

How PURCHASE PROBABILITY SCALES Can Shed Light on CONSUMER PURCHASE INTENTIONS

BY RENE BEFURT AND ALVIN J. SILK

Market researchers generally, and survey experts specifically, study consumers to learn about their behavior: What are consumers' opinions, attitudes, thoughts, and actions at the various stages of the buying process? Especially in litigation cases, these and other product purchase-related questions arise in the context of matters related to topics such as intellectual property, trademark infringement, false advertising, and antitrust and competition. Carefully crafted surveys are a powerful tool for assessing consumer behavior, preferences, and purchase intent, especially if real-life observations or purchase data are not readily available.

Assessing Consumer Purchase Intent

In many cases, in contrast to asking consumers what they think of a product, we may want to know whether they are actually willing to put money on the table and make a purchase. A frequently posed question in litigations involving consumer behavior asks whether or not consumers will purchase a good or service in the but-for world, i.e., a world that has never been observed, and which describes what would have happened in a difference circumstance. Such questions explore consumers' "purchase intent" using purchase probability scales, such as the Juster scale discussed later in this article.

Marketing, economics, and public opinion researchers have adapted scaling theories and methods from the fields of psychometrics and statistics. All these disciplines share a mode of observation: collecting and analyzing human responses elicited through interviews and/or self-administered surveys, in which scales are often used. Scales are not novel concepts to

consumers, as we encounter them in everyday life. For example, in many brick and mortar stores, restaurants, and other service venues, consumers have the opportunity to rate their experience on a scale from one to five. Some scales we may use regularly, and almost unbeknownst to us—we provide ratings for ride services, online shopping experiences, or shows broadcasted through online streaming services.

Scales are also frequently used in market research. For example, surveys often rely on purchase intent scales to determine whether or not consumers are inclined to buy a product or service. While it may be tempting to ask directly whether a respondent expects to buy a product during the next six months, research has shown that these types of direct questions are limited in their reliability.¹ Instead, survey experts have developed methods to elicit a *probability-based measure of purchase intent*, or "purchase probability scales."

Purchase probability scales can be used to answer a variety of question types. Does a certain advertising message make consumers more likely to purchase the product? How does the presence of a particular trademark affect consumers' choices? How would consumers react to potentially different prices or to product or service options after a merger of two companies? Do consumers consider fewer products, and are their choices ultimately affected?

In an ideal world, we would be able to answer these questions by following consumers in real life as they make decisions and purchases, or track their purchases across various databases. However, these data are not always readily available, or are impossible to track in the case of products that are described in a but-for world. In such situations, purchase probability scales in carefully crafted surveys are a proven method for assessing purchase intent.

Criteria	Nature of Error	Typical Statistical Indicators of the Quality of Measurement Scales
Reliability	Random Measurement Error	Association between scores when method is replicated under maximally similar conditions.
Construct Validity	Systematic Measurement Error	Association between different measures of same construct as compared to agreement between different construct measured in same way.
Predictive Validity	Change in True Intention	Association between intentions measured at time t and purchase behavior occurring in the interval $t + d$.

Table 1

What Are Purchase Probability Scales?

In a basic form, purchase intent questions attempt to measure whether consumers intend to acquire a good or service by asking directly about “intent,” “expectations,” or “plans.” However, respondents may not have 100 percent certainty about their intentions, expectations, or plans; alternatively, respondents might have no plans to buy a certain product, but can still acknowledge that there is some probability that given future circumstances they could buy that product. As such, while categorizations of purchase intent are appropriate measures in some situations, numerous studies have demonstrated that the *purchase probability* framing can achieve greater precision and reliability.²

Purchase probability scales are designed to frame purchase intention as the chance or likelihood that a purchase will occur, rather than as only categorizations of purchase intent. For example, if the respondent expresses a 10 percent purchase probability on a slider scale, we can conclude that the consumer is telling us that the chance the consumer will make the purchase comes close to 10 percent (one in 10). Indeed, if all consumers felt this way, and they accurately gauged their purchase intentions, roughly 10 percent of consumers would purchase. This type of question asks consumers to estimate the chance of their own future behavior.³

Purchase probability scales arose as an additional level of sophistication to purchase intent measures, consistently achieving higher correlation with future behavior (i.e., whether consumers actually buy or do not buy a product or service) than simpler categorizations of purchase intent.⁴ Purchase probability scales have been widely utilized in consumer research.⁵ Among purchase probability scales, one has been demonstrated repeatedly as one of the most reliable and accurate predictors of future consumer behavior: the 11-point Juster scale.

The Juster Scale

The Juster scale was developed as a way to measure purchase intent with increased precision and reliability. The scale as proposed and tested by Juster in his 1966 seminal paper is as follows, including numerical, textual, and probabilistic descriptions⁶:

- 10 Certain, practically certain (99 in 100)
- 9 Almost sure (9 in 10)
- 8 Very probable (8 in 10)
- 7 Probable (7 in 10)
- 6 Good possibility (6 in 10)
- 5 Fairly good possibility (5 in 10)
- 4 Fair possibility (4 in 10)
- 3 Some possibility (3 in 10)
- 2 Slight possibility (2 in 10)
- 1 Very slight possibility (1 in 10)
- 0 No chance, almost no chance (1 in 100)

Of note, in designing the text of the scale, Juster first tested a similar 11-point scale with phrasing centered around the midpoint as “About even chance (50-50),” demonstrating that this phrasing yielded an artificial peak at the midpoint. In further testing, Juster’s solution was the above language where each phrase has similar “visibility” to respondents.⁷

Reliability and Validity of Purchase Probability Scales

Generally, when researchers evaluate measures such as purchase probability scales, they focus on three criteria: reliability, construct validity, and predictive validity (see tbl. 1). Since Juster’s seminal paper, the reliability of purchase probability scales has been carefully studied.⁸ It is one of the most used and verified market research scales.

Evaluated repeatedly under the psychometric criteria of reliability, construct validity, and predictive validity, the Juster scale has been found to be effective and reliable in both face-to-face interviews and self-completion surveys, such as Internet survey instruments, and for forecasting time periods ranging from three months to a year.⁹ While the Juster scale may not be a perfect predictor of behavior in each and every situation, it has been effectively applied in numerous scenarios, including durables, fast-moving consumer goods, and services, such as renewal and churning of professional football season tickets¹⁰ and customer defection in personal retail banking.¹¹ Researchers have carefully documented when the Juster scale predicts well and when predictions need adjustment. For example, scenarios which benefit from applying an adjustment factor to Juster scale-based measurements include the case of unfamiliar products, extremely

Rene Befurt is an expert in applying marketing research methodologies to both litigation and strategy case work. He has experience with matters related to false advertising, consumer packaged goods, and communications technology, among others. He can be reached at rene.befurt@analysisgroup.com. **Alvin J. Silk** is Lincoln Filene Professor of Business Administration Emeritus at the Harvard Business School. His research interests are in the economics of the advertising and marketing services industry, the development and management of advertising campaigns, and decision support systems in marketing. He can be reached at asilk@hbs.edu.

Factor	Coefficient	p-Value	Correlation between Purchase Intent and Subsequent Behavior
Product Type (New vs. Existing)	0.3733	0.0284	Higher for existing products than for new products.
Product Type (Nondurable vs. Durable)	0.4537	0.0494	Higher for durables than for nondurables.
Product Level (Specific Brands or Product Models vs. Category)	-0.9579	0.0039	Higher for specific brands or product models than for broader categories.
Time between Intent and Behavior	0.5552	0.0368	Higher for shorter time horizons than for longer time horizons.

Table 2

long time horizons between stated purchase intentions and actual purchases, or broad product categories that require more precise definitions or reference to specific brands to improve the reliability and predictive validity of purchase intention measurements.¹²

In evaluating the circumstances under which the Juster scale performs best, researchers analyzed data from 40 studies conducted independently, in which data on both purchase intentions and subsequent purchase behavior had been collected. The studies differed widely with respect to nature of the products studied (new vs. existing, durable vs. nondurable, specific brand or model vs. product category) and the length of time interval between when purchase intentions were measured and when purchasing behavior was observed. The researchers reported a regression analysis indicating how the relationship between purchase intentions and purchase behavior (i.e., predictive validity) varied according to the type of purchase behavior and the length of the time horizon.

Table 2 is adapted from the aforementioned meta-analysis of purchase intent studies and summarizes the best-performing combination of characteristics for which purchase intentions are highly correlated with subsequent behavior—in other words, to what extent each specific factor drives this correlation.¹³ The p-values for these factors are below the standard threshold for statistical significance ($p < 0.05$), indicating that each of the factors described is associated with increased correlation between the purchase intent measure and subsequent purchase behavior.

Despite this published, long-term track record of the Juster scale’s reliability and predictive validity, in the recent case *United States v. AT&T, Inc.*, the U.S. District Court for the District of Columbia focused on alleged shortcomings of the application of that scale but made no reference to the body of evidence addressing those concerns and serving to affirm the reliance on the Juster scale.¹⁴ The defendants and ultimately the judge were concerned with (1) consumers’ alleged lack of experience with a blackout on TV that would affect consumers’ purchase intent statements regarding a switch to a competing TV provider that would not suffer from blackouts;¹⁵ and (2) consumers’ not understanding the descriptions on the Juster scale that express their perceived purchase likelihood.¹⁶

Of note, the court did not weigh these concerns against arguments for the reliability of the Juster scale. For example, the court’s first concern pertaining to consumers’ lack of experience with blackouts does not reflect the pervasive messaging from TV providers who blame failed negotiations for an evidently black screen on channels for which a carriage agreement had not been reached.¹⁷ Similarly, the concern that consumers

may not understand the likelihood expressions shown to them by the Juster scale was tied to consumers’ potentially distorted understanding of likelihoods for rare, dangerous events such as car accidents.¹⁸ However, the reliability of consumers’ statements of their purchase likelihoods has been established on numerous occasions. It is misleading to connect them to the difficulty of estimating the likelihood of rare events such as car accidents, house fires, or robberies.¹⁹ For purchase intention probabilities, 60 years of practice and academic research has assured that consumers understand the verbal descriptions and that the verbal descriptions are good indicators of probabilities (albeit applying an adjustment factor when necessary).

In contrast to the rather skeptical and selective consideration of the Juster scale’s application in *AT&T*, courts have relied on purchase intent surveys in intellectual property cases such as *Harolds Stores, Inc. v. Dillard Department Stores, Inc.*²⁰ and *Doctor’s Associates, Inc. v. QIP Holder LLC*.²¹ Because purchase probability scales combine reliability and predictive validity with (relative) ease of implementation, they are poised to remain one of the main tools to measure how advertisement claims, product features, missing disclosures, or (anti-)competitive arrangements affect consumer behavior.

Conclusion

Ultimately, well-implemented purchase probability scales have a proven track record in market research and in academia, and are a well-vetted solution for surveys that seek to examine consumer purchase behavior. ■

Endnotes

1. F. Thomas Juster, *Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design*, 61 J. AM. STAT. ASS’N 658, 663–64 (1966) (“Let us now examine the typical survey question about intentions to buy. The respondent is asked whether he ‘expects’ or ‘plans’ to buy a car during the next six or twelve months, and the interviewer codes the answer into categories such as definitely will buy, probably will buy, don’t know, no, etc. What are we to make of these responses? . . . [T]he most reasonable general interpretation is that plans or intentions to buy are a reflection of the respondent’s estimate of the probability that an item will be purchased within the specified time period.”).

2. Dianne Day et al., *Predicting Purchase Behaviour*, 2 MARKETING BULL. 18 (1991); Donald H. Granbois & John O. Summers, *Primary and Secondary Validity of Consumer Purchase Probabilities*, 1 J. CONSUMER RES. 31 (1975); Juster, *supra* note 1; Manohar

- U. Kalwani & Alvin J. Silk, *On the Reliability and Predictive Validity of Purchase Intention Measures*, 1 *MARKETING SCI.* 243 (1982).
3. Donald H. Granbois & Ronald P. Willet, *An Empirical Test of Probabilistic Intentions and Preference Models for Consumer Durables Purchasing*, in *MARKETING AND THE NEW SCIENCE OF PLANNING* 401 (Robert King ed., 1968); Paul R. Warshaw, *Predicting Purchase and Other Behaviors from General and Contextually Specific Intentions*, 17 *J. MARKETING RES.* 26 (1980); Malcolm Wright et al., *Market Statistics for the Dirichlet Model: Using the Juster Scale to Replace Panel Data*, 19 *INT'L J. RES. MARKETING* 81 (2002).
 4. André Gabor & C.W.J. Granger, *Ownership and Acquisition of Consumer Durables: Report on the Nottingham Consumer Durables Project*, 6 *EUR. J. MARKETING* 234 (1972); J.F. Pickering & B.C. Isherwood, *Purchase Probabilities and Consumer Durable Buying Behaviour*, 16 *J. MKT. RES. SOC'Y* 203 (1974); Malcolm Wright & Murray MacRae, *Bias and Variability in Purchase Intention Scales*, 35 *J. ACAD. MARKETING SCI.* 617 (2007).
 5. Wright & MacRae, *supra* note 4. While the results of purchase probability scales are reliable at the time of respondents answering the questions, future events that were unrealized or unanticipated at survey time could cause some shifts between survey results and subsequent actual purchases. C. Joseph Clawson, *How Useful Are 90-Day Purchase Probabilities?*, 35 *J. MARKETING* 43 (1971); Charles F. Manski, *The Use of Intentions Data to Predict Behavior: A Best-Case Analysis*, 85 *J. AM. STAT. ASS'N* 934 (1990).
 6. Today, web-based questionnaires improve the format, often with a slider scale and an additional visual anchor.
 7. Juster, *supra* note 1.
 8. See, e.g., Kalwani & Silk, *supra* note 2; Donald G. Morrison, *Purchase Intentions and Purchase Behavior*, 32 *J. MARKETING* 65 (1979); Vicki Morwitz, *Consumers' Purchase Intentions and Their Behavior*, 7 *FOUND. & TRENDS MARKETING* 181, 194 (2012) (“[T]he literature to date suggests that purchase intentions measures that ask consumers to assess and report their probability or purchase (e.g., [the Juster scale]), most accurately predict subsequent purchase behavior.”).
 9. Day et al., *supra* note 2.
 10. Heath McDonald et al., *Predicting Which Season Ticket Holders Will Renew and Which Will Not*, 14 *EUR. SPORT MGMT. Q.* 503 (2014).
 11. Ron Garland, *Estimating Customer Defection in Personal Retail Banking*, 20 *INT'L J. BANK MARKETING* 317 (2002).
 12. Vicki G. Morwitz et al., *When Do Purchase Intentions Predict Sales?*, 23 *INT'L J. FORECASTING* 347, 358 (2007).
 13. *Id.*
 14. 310 F. Supp. 3d 161, 234 (D.D.C. 2018) (“All in all, I can’t help but conclude that the internet survey’s methods [using the Juster scale] are unreliable and that its results fly in the face of real-world evidence regarding the effect of programming blackouts.”).
 15. *Id.* (“And even that unsupported correlation ‘basically disappears’ when respondents are asked to predict their behavior with respect to new products or situations—such as a permanent Turner blackout.”).
 16. *Id.* at 233 (“[The] Juster scale had two critical flaws: first, its text descriptions were ‘out of w[h]ack with the numbers,’ and, second, Juster scales are particularly unreliable in quantifying consumer choices of this kind [with respect to new products or situations.]” (alteration in original) (citations omitted)).
 17. See, e.g., Meg James, *Blackout Ends: Tribune Media TV Stations, including KTLA, Return to Charter Spectrum*, *L.A. TIMES* (Jan. 11, 2019), <https://www.latimes.com/business/hollywood/la-fi-ct-tribune-charter-blackout-over-20190111-story.html> (“Spectrum customers were caught in the crossfire. Instead of programming, Charter aired a message that blamed Tribune Media for demanding a ‘ridiculous increase’ in the fees that it charged for the right to rebroadcast its stations’ over-the-air signals.”).
 18. *AT&T*, 310 F. Supp. 3d at 233 (“[The defendant’s expert] put in plain terms: ‘Now if I told you that I thought there was a very slight possibility that I would get into a car accident driving from Washington to Baltimore on the Baltimore Washington Parkway this evening, I don’t think you would say that was one out of every ten times I attempted that. You might say one out of every thousand or more. So the text description is out of whack with the numbers. And that’s true throughout the scale.’”).
 19. Daniel Kahneman & Amos Tversky, *Prospect Theory: An Analysis of Decision under Risk*, 47 *ECONOMETRICA* 263 (1979).
 20. 82 F.3d 1533 (10th Cir. 1996).
 21. No. 3:06-cv-1710 (D. Conn. Feb. 19, 2010).