Political Analysis*, *22*(1), 1–30. https://doi.org/10.1093/pan/mpt024

Hartmann, J., Heitmann, M., Schamp, C., & Netzer, O. (2020). "The power of brand selfies in consumer-generated brand images". Available at SSRN.

Hasan, Z.. (2019). Review: 'The Great Hack' will give you a sense of encroaching dread, *San Francisco Chronicle*, July 23.

Hensel-Börner, S., & Sattler, H. (2000). Validity of customized and adaptive hybrid conjoint analysis. In *Classification and Information Processing at the Turn of the Millennium* (pp. 320–329). Springer.

Hersh, E. D., & Schaffner, B. F. (2013). Targeted campaign appeals and the value of ambiguity. *The Journal of Politics*, *75*(2), 520–534. https://doi.org/10.1017/S0022381613000182

Hu, Y., Tafti, A., & Gal, D. (2019). Read this, please? The role of politeness in customer service engagement on social media. In *Proceedings of HICCS*.

Hu, Y., Xu, A., Hong, Y., Gal, D., Sinha, V., & Akkiraju, R. (2019). Generating business intelligence through social media analytics: Measuring brand personality with consumer-, employee-, and firm-generated content. *Journal of Management Information Systems*, *36*(3), 893–930. https://doi.org/10.1080/07421222.2019.1628908

Huber, G. P., Daneshgar, R., & Ford, D. L. (1971). An empirical comparison of five utility models for predicting job preferences. *Organizational Behavior and Human Performance*, *6*(3), 267–282. https://doi.org/10.1016/0030-5073(71)90017-1

Itsubo, N., Sakagami, M., Washida, T., Kokubu, K., & Inaba, A. (2004). Weighting across safeguard subjects for LCIA through the application of conjoint analysis. *The International Journal of Life Cycle Assessment*, *9*(3), 196–205. https://doi.org/10.1007/BF02994194

Jo, J., & Bengio, Y. (2017). Measuring the tendency of CNNs to learn surface statistical regularities. *arXiv Preprint arXiv:1711.11561*.

Johnson, E.. (2019). Cambridge Analytica made "ethical mistakes" because it was too focused on regulation, former COO says, *Recode*, July 31.

Johnson, G. (2020). Inferno: A guide to field experiments in online display advertising. *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.3581396

Johnson, G., Lewis, R. A., & Nubbemeyer, E. (2017a). The online display ad effectiveness funnel & carryover: Lessons from 432 field experiments. Available at SSRN 2701578.

Johnson, G. A., Lewis, R. A., & Nubbemeyer, E. I. (2017b). Ghost ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research*, *54*(6), 867–884. https://doi.org/10.1509/jmr.15.0297

Johnson, G. A., Lewis, R. A., & Reiley, D. H. (2017). When less is more: Data and power in advertising experiments. *Marketing Science*, *36*(1), 43–53. https://doi.org/10.1287/mksc.2016.0998

Judge, T. A., Klinger, R., Simon, L. S., & Yang, I. W. F. (2008). The contributions of personality to organizational behavior and psychology: Findings, criticisms, and future research directions. *Social and Personality Psychology Compass*, *2*(5), 1982–2000. https://doi.org/10.1111/j.1751-9004.2008.00136.x

Jung, J., Concannon, C., Shroff, R., Goel, S., & Goldstein, D. G. Simple rules to guide expert classifications. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *183*(3), 771–800. https://doi.org/10.1111/rssa.12576

Kahneman, D. (2017). "Will AIs replace human decision-making?" *NBER Economics of AI Workshop*. Toronto, Canada. September 14.

Kalla, J. L., & Broockman, D. E. (2018). The minimal persuasive effects of campaign contact in general elections: Evidence from 49 field experiments. *American Political Science Review*, *112*(1), 148–166. https://doi.org/10.1017/S0003055417000363

Kamakura, W., & Ozer, M. (2000). *A multi-trait multi-method validity test of partworth estimates. In Conjoint Measurement* (pp. 225–251). : Springer.

Kizilcec, R. F., Reich, J., Yeomans, M., Dann, C., Brunskill, E., Lopez, G., & Tingley, D. (2020). Scaling up behavioral science interventions in online education. *Proceedings of the National Academy of Sciences*, *117*(26), 14900–14905. https://doi.org/10.1073/pnas.1921417117

Kosinski, M., Bachrach, Y., Kohli, P., Stillwell, D., & Graepel, T. (2014). Manifestations of user personality in website choice and behaviour on online social networks. *Machine Learning*, *95*(3), 357–380. https://doi.org/10.1007/s10994-013-5415-y

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802–5805. https://doi.org/10.1073/pnas.1218772110

Kupor, D., & Tormala, Z. (2018). When moderation fosters persuasion: The persuasive power of deviatory reviews. *Journal of Consumer Research*, *45*(3), 490–510. https://doi.org/10.1093/jcr/ucy021

Kuzmanovic, M., & Martic, M. (2012). An approach to competitive product line design using conjoint data. *Expert Systems with Applications*, *39*(8), 7262–7269. https://doi.org/10.1016/j.eswa.2012.01.097

Lewis, R. A., & Rao, J. M. (2015). The unfavorable economics of measuring the returns to advertising. *The Quarterly Journal of Economics*, *130*(4), 1941–1973.

Lichtenstein, S., & Slovic, P. (Eds.) (2006). *The construction of preference*. Cambridge University Press.

Liu, L., Dzyabura, D., & Mizik, N. (2020). Visual listening. In: Extracting brand image portrayed on social media. *Marketing Science* (forthcoming).

Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science*, *62*(12), 3412–3427.

Ludewig, M., Mauro, N., Latifi, S., & Jannach, D. (2019 September). Performance comparison of neural and non-neural approaches to session-based recommendation. In *Proceedings of the 13th ACM Conference on Recommender Systems* (pp. 462–466). .

Malhotra, N. (2010). *Marketing Research: An Applied Orientation*, chapter 8. : Pearson.

Masnick, M. (2012). Why Netflix never implemented the algorithm that won the Netflix $1 million challenge. *TechDirt* (April 13).

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proceedings of the National Academy of Sciences*, *114*(48), 12714–12719.

Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2018). Reply to Eckles et al: Facebook's optimization algorithms are highly unlikely to explain the effects of psychological targeting. *Proceedings of the National Academy of Sciences*, *115*(23), E5256–E5257.

Matz, S. C., & Netzer, O. (2017). Using big data as a window into consumers' psychology. *Current Opinion in Behavioral Sciences*, *18*, 7–12.

Mischel, W. (1968). *Personality and Assessment*. Lawrence Erlbaum Associates.

Moise, N., Wood, D., Cheung, Y. K. K., Duan, N., Onge, T. S., Duer-Hefele, J., Pu, T., Davidson, K. W., Kronish, I. M., & Personalized Trial Collaboratory. (2018). Patient preferences for personalized (N-of-1) trials: A conjoint analysis. *Journal of Clinical Epidemiology*, *102*, 12–22.

Moore, W. (2004). A Cross-validity Comparison of Rating-based and Choice-based Conjoint Analysis Models. *International Journal of Research in Marketing.*, *21*, 299–312.

Murphy, K. R. (2005). Why don't measures of broad dimensions of personality perform better as predictors of job performance? *Human Performance*, *18*(4), 343–357. https://doi.org/10.1207/s15327043hup1804_2

Narayanan, A. (2018). *The web tracking arms race: Past, present, and future*. Usenix Association. Retrieved from https://www.usenix.org/node/208180

Narayanan, 2019Narayanan, A. (2019). How to recognize AI snake oil. *Princeton University, Department of Computer Science*.

Natter, M., & Feurstein, M. (2002). Real world performance of choice-based conjoint models. *European Journal of Operational Research*, *137*(2), 448–458. https://doi.org/10.1016/S0377-2217(01)00147-3

Nayakankuppam, D., Priester, J. R., Kwon, J. H., Donovan, L. A. N., & Petty, R. E. (2018). Construction and retrieval of evaluative judgments: The attitude strength moderation model. *Journal of Experimental Social Psychology*, *76*, 54–66. https://doi.org/10.1016/j.jesp.2017.12.005

Neumann, N., Tucker, C. E., & Whitfield, T. (2019). Frontiers: How effective is third-party consumer profiling? *Evidence from Field Studies. Marketing Science*, *38*(6), 918–926.

Ortigosa, A., Carro, R. M., & Quiroga, J. I. (2014). Predicting user personality by mining social interactions in Facebook. *Journal of Computer and System Sciences*, *80*(1), 57–71. https://doi.org/10.1016/j.jcss.2013.03.008

Pujol, E., Hohlfeld, O., & Feldmann, A. (2015 October). Annoyed users: Ads and ad-block usage in the wild. In *Proceedings of the 2015 Internet Measurement Conference* (pp. 93–106).

Rafieian, O., & Yoganarasimhan, H. (2020). Targeting and privacy in mobile advertising. Available at SSRN 3163806.

Reich, T., & Maglio, S. J. (2020). Featuring mistakes: The persuasive impact of purchase mistakes in online reviews. *Journal of Marketing*, *84*(1), 52–65. https://doi.org/10.1177/0022242919882428

Robinson, G. B. (1999). *U.S. Patent No. 5,918,014*. : U.S. Patent and Trademark Office.

Sahni, N. S., Narayanan, S., & Kalyanam, K. (2019). An experimental investigation of the effects of retargeted advertising: The role of frequency and timing. *Journal of Marketing Research*, *56*(3), 401–418. https://doi.org/10.1177/0022243718813987

Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., Altschul, D. M., Brand, J. E., Carnegie, N. B., Compton, R. J., Datta, D., Davidson, T., Filippova, A., Gilroy, C., Goode, B. J., Jahani, E., Kashyap, R., Kirchner, A., McKay, S., … McLanahan, S. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences*, *117*(15), 8398–8403. https://doi.org/10.1073/pnas.1915006117

Schmidt, J., & Bijmolt, T. H. (2020). Accurately measuring willingness to pay for consumer goods: A meta-analysis of the hypothetical bias. *Journal of the Academy of Marketing Science*, *48*(3), 499–518. https://doi.org/10.1007/s11747-019-00666-6

Schoenmueller, V., Netzer, O., & Stahl, F. (2020a). "Polarized America: From political partisanship to preference partisanship. *SSRN Electronic Journal*, https://doi.org/10.2139/ssrn.3471477

Schwarz, N. (2006). Attitude research: Between Ockham's razor and the fundamental attribution error. *Journal of Consumer Research*, *33*(1), 19–21. https://doi.org/10.1086/504124

Schwarz, N. (2007). Attitude construction: Evaluation in context. *Social Cognition*, *25*(5), 638–656. https://doi.org/10.1521/soco.2007.25.5.638

Schweidel, D. A., & Moe, W. W. (2016). Binge watching and advertising. *Journal of Marketing*, *80*(5), 1–19. https://doi.org/10.1509/jm.15.0258

Sidak, J. G., & Skog, J. O. (2015). Using conjoint analysis to apportion patent damages. *Federal Circuit Bar Journal*, *25*, 581.

Simon, H. (2018). Irrationals Verhalten. Interview. *Harvard Business Manager*, *40*(8), 52–54.

Simonson, I. (1989). Choice based on reasons: The case of attraction and compromise effects. *Journal of Consumer Research*, *16*(2), 158–174. https://doi.org/10.1086/209205

Simonson, I. (2005). Determinants of customers' responses to customized offers: Conceptual framework and research propositions. *Journal of Marketing*, *69*(1), 32–45. https://doi.org/10.1509/jmkg.69.1.32.55512

Simonson, I. (2008). Will I like a "medium" pillow? Another look at constructed and inherent preferences. *Journal of Consumer Psychology*, *18*(3), 155–169. https://doi.org/10.1016/j.jcps.2008.04.002

Simonson, I., & Rosen, E. (2014). *Absolute value: What really influences customers in the age of (nearly) perfect information*. Harper Collins.

Srinivasan, V. (1988). A conjunctive-compensatory approach to the self-explication of multiattributed preferences. *Decision Sciences*, *19*(2), 295–305. https://doi.org/10.1111/j.1540-5915.1988.tb00268.x

Srinivasan, V., & Park, C. S. (1997). Surprising robustness of the self-explicated approach to customer preference structure measurement. *Journal of Marketing Research*, *34*(2), 286–291. https://doi.org/10.1177/002224379703400208

Stephens-Davidowitz, S. (2017). *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are*. HarperCollins Publishers.

Terlep, S., & Seetharaman, D. (2016). P&G to scale back targeted Facebook ads. *The Wall Street Journal*, 17.

Tomlin, J., & Zeithammer, R. (2018). *INSIGHT: Product Labeling Class Actions—Identifying the 'Con' in Conjoint Surveys.* : Bureau of National Affairs Inc.

Tversky, A., & Thaler, R. H. (1990). Anomalies: Preference reversals. *Journal of Economic Perspectives*, *4*(2), 201–211. https://doi.org/10.1257/jep.4.2.201

Van Laer, T., Edson Escalas, J., Ludwig, S., & Van Den Hende, E. A. (2019). What happens in Vegas stays on TripAdvisor? A theory and technique to understand narrativity in consumer reviews. *Journal of Consumer Research*, *46*(2), 267–285.

Wright, P., & Kriewall, M. A. (1980). State-of-mind effects on the accuracy with which utility functions predict marketplace choice. *Journal of Marketing Research*, *17*(3), 277–293. https://doi.org/10.1177/002224378001700301

Yaffe-Bellany, D. (2019). Would you like fries with that? McDonald's already knows. *New York Times* (10/23/19)

Zhao, P., & Lee, D. L. (2016). July). How much novelty is relevant? it depends on your curiosity. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval* (pp. 315–324).

Zhao, W. X., Li, S., He, Y., Wang, L., Wen, J. R., & Li, X. (2016). Exploring demographic information in social media for product recommendation. *Knowledge and Information Systems*, *49*(1), 61–89. https://doi.org/10.1007/s10115-015-0897-5

Zhuang, M., Cui, G., & Peng, L. (2018). Manufactured opinions: The effect of manipulating online product reviews. *Journal of Business Research*, *87*, 24–35. https://doi.org/10.1016/j.jbusres.2018.02.016

Zuboff, S. (2020). You are now remotely controlled. *New York Times*.