Marketing Science Institute Working Paper Series 2021

Report No. 21-139

Predicting Sales of New Consumer Packaged Products with fMRI, Behavioral, Survey and Market Data

Marton Varga, Anita Tusche, Paulo Albuquerque, Nadine Gier, Bernd Weber, and Hilke Plassmann

PREDICTING SALES OF NEW CONSUMER PACKAGED PRODUCTS WITH fMRI, BEHAVIORAL, SURVEY, AND MARKET DATA

Marton Varga[1] Anita Tusche[2] Paulo Albuquerque[1*]
Nadine Gier[3] Bernd Weber[4] Hilke Plassmann[1*]

[1] INSEAD, France
[2] Queen's University, Canada
[3] Heinrich Heine University, Dusseldorf, Germany
[4] Rheinische Friedrich-Wilhelms-University of Bonn, Germany
* Corresponding authors

Abstract: This paper studied the relevance of different types of data for a retailer's ability to predict sales of new products before their launch. Our approach combined four information sources: (1) in-house observable market data such as price and promotion level, (2) customer attitudes based on a representative survey, (3) incentivized purchasing decisions, and (4) functional magnetic resonance imaging (fMRI) data from a relatively small sample of individuals collected in a laboratory. We used a large German retailer's weekly sales data to define an estimation data set containing 34 packaged foods and drinks. This estimation data set was used to estimate the parameters of our model. We then used the parameter estimates to predict sales of 17 different products before they were launched. Results indicate that using fMRI data to forecast sales of new products significantly increased forecasting accuracy: It led to a 28.6% better forecast than a naïve model that considered historic sales data only, while the model combining all data led to an improvement of 38.6%. Using our approach, managers can quantify the benefits of collecting different types of data beyond observable market data—including neuroscientific data—to predict the market success of new products.

Acknowledgments: We acknowledge funding from INSEAD's Octapharma Chair awarded to HP and would like to thank Peter Kenning for developing a pilot study and sharing its data for this project.

INTRODUCTION

One of the key decisions of marketing managers is whether to launch a new product (Beard & Easingwood, 1996; Biyalogorsky, Boulding, & Staelin, 2006) to maintain or grow market share or to conquer new markets (Hultink et al., 2000). The launch of a new product depends on significant amounts of resources allocated to that initiative before and during the launch (Bhaskaran & Krishnan, 2009). It involves a broad range of firm decisions, such as promotions and pricing on the marketing side, and capacity planning, production, and inventory scheduling on the supply chain side (Cooper, 1979; Petersen, Handfield, & Ragatz, 2005). For example, in 2006 the French company Danone spent over €10 million introducing the new yogurt brand Essensis, which later failed and was removed from the market (Bruno & Plassmann, 2014). About 40% of new products fail at launch, even after extensive evaluation, and only one out of 10 innovations achieves commercial success (Cooper, 2011; Cooper, Edgett, & Kleinschmidt, 2004). Thus, correctly predicting the success of new products is crucial and of great interest to firms (Cooper, 1979; Cooper & Kleinschmidt, 1987, 1995; Rothwell, 1974; Rothwell et al., 1974; Ryans, 1988).

Given this importance, there is a continuous search for new methods and information sources that can improve the accuracy of forecasts of commercial success (Kahn & Chase, 2018). Compelling work in consumer neuroscience has shown that neuroscientific data, such as functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) variables, can predict market-level outcomes such as music sales (Berns & Moore, 2012), movie box office sales (Boksem & Smidts, 2015), and advertising elasticities (Venkatraman et al., 2015). These findings suggest that neuroscientific data from a few participants might outperform traditional marketing research measures such as attitudes and preferences (Knutson & Genevsky, 2018).

The main objective of this paper is to investigate the contribution of fMRI data in combination with other data types that marketers typically use to predict sales of new products. With the collaboration of a large German food retailer, we obtained weekly sales and market data (including prices and promotional activities) for 56 different food and beverage items. We also used surveys to collect information about consumer attitudes toward these products from a representative set of customers of the retailer. Finally, we conducted a brain imaging study to collect fMRI data and non-hypothetical, incentive-compatible purchase decisions regarding these products. Here, a small number of customers not representative of the retailer's customer base were exposed to images of the product and price information and indicated their incentivized willingness to buy. This setup allowed us to measure the impact of each of the four data types we collected (market, survey, incentivized purchase behavior, and fMRI data), on its own or in combination with the other data, on the prediction of sales of new products above and beyond our baseline model, which used average weekly historical sales data.

After data cleaning (excluding 5 out of the 56 products), our sample of products was divided into an estimation data set (34 products) and a prediction data set (17 products). We assessed the change in forecast accuracy of our models in terms of the mean average percentage error (MAPE). Using regression models,[1] our results show that using fMRI data led to an improvement of 28.6% in prediction accuracy compared to a naïve model that considered only the average sales of old products to forecast new product performance. When considered in isolation as the only information source, the fMRI variables did better than models that considered the other available data types (i.e., market data, surveys, and incentive-compatible purchase decisions) on their own. When all data were combined, the improvement in prediction accuracy reached 38.6%

[1] We also implemented random forest models, with substantively similar results. We discuss them in the results section.

compared to the naïve model. Model predictions and additional information about the costs of collecting each data type provide insights into the value of each source of information for the firm. Taken together, our results can assist managers in justifying the acquisition of the different data types to improve forecasts. This is especially important for fMRI data, with which managers are likely to have less experience and thus less knowledge of costs and return on investment.

LITERATURE REVIEW

Our paper is related to two streams of past work: (1) the contribution of brain imaging data to predict consumer choices and (2) the prediction of the performance of new products through the use of different types of data. In what follows, we summarize previous work in these two streams.

Market-level predictions using brain imaging data

One promise of the nascent field of consumer neuroscience has been to improve predictions about what consumers like and thus decide to buy (Plassmann et al., 2015). Being able to more accurately predict whether consumers will buy a product has important marketing applications for new product development (Ariely & Berns, 2010). Table 1 presents an overview.

In a seminal paper, Knutson et al., (2007) developed an fMRI purchasing task in which participants evaluated the desirability of consumer products, considered whether they were worth the price, and decided to buy or not (see supplemental Figure S1). Brain responses obtained in this task improved the prediction of the sample's purchase decisions above and beyond self-reported liking of these products, albeit only marginally. The authors identified three brain regions that were predictive of purchasing decisions: (1) the ventral striatum (vStr), (2) the ventromedial prefrontal cortex (vmPFC), and (3) the anterior insula (aI). Evidence on the ability of these brain regions to predict consumer preferences and choices has been replicated and extended across studies and

various product categories (Genevsky & Knutson, 2015; Tong et al., 2020; Tusche, Bode, & Haynes, 2010). Our selection of brain regions from which to extract fMRI data was based on this evidence, as detailed in the methods section.

These initial empirical findings showcase the consistency of brain regions involved in purchasing decisions on the level of single individuals. More recent papers (summarized in Table 1) have demonstrated the ability of neuroscientific data to predict out-of-sample behavior at the market level—a new method commonly referred to as neuroforecasting (Knutson & Genevsky, 2018). Berns & Moore (2012) provided early evidence in favor of neural predictions of market-level outcomes. They found that brain imaging data from a few music listeners (N = 27) could predict whether a song would become a national hit three years later, as indicated by commercial sales data from Nielsen SoundScan. Data from brain activity in the vStr—obtained using fMRI while subjects listened to music—were successfully used to predict the future sales of those songs, while self-reported liking ratings taken at the fMRI experiment showed no significant correlation with future sales. This study was the first to suggest that brain data from a relatively small sample of individuals could predict commercial sales at the market level better than self-reported liking ratings.

Table 1: Summary of neuroforecasting literature

	Data		Prediction		Measures
	Before launch	After launch	Variables	Evaluate data/models	
Berns & Moore (2012)	survey N=27, fMRI N=27 listening to 120 unknown songs	Nielsen SoundScan for 87 songs	number of albums sold containing that song, three years later	self-reported liking rating and fMRI measures of same participants	moderation analysis
Falk, Berkman, & Lieberman (2012)	survey N=30, fMRI N=30 watching three anti-smoking announcements	call volume to 1-800-QUIT-NOW	ad effectiveness=difference in call volume before and after PSA launch	self-reported liking and ad effectiveness rating and fMRI measures of same participants	weighted Kendall's tau
Dmochowski et al. (2014)		Study 1: EEG ISC, N=16 while seeing scenes from TV show tweet volume and audience size, study 2: EEG ISC N=12 Superbowl ads, survey N=12, liking	time-stamped tweet volume from Crimson Hexagon and Nielsen's audience size while TV show was nationally aired and Facebook-*USA Today* Ad Meter liking of ads	self-reported ad liking and EEG ISC for study 2	explained variance
Boksem & Smidts (2015)		survey N=29, EEG beta and gamma oscillations N=29 while watching trailers of 18 never seen movies, U.S. box office movie sales	U.S. box office movie sales	self-reported liking, ranking, WTP of movies, EEG gamma band	explained variance
Genevsky & Knutson (2015)		survey N=28, fMRI N=28 while doing a microlending task, internet lending rates	internet lending rates for requests from kiva.com	self-reported affect, lending choices, and fMRI measures of same participants	explained variance, AIC, classification accuracy
Venkatraman et al. (2015)		survey N=186, IAT N=80, eye-tracking, heart rate, and SCR N=29, fMRI N=33 while viewing ads and ad elasticity	ad elasticity for 37 ads	IAT N=80, eye-tracking, heart rate, and SCR N=29, fMRI N=33 while viewing ads above and beyond the survey data	explained variance
Falk et al., (2016)	survey N=36 & N=19 from MTurk, fMRI N=47	click-through rate citywide anti-smoking email campaign		self-reported affect, image strength and ad effectiveness rating, and fMRI measures	explained variance
Kühn, Strelow, & Gallinat (2016)	fMRI N=18, seeing six chocolate ads	daily sales of advertised product for six weeks, ad displayed at POS	daily sales for one week in one supermarket that used each ad at the point of sale	compared contribution of different brain regions	discussion of coefficients
Barnett & Cerf (2017)	survey N=122, mobile EEG ISC, N=58 and SCR, 13 movie trailers	U.S. box office movie sales	average weekly movie ticket sales	self-reported WTP, liking and free recall, and ISC EEG measures	discussion of coefficients
Genevsky, Yoon, & Knutson (2017)	survey N=30 +35, fMRI N=30 + 35	market-level crowdfunding outcomes)	crowdfunding decisions on kickstarter.com	self-reported affect, success, funding choices, and fMRI measures of same participants	explained variance, AIC, classification accuracy
Scholz et al. (2017)	survey=41, fMRI N=41 & N=39	online sharing of news article captured by NYTimes API	online sharing via Facebook and Twitter of news article	self-reported intention to share and fMRI measures of same participants	explained variance
Cha et al. (2019)		fNIRS N= 56, average number of daily hits on YouTube		neural measures with fNIRS	discussion of coefficients
Shestyuk et al. (2019)		EEG measures, N=38 while seeing scenes from a TV show, tweet volume and audience size	tweet volume and audience size during first airing of TV show based on Nielsen	correlation between twitter activity and audience size vs. EEG components	explained variance
Motoki et al. (2020)	survey N=40, fMRI N=40		sharing ads in social media	self-report and neural measures	AIC, MSE
Tong et al. (2020)		survey N=36, fMRI N=36, YouTube video view frequency and duration	metadata extracted from internet: aggregate view frequency and aggregate video engagement	choices, self-reported affect ratings and fMRI measures in the same participants	explained variance, AIC, RMSE, classification accuracy

Notes: ISC = Inter-subject correlations, fMRI = functional magnetic resonance imaging, EEG = electroencephalography, SCR = skin conductance response, IAT = implicit association test, fNIRS = functional near-infrared spectroscopy

A related pioneering fMRI paper asked smokers who intended to quit ($N = 30$) about their liking and perceived effectiveness of three different anti-smoking campaigns after their brains were scanned while watching them repeatedly (Falk, Berkman, & Lieberman, 2012). Neural activity in the vmPFC predicted the overall success of the three campaigns, measured in call volume of the advertised quit hotline. Behavioral rankings from the same participants made less accurate predictions.

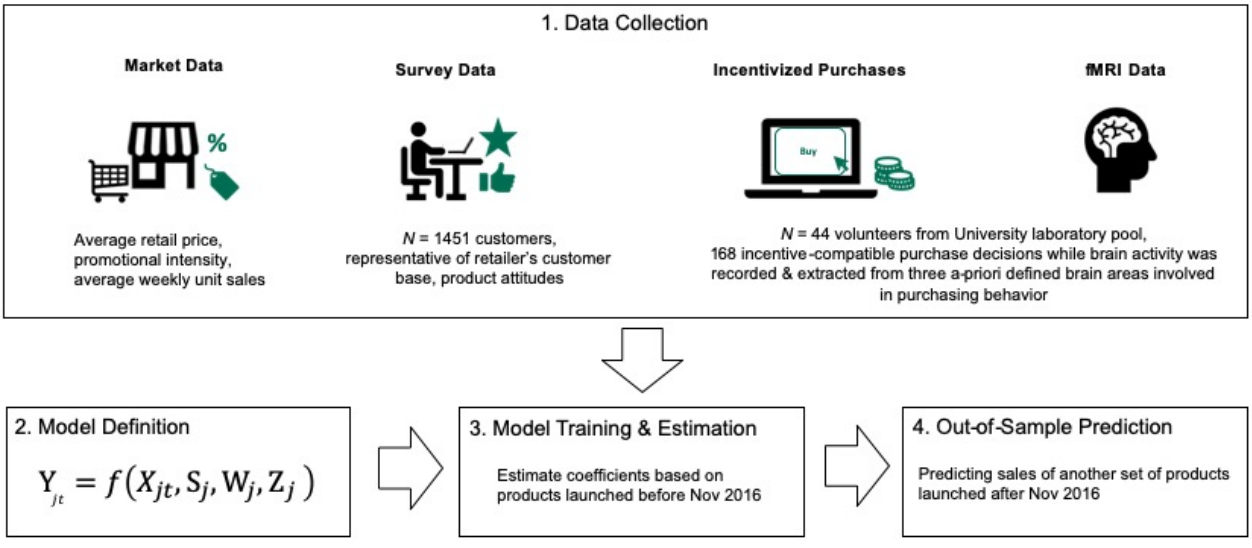
The efficacy of brain data for forecasting market-level outcomes extends beyond fMRI data. For instance, several papers demonstrated that brain activity measured using EEG predicted market-level outcomes such as U.S. box office sales (Barnett & Cerf, 2017; Boksem & Smidts, 2015) and TV audience size (Dmochowski et al., 2014), above and beyond self-reported liking and related preference measures. These studies used a greater variety of methodological approaches and metrics to capture people's brain activity in response to the marketing stimuli, such as different oscillation bands, different components of time-locked EEG signals, and how much participants' brains had the same reaction (using correlations between participants' EEG signals). Thus, less consistency exists regarding the type of EEG signal best suited for which type of neuroforecasting exercise (see Hakim & Levy, 2019) for a review).

The idea that "brain beats behavior" in predicting market-level success has since been conceptually replicated and generalized across product categories—examples include forecasting the success of microloan appeals (Genevsky & Knutson, 2015), advertising elasticities (Venkatraman et al., 2015), movie sales (Boksem & Smidts, 2015), chocolate sales (Kühn, Strelow, & Gallinat, 2016), news article popularity (Scholz et al., 2017), crowdfunding appeal success (Genevsky, Yoon, & Knutson, 2017), and YouTube viewing frequency and duration (Tong

et al., 2020)—and also across different brain imaging techniques (for a review see Knutson & Genevsky, 2018).

All these studies compare different data types from a few individuals in a laboratory environment with their brain imaging data (except Venkatraman et al., 2015). To advance the neuroforecasting literature and demonstrate the value of consumer neuroscience for marketing managers and neuromarketing companies, the comparisons need to include richer data sets that companies typically have access to or acquire to predict sales and success. Against this background, this paper investigates whether the combination of different data types can predict sales of newly introduced food and beverage products. These data sources are (1) market data such as price and promotion level that are accessible for retailers and manufacturers, (2) representative surveys asking customers about their attitudes and intentions, (3) incentivized purchasing decisions, and (4) functional magnetic resonance imaging (fMRI) data from a relatively small sample of individuals collected in a laboratory. Figure 1 gives an overview of the general methodological approach underlying this paper.

Figure 1: Overview Methodological Approach



Combining different data sets to predict the performance of new products

In marketing, work on new product performance began by using data from initial sales of a launched product to predict whether that product was going to be successful in the long run, mostly drawing from repeat-purchase patterns and loyalty rates (Fourt & Woodlock, 1960). Early papers on new product performance prediction reported that sales of fast-moving consumer goods were easier to predict than those of other product categories, due to the repetition of purchase decisions. In parallel, Bass (1969) established that the consumer's initial purchase decision is a function of the number of previous buyers of the product, and since his seminal work, papers using diffusion models to study the success of new products have become commonplace in the marketing literature (e.g., Chandrasekaran & Tellis, 2017; Fan, Che, & Chen, 2017). Given the focus of our research question, we next discuss a subset of the subsequent literature on prediction of performance of new products, concentrating our attention on papers that examined how different types of data can be used or combined to improve the accuracy of predictions of new product sales.

Given that more data—in terms of both quantity and variety—have become increasingly available and at a faster pace, researchers have made efforts to answer the question of how to combine alternative data types and sources in a managerially relevant way. Kahn (2002) suggested that surveys, expert opinions, and average sales of comparable products are the most widespread techniques for predicting demand of new products, highlighting that these methods are popular due to their interpretability. As Armstrong, Green, & Graefe (2015) argued, practitioners should be overly conservative when they do not understand the forecasting procedures. Our aim is to provide a parsimonious method of combining different data, with the intent of investigating which data set or data sets can best improve the prediction of sales of new products.

The objective of combining data is to make use of the advantages of each data type while reducing the disadvantages. Phaneuf, Taylor, & Braden (2013) provided a review of how data on revealed preferences and stated preferences have been combined in marketing, transportation, and environmental economics literature with this purpose in mind. While the main advantage (disadvantage) of revealed preferences data is that it is based on real choices (it is historic in nature), the main advantage (disadvantage) of stated preferences data is that it is flexible in scenario creation (it is hypothetical in nature). Morikawa, Ben-Akiva, & McFadden (2002) also highlighted this and the fact that the two types of data have complementary characteristics and proposed a methodology to use multiple types of data to estimate discrete choice models. The combination of the different data sets allows for a better prediction of scenarios, such as new product introduction (Phaneuf, Taylor, & Braden, 2013), that go beyond the scope of the revealed preferences data, in our case previous sales and price data, and consider possible trends or behavioral perspectives from survey participants.

Several papers have tackled similar research questions. In their seminal paper, Rossi, McCulloch, & Allenby (1996) combined data on past choices, causal variables (such as price, display, and feature), and demographics to better predict individual price and promotional elasticities, which is essential information for targeting marketing activities. The authors showed that previous choices are very informative about consumer preferences. Urban, Weinberg, & Hauser (1996) described how pre-market forecasting can be done for automobiles, using methods with a multimedia virtual-buying environment (an experiment with about 600 participants) to simulate a user experience, combined with tasks where consumers could seek more information about the product, surveys about their purchase intent, and the use of diffusion models and conjoint analysis. The authors quantified the value of each type of data by comparing implementation costs

with benefits regarding the final launch decision of the product. We use a similar approach: the collection and implementation of several studies that allow us to obtain data, which is then used to predict the success of new products.

Feit, Beltramo, & Feinberg (2010) combined different data sets to better predict market shares of products with different levels of attributes. The authors argued that estimates of the importance of product attributes that rely solely on hypothetical choice experiments (for example, conjoint analysis) frequently show inconsistencies that can and should be corrected through the combination of these data with individual-level purchase data. The authors applied a general framework using Bayesian models and individual-level data to the evaluation of attributes in the U.S. minivan market, predicting holdout purchases better than an approach that excluded individual characteristics and motivations.

The data used in some papers goes beyond the traditional revealed and stated preferences data. For example, Mueller et al. (2010), in a two-stage approach, applied an online discrete choice experiment combined with product consumption tasks to understand the interplay between sensory (e.g., taste) and product (e.g., packaging) characteristics to predict liking and repurchase intention of Australian red wines. The study was designed in such a way as to integrate the entire purchase process, from the initial choice through the consumption process and the repurchase decision, with the intent of predicting repurchase decisions. The authors found that data on both types of characteristics are important in explaining repurchase decisions, although the findings in terms of the combination of the data seem to have limited suitability to find the drivers of purchase decisions, in part because wine might be too complex a product for consumers to base their repurchase intention on taste (Mueller et al., 2010). Schneider & Gupta (2016) used both numeric

and textual data from consumer reviews to predict the sales of existing and new products, using a parsimonious linear regression approach, in a similar way as our proposed approach.

Beyond marketing, other fields such as healthcare have also benefited from similar methods. For example, Harris & Keane (1998) studied elderly consumers' choice among health plans using attitudinal data and choice data, showing that the combination of these data sets provided more reliable estimates of their preferences for and perceptions of the attributes of choice alternatives. Kappe, Venkataraman, & Stremersch, (2017) combined historic data on prescriptions and firm detailing efforts with data from subject-matter experts obtained through a conjoint experiment to predict how firms would react to unprecedented detailing changes in the pharmaceutical industry.

To summarize, motivated by these papers we collected data from several information sources, estimated a parsimonious model that allowed us to predict sales and do a hold-out prediction evaluation, and conducted a cost-benefit analysis of each type of data, providing insights to managers regarding which studies might be relevant.

We finish the discussion of the literature on prediction by highlighting that there are alternative methods for prediction and measures to evaluate the accuracy of predictions. In terms of modeling approaches and their applicability to forecasting sales of new products, Hardie, Fader, & Wisniewski (1998) found that simple models provide significantly better forecasts than complex model specifications. Although there have been recent attempts to predict sales of new products with complex approaches (Chong, Han, & Park, 2017; Kulkarni et al., 2012), Lee et al. (2012) showed that the simple logistic regression model is often a better choice than the more complex neural network approaches for forecasting the sales of fresh foods. Hence, and in line with other papers that use neuroscience data, we use linear regressions as the main method.

For the measures used to evaluate prediction, we followed Hardie, Fader, & Wisniewski (1998) and used the mean absolute percentage error (MAPE) as main criterion, defined as

$$MAPE = \frac{\sum_{j=1}^{J}\sum_{t=1}^{T}\left|\left(Y_{jt} - \hat{Y}_{jt}\right)/Y_{jt}\right|}{JT},$$

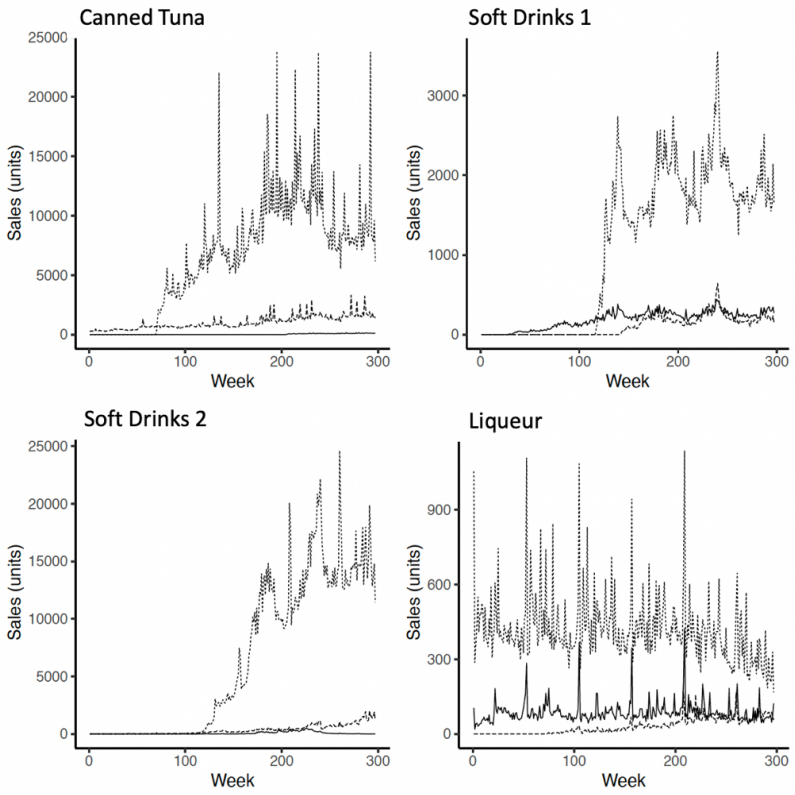
where J is the number of products, T is the number of time periods (weeks), Y_{jt} is the value of actual sales per retailer of product j in week t, and $\hat{Y}_{jt}$ is the respective estimated value. In Hardie, Fader, & Wisniewski (1998), the authors discussed which measure of prediction accuracy is best suited to product sales forecasting tasks and concluded that MAPE is recommended (see also Makridakis, 1993). Divakar, Ratchford, & Shankar (2005) also used MAPE as measure of forecast accuracy in their paper on the practical applications of forecasting models. The authors highlighted that a careful balance between modeling sophistication and practical relevance is key to achieving accurate predictions, with MAPE being one of the easiest measures to understand and interpret. In addition, MAPE has been proven to be very appropriate in planning and budgeting situations (Makridakis, 1993). A number of recent applications have used MAPE, including Prayudani et al. (2019), Jadhav, Chinnappa Reddy, & Gaddi (2017), ArunKumar et al. (2021), Kaewtapee et al. (2021), and Wickramasinghe et al. (2021).

SETTING AND DATA DESCRIPTION

One of Germany's largest food retailers provided us with data on 56 products (23 beverages and 33 food items). Product selection by the retailer's marketing managers ensured representation across 18 product categories (e.g., canned tuna, carbonated soft drinks; on average three products each) and sufficient variation in launch dates.

For each product, we observed the average weekly number of units sold per retailer and the number of retailers that decided to carry each product. For products launched before January 2014, this data covered close to six years (until September 2019). For products launched after January 2014, we observed weekly sales and number of adopted retailers since their launch date. Figure 2 shows the sales evolution of three products in four product categories, as an illustration. It highlights the significant variation in the level of sales, even within each category, suggesting that historical sales data for previously introduced products are limited in their predictive usefulness for the sales of newly launched products. High variance in sales performance characterizes most categories in our data set and partly motivates the retailer's managers to use multiple data sets to predict sales.

Figure 2: Evolution of product sales in four product categories



We divided the products into two sets to implement our analysis approach and evaluate the predictive utility in new product sets. Products launched before November 2016 (35 products) were part of an estimation set, while products launched after November 2016 were the prediction or test set (21 products; see Table S1 in the online appendix for products and launch dates). The threshold date was chosen for practical and data analytical reasons, as the retailer introduced several products soon after this date. Moreover, it yielded estimation and prediction sets of a size consistent with standards for cross-validation and out-of-sample predictions in the field (Berrar, 2019). In a way, adopting a threshold date mimics a manager's challenge to forecast the commercial success of not-yet-launched products, using the information on overall sales of products in the food and drinks categories—and additional data sources at her disposal—at this point.

Besides sales information, four different types of data formed our explanatory variables: (1) market data of all products, including prices and promotional activities; (2) attitudes toward the products obtained using a survey from an online sample representative of the general customer population of the supermarket chain ($N = 1451$); (3) the incentive-compatible purchase decisions of laboratory student participants while their brain responses were measured using fMRI ($N = 44$); and (4) their neural correlates of purchasing the different products, obtained in the same fMRI study. A detailed protocol description is available on the Open Science Framework (OSF).[2] We received ethical approval from the institutional review board of a German university's medical school. The study was conducted in accordance with the Declaration of Helsinki, and all participants gave informed consent for their participation. In what follows, we describe the four data sources in more detail.

[2] https://osf.io/6du3r/?view_only=f130aa7005af42bfa86ea424f2a03069

Market Data

Along with the weekly average sales per retailer, the retailer provided information about the average price of each product for a given week and the weekly frequency with which the product was on promotion across the retailer's stores in Germany. In the estimation model, we also included a dummy variable for food versus beverages (taking drinks as base), to control for the different market size of the two types of items. We refer to this set of variables (price, promotions, food category dummy variable) as market data.[3]

Table 2 presents the summary statistics of the market variables of the 35 products in our estimation data set (launched before November 2016) and the 21 products in the prediction set (launched after that date). The table describes the average price (in euros), average promotional level (in euros), and weekly sales (in thousands units) per retailer. We display the mean and the standard deviations. Products in the estimation set have a higher mean price, a lower mean promotional level, and higher mean sales, due to retailer decisions to support newer products with more aggressive pricing strategies. A typical product costs about €6, is not promoted, and has a sales volume of about four units.

Representative Survey

We recruited 1,451 customers of the supermarket chain using the Qualtrics online panel to be representative of the chain's customer base (see Table S2 in the online appendix for a description). This survey was done in June and July of 2018; the sample size was determined to match sample sizes traditionally used by the retailer when conducting similar surveys.

We note that the survey (and the fMRI experiment—see below) was done at a later stage than the threshold date chosen to define the estimation and prediction sets. We recognize that this

[3] We note that we could have included more dummy variables to capture the differences in product categories within food and drinks. However, with only 35 products in the estimation set, we found that that model overfitted the data.

might be a concern because consumers may have been familiar with the products chosen to predict sales. However, looking at participants' responses about their product purchase before the experiment, only 5% of the respondents indicated previous purchase of the products. Thus, there was very low familiarity with these more recent products (significantly lower than with the products in the estimation set). This is a limitation of our data and not of the approach, driven by the time periods covered by the market data.

Table 2: Summary statistics of market variables by product set

	Products in the estimation set				Products in the prediction set				
	Mean	St. dev.	Min.	Max.	Mean	St. dev.	Min.	Max.	T-values
Price	6.44	11.06	0.87	39.35	3.16	2.83	1.22	13.56	18.20***
Promotional level	0.81	1.50	0	6.53	2.51	2.37	0.07	7.56	9.02***
Sales per retailer	4.23	3.02	1.38	13.56	3.65	2.08	1.49	8.90	5.43***
Number of observations	3,437				2,314				

Note: T-statistics for the equality of means in the last column: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Each participant evaluated 12 products across four different product categories, yielding a questionnaire length of about 30 minutes suitable for an online survey. On average, we obtained 311 evaluations per product, varying between 293 and 331. Participants answered several questions about their attitudes toward the products and their personality and socio-demographic status, and completed an instructed attention manipulation check adapted from Oppenheimer, Meyvis, & Davidenko (2009), which was used to exclude participants who did not pay attention.[4] The order in which the products were shown to participants was randomized.

[4] The questionnaire is available on Open Science Framework (OSF): https://osf.io/6du3r/?view_only=f130aa7005af42bfa86ea424f2a03069. There are more details in the online appendix.

The survey included questions about each product's desirability, measured through four questions about product liking, product attractiveness, packaging attractiveness, and intention to buy the product (translated from a scale by Cho, Lee, & Tharp, 2001). For all of these items, participants' evaluations were based on a 7-point Likert scale ranging from 1 ("fully agree") to 7 ("don't agree at all"). Respondents answered these questions without knowing the product's price. Given the high positive correlations among these measures, we computed an average of these four variables per individual and used this as a composite measure of the *desirability* of the product (Cronbach's alpha = 0.955). Product desirability scores were reversed, so that higher numbers represent more positive product attitudes.

The survey also asked participants about their perception of the product's success by translating and shortening the success scale from Zhang & Schmitt (2001). More specifically, respondents indicated whether they believed that many customers would purchase the product, that it was an enrichment to the category, and that it would have lasting popularity among buyers. They used a 7-point Likert scale ranging from 1 ("fully agree") to 7 ("don't agree at all"). We averaged these three indicators to reflect the *perceived success* of the product in the eyes of survey participants (Cronbach's alpha = 0.868) and reverse coded it. After the product's recommended retail price was revealed to the respondents, they also indicated their hypothetical *purchase intention* if the product was sold at that price.

Table 3 presents the summary statistics related to the measures included in the model. Comparing the products in the estimation and prediction sets, we see that products from the prediction set were perceived to be somewhat more desirable and more successful.

Table 3: Summary statistics of survey variables by product set

	Products in the estimation set				Products in the prediction set				
	Mean	St. dev.	Min.	Max.	Mean	St. dev.	Min.	Max.	T-value
Desirability	4.09	0.43	3.22	4.91	4.55	0.51	3.78	5.41	3.41***
Perceived success	4.37	0.32	3.78	5.04	4.70	0.33	4.05	5.36	3.55***
Purchase intention	2.19	0.50	1.38	3.18	2.36	0.46	1.50	3.10	1.27
# of observations		35				21			

Note: T-statistics for the difference of means in the last column: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

fMRI Experiment

Data from 44 participants were included in the analysis (49.1% female, average age of 27.2 years; see online appendix for more details). This sample size is in line with previous neuroforecasting studies (37% higher than that of the average sample size of the papers reviewed in Table 1) and with current standards in cognitive neuroscience (Yarkoni, 2009). The study was conducted from June to August of 2018.

The participants were asked to make purchase decisions for each of the 56 products at three different price levels, resulting in 168 purchase decisions (56 products x 3 price levels). Using a theory-driven approach, we included average brain activity in the three brain regions previously found to be involved in purchasing decisions (i.e., the neural correlates of product desirability and value (1) the vmPFC, (2) the vStr, and (3) the bilateral aI, see Figure S1). The online appendix describes in detail the fMRI data acquisition, analyses, and detailed definition of brain regions of interest (ROIs).[5]

[5] It is important to note that these three brain regions are also involved in other mental processes unrelated to purchasing; most brain regions are involved in more than one function (Poldrack, 2011). However, given the meta-analysis on value coding (www.sas.upenn.edu/~mcguirej/meta-analysis.html) and emotional intensity coding (www.neurosynth.org) and the sanity check described in the online appendix (Tables S3 and S4 and Figures S2 and S3), we are confident that our three regions of interest are indeed involved in the formation of purchase decisions in our experiment.

As a sanity check, we also tested whether the data from these brain regions are correlated with product desirability during the product consideration phase and with purchase decision and willingness to pay in the product and price consideration phase. Our analysis replicated previous findings that (1) the vStr and aI encoded product desirability while subjects considered products and (2) the vmPFC and aI encoded the subsequent purchase decisions ("strong no" to "strong yes," referred to as decision value) for these items. (See supplemental Figures S3 and S4 and supplemental Tables S2 and S3, and supplemental *sanity check and supplemental whole-brain analyses* section for further details.)

Table 4 presents relevant summary statistics related to these measures. We observed that activity changes in the aI during product consideration were significantly lower for the products in the prediction set as compared to the estimation set.

Table 4: Summary statistics of mean fMRI parameter estimates by product set

	Products in the estimation set				Products in the prediction set				
Brain region and period	Mean	St. dev.	Min.	Max.	Mean	St. dev.	Min.	Max.	T-value
vStr during product consideration	-0.170	0.125	-0.396	0.180	-0.187	0.137	-0.340	0.229	0.46
aI during product consideration	-0.021	0.078	-0.206	0.174	-0.065	0.075	-0.164	0.170	2.11**
aI during product & price consideration	0.433	0.083	0.257	0.653	0.441	0.069	0.291	0.567	0.41
vmPFC during product & price consideration	-0.255	0.125	-0.634	0.080	-0.297	0.095	-0.437	0.020	1.43
Number of observations	35				21				

Note: T-statistics for the equality of means in last column: ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Finally, we observed that the incentivized purchase decisions by participants in the fMRI experiment were not significantly different ($t = 0.794$, $p = 0.431$) for products in the estimation set (M = 2.274, SD = 0.256) and the prediction set (M = 2.329, SD = 0.238).

DATA MODELING AND STATISTICAL ANALYSES

General Approach

Our prediction approach has two stages. In the first stage, we estimated the parameters of our model with information from the estimation set, composed of 35 products launched before November 2016. In the second stage, we used the coefficients of the estimated model to predict the sales of products in the prediction set, which were launched after November 2016. This approach simulates the managerial challenge of predicting the success of newly launched products in the marketplace based on existing data. Our approach also allowed us to compare the predictive utility of the four available data types. More precisely, it enabled quantifying the added value of each data source in terms of model fit and, more important, out-of-sample prediction of commercial success of new products.

We modeled the sales per retailer of product j during week t as a flexible function of the covariates contained in the four different data types,

$$Y_{jt} = f\left(X_{jt}, S_j, Z_j, W_j\right). \quad (1)$$

In equation 1, X_{jt} includes market variables (market price, promotional activities, product type); S_j stands for the variables included in the representative survey (average perceived product desirability, product success, and respondents' intent to purchase at the recommended retail price); Z_j consists of the fMRI data (product-specific parameter estimates during the product consideration phase in the vStr and aI, and during the price and product consideration phase in the vmPFC and aI); and W_j refers to the incentive-compatible purchase decisions during the purchasing task in the fMRI experiment.

To compare the benefit of including different data sources for sales forecasts, we specified models based on different subsets of the data. For example, a model that uses only market data will take the specification

$$Y_{jt} = f_1(X_{jt}), \quad (2)$$

while a model in which we augment the market data with data from the representative survey will be

$$Y_{jt} = f_2(X_{jt}, S_j), \quad (3)$$

and so forth for other combinations of the several data sets. This approach enables us to investigate which combination of data type optimizes the prediction of sales and to quantify the benefit of adding other types of data.

Empirical Specification

In our analysis, we use as a dependent variable the average weekly unit sales of the product per retailer that has decided to sell the product on its shelves (Y_{jt}). Although we could have instead used the overall sales of a product, we decided against it because the volume of sales depends on both consumers' demand and retailers' decision to carry. Given that we do not have information about retailer characteristics or about the decision process retailers go through to adopt a new product, we decided to focus our analysis on explaining and predicting consumer demand, conditional on the retailer offering the product.

Our approach to estimating the parameters is the ordinary least squares (OLS) method, which is simple, widely used, and easy for managers to understand, with a linear form for the *f(.)* function in equation 1. Given that we have a limited number of products, divided into estimation and prediction sets, more complex models that allow for interactions between variables are likely

to increase overfitting, as the degrees of freedom go down with more explanatory variables (Hawkins, 2004).[6]

Data Details for Estimation

To estimate the model, we used weekly data from the 60 weeks before the threshold launch date, from October 2015 to November 2016. To test the model's predictive ability, we predicted sales per retailer for 60 weeks after the launch date, from November 2016 to January 2018. Given that four products were launched after January 2018, our data set dropped to 52 products, for which we have a total of 2,407 observations.

For each product, we excluded the first eight weeks immediately after product launch, as these weeks are typically marked mostly by stocking up and placement decisions by the retailers, leading to more variation in sales not related to the overall performance of the product. This left us with 2,247 observations and 51 products, as one product was launched in November 2016 (so this filter eliminated it from the estimation set).

We also excluded outlier observations, defined as time periods when promotional activity was above the 95th percentile across observations in the estimation set. These are periods when managers likely combined the promotional activity with unobserved-to-the-researcher activities that supported sales. Hence, these extreme cases can influence predictive outcomes, although in our case the results are not substantively different if they are included.

After applying these exclusion criteria, the estimation set consisted of 34 products and 1,600 observations, and the prediction set included 17 products and 505 observations.

[6] We also estimated our model with a random forest approach. This approach benefits from allowing for interactions between the different terms, and so it is more appropriate when several data sets are included in the estimation. In the full model, we found that the OLS approach produced better predictions, in terms of lower MAPE, than the random forest, and hence we decided to present here only the more parsimonious approach.

RESULTS

We first describe the fit and predictive accuracy of the OLS model, using different sets of data. We then discuss the coefficients of the explanatory variables for the subset of best models.

Model Fit and Predictive Accuracy

Following the approach outlined in the previous section, we estimated several models using all possible combinations of the four sets of data and collected as fit measures the adjusted R-square and the in-sample MAPE. To evaluate predictive accuracy, we computed the MAPE for the out-of-sample prediction values, using the estimated OLS coefficients and the data available for products launched after November 2016.

Table 5 shows the estimated and predicted errors, using all possible combinations of data types (smaller numbers—i.e., smaller errors—represent better outcomes). We group the models based on the amount of data available, from models that use a single data set to the full model, which uses the four different types of data. Besides the MAPE, we also compute a measure of how much the fit and predictive accuracy changes, shown as a percentage, when compared to a baseline model in which only the constant is included, defined as $\left(-\left({}^{MAPE_{model}}/_{MAPE_{baseline}}-1\right)\right)$. Our baseline model, a model that represents the manager's best guess on the performance of newly launched products, was based solely on the average of sales per retailer of previously launched products, and assuming no access to any additional data. We observe an in-sample MAPE of 0.72 and an out-of-sample MAPE of 0.84 for this baseline model.

As one might expect, more data is better in terms of the estimated in-sample performance, with the in-sample MAPE improving in all cases with additional data. Looking at the models estimated with a single data set, we observe that market and fMRI data sets provide the best

increments in in-sample accuracy. For instance, adding market data (fMRI data) to a constant-only model reduced the in-sample MAPE to 0.52 (0.62), an improvement of 28.3% (14.6%).

The combination of these two data sets leads to a large reduction in the in-sample MAPE, to a value of 0.50, an improvement of 30.6%. Interestingly, the combination of fMRI and survey data also leads to a similar improvement of the in-sample MAPE. However, it seems that this combination leads to an overfitting of the model, as in out-of-sample, the respective MAPE is worse than in the baseline model. Finally, in terms of in-sample results, when looking at the cases when the three and four types of data are combined, the values again show in-sample improvements above the 30% mark, with the full model using all data reaching the value of 34%.

Looking at the out-of-sample prediction results, and starting with the single-data models, we find that the fMRI data alone performs better than other single-data models—and even better than some of the models that use two or three data sources—with a prediction MAPE of 0.60, an improvement of 28.6% compared to the baseline model. This finding suggests that the fMRI data from a few participants were powerful predictors of the sales of not-yet-launched products at the market level. This finding is even more notable given that participants in the fMRI study were not representative of the retailer's customer base (a student convenience sample).

When combining two data types, the fMRI and market data provide the best improvements in terms of out-of-sample MAPE, with a value of 0.55, an improvement of 34.1%. This result provides evidence that capturing additional information directly related to the product and marketing decisions—in this case the type of product, price, and promotional activities—can complement the data variation captured by the fMRI experiment and account for additional in-store elements that are at least in part under the control of the manager or the retailer, and that are relevant to explain sales of new products.

Table 5: Fit and prediction accuracy

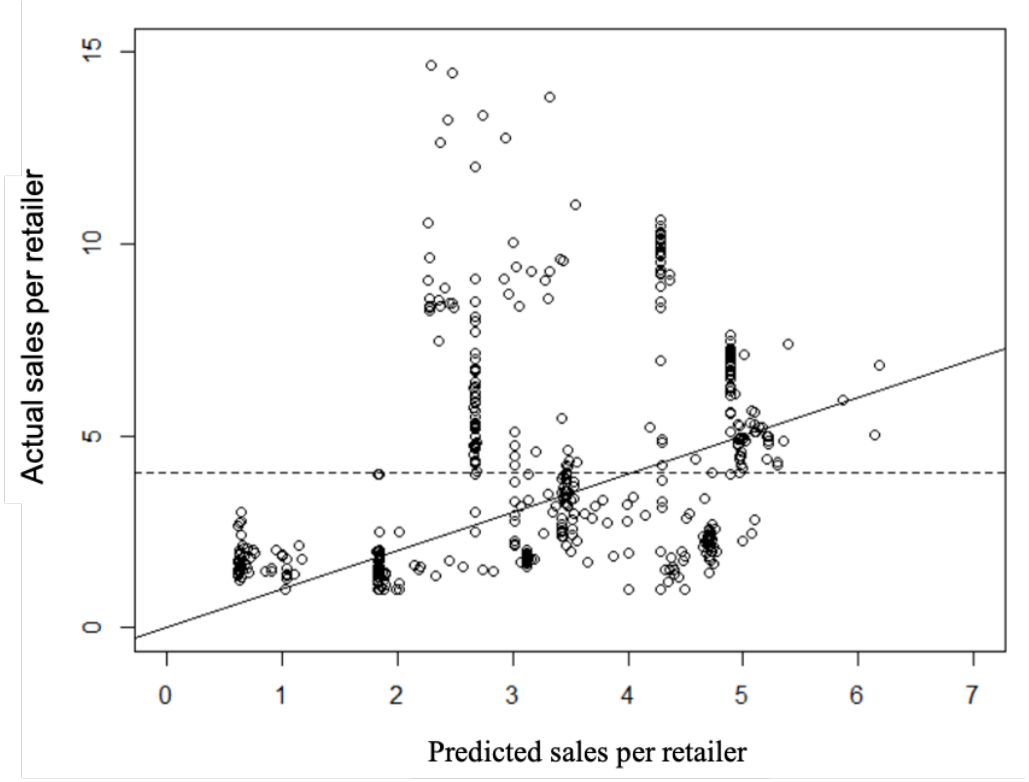
	In-sample prediction error			Out-of-sample prediction error	
	Adj. R-square	MAPE	% improv.	MAPE	% improv.
Baseline (constant only)	0.00	0.72		0.84	
Single data type					
Behavioral data (incentivized purchases)	0.08	0.63	12.7%	0.82	1.7%
Market data variables only (price, promotion, food product)	0.28	0.52	28.3%	0.69	18.0%
Survey data variables (success, desirability, purchase intention)	0.04	0.66	9.4%	0.81	3.3%
fMRI data variables (vStr, aI, vmPFC)	0.11	0.62	14.6%	**0.60**	**28.6%**
Combination of two data types					
Behavioral data + market data	0.28	0.51	29.6%	0.68	18.2%
Behavioral data + survey data	0.16	0.56	23.1%	0.70	16.5%
Behavioral data + fMRI data	0.19	0.59	18.2%	0.59	29.4%
Market data + fMRI data	0.36	0.50	30.6%	**0.55**	**34.1%**
Market data + survey data	0.30	0.54	25.9%	0.71	14.7%
Survey data + fMRI data	0.23	0.50	30.9%	0.85	-2.1%
Combination of three data types					
Behavioral data + market data + fMRI data	0.37	0.48	33.1%	**0.54**	**35.5%**
Behavioral data + market data + survey data	0.33	0.50	30.3%	0.61	27.2%
Behavioral data + survey data + fMRI data	0.28	0.46	36.7%	0.75	10.5%
Market data + survey data + fMRI data	0.37	0.50	31.3%	0.55	33.8%
Combination of four data types (all data)	0.38	0.48	34.1%	**0.51**	**38.6%**

It is important to note that some models overfit the data when additional data is included. This overfitting seems driven mainly by the survey data variables. For instance, adding the survey data variables to a model combining them with one other data type, including the incentivized purchase decisions and fMRI data variables, leads to poorer predictions and increases the MAPE in most cases. One possible explanation is the lack of an incentive when respondents answer the survey in a laboratory and/or their lack of experience with the category. Similar results in which

stated or liking preferences do not match with market outcomes have been found in previous studies (e.g., Kühn, Strelow, & Gallinat, 2016; Phaneuf, Taylor, & Braden 2013).

Overall, combining all data sets maximized the forecast accuracy, leading to a MAPE of 0.51, an improvement of close to 38.6% over the baseline model. For this specification, Figure 3 illustrates the match of predicted and actual sales per retailer. Overall, sales were well captured by this combination of data sources and align better than in the naïve model. The dotted line represents the intercept of that model (i.e., the average sales per retailer of products used in the estimation data set), with an intercept of 4.03 and a standard error of 0.49. The "missed" predictions on the top left center of the figure are all from one product that did extremely (and, according to retailer managers, unexpectedly) well in the market.

Figure 3. Prediction accuracy for newly launched products in the best-performing model (all data types, out-of-sample)



Note: Circles represent weekly sales per retailer values (in 1,000 units) for the products in the out-of-sample set. The solid line represents a perfect forecast. The dashed line shows predictions of the baseline model (with intercept only).

Parameter Estimates

Table 6 presents the coefficients from the OLS models for several combinations of data types, obtained using the estimation data set. These coefficients measure the marginal effect of variables on the dependent variable (sales per retailer in a given week) using solely the data until November 2016 for estimation. The first four columns present the specifications using only one data source for the estimation, with each data set considered separately. In columns 5 and 6, we show the results of the best models (in terms of MAPE prediction) for the combination of two and three data types. The last column shows coefficient estimates using all available data types.

The estimates were consistent overall across models. For the *incentivized purchases* collected during the fMRI task, we observed a significant positive coefficient when the data was used alone. This finding suggests that incentivized purchase decisions observed in a small (nonrepresentative) sample in a well-controlled laboratory context reveal relevant information about product sales on the market level. The significance goes away when other variables are included in the model.

The *market data variables* show effects in line with commonly held notions in the field: The positive coefficients for the promotion level suggest that a more heavily promoted product is likely to have more sales. The significant negative coefficients for average price indicate that a more expensive product is likely to have lower sales. These findings suggest that, to some degree, the store manager has control over the success of new products, using different levels of marketing mix variables.

Table 6: Parameter estimates of selected combinations of data types within the estimation set

Variables	Single data type				Combination of two data types	Combination of three data types	All data
Constant	-3.46	6.49***	10.35	-2.35	-0.47	-3.24	-6.32
	(4.03)	(0.91)	(10.87)	(3.13)	(3.08)	(4.21)	(10.07)
Behavioral data							
Incentivized purchases	3.27*					1.13	2.92
	(1.72)					(1.47)	(2.16)
Market data							
Promotional activities		0.17			0.18	0.18	0.17**
		(0.19)			(0.16)	(0.16)	(0.08)
Market price		-0.12***			-0.13***	-0.13***	-0.12***
		(0.03)			(0.03)	(0.04)	(0.04)
Food product		-3.26***			-2.85***	-2.57***	-2.57**
		(0.95)			(0.86)	(0.84)	(1.26)
Survey data							
Success			-3.05				2.01
			(4.97)				(4.77)
Desirability			0.88				-2.28
			(2.81)				(2.70)
Purchase intention			1.55				0.01
			(0.99)				(1.52)
fMRI data							
vStr at product consideration				-12.86***	-9.54**	-9.58**	-8.20**
				(4.69)	(4.21)	(4.05)	(3.93)
aI at product consideration				12.45*	10.63*	10.50**	8.89
				(6.73)	(5.44)	(5.14)	(5.24)
aI at product & price consideration				9.30**	11.54**	11.80***	11.17**
				(4.58)	(4.63)	(4.50)	(5.40)
vmPFC at product & price consideration				-1.63	-1.70	-1.18	-0.73
				(5.48)	(3.36)	(3.55)	(3.85)
Adj. R-square	0.082	0.280	0.041	0.110	0.368	0.375	0.387
Number of observations	1600						

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Robust standard errors clustered at the product level in parentheses.

The *survey data variables* did not show a significant relationship with sales in the estimation set in our estimation model, suggesting that consumer attitudes, liking responses, and hypothetical purchase intentions are less powerful to predict the sales of new products.

Finally, we observed that the *fMRI data* were a significant predictor of sales, except for estimated brain data in the vmPFC. This finding is in line with recent results that while activity in the vmPFC predicts purchases within the same individual (Chib et al., 2009; Knutson et al., 2007; Litt et al., 2011; Tusche, Bode, & Haynes, 2010), it is likely less suited to predict other people's purchases (Genevsky, Tong, & Knutson, n.d.).

The vStr during product consideration correlated negatively with the product's sales per retailer. This is notable because the vStr has positively correlated with purchases and product desirability within the same individual in previous literature (Knutson et al., 2007) as well as in our data (see Table S3 in the online appendix). It has also consistently shown a positive coefficient in similar in-sample regressions in the neuroforecasting literature that directly investigate not purchasing of packaged goods, but other related behaviors such as sales of songs in the U.S. charts (Berns & Moore, 2012), advertising elasticities (Venkatraman et al., 2015), and promotional sales after ad exposure (Kühn, Strelow, & Gallinat, 2016). However, if we consider vStr activation during the product and price consideration phase instead of the consideration phase, as done in the previous literature (Genevsky & Knutson, 2015; Genevsky, Tong, & Knutson, n.d.; Genevsky, Yoon, & Knutson, 2017; Knutson et al., 2007), we do find that the vStr during this phase is a significantly positive predictor (see Table S5), and including the vStr during this phase in the prediction model does not substantially change our prediction results (see Table S6).

The aI activation has a significantly positive relationship with per-retailer sales, during both the product phase and the product and price consideration phase. This brain region has been

repeatedly linked with consumer choices. Yet evidence regarding the directionality of the effect is mixed (Knutson et al., 2007; Tusche, Bode, & Haynes, 2010). The neuroforecasting literature has generally paid less attention to this brain region. Notable exceptions are studies outside the consumer domain that predict microlending rates (Genevsky & Knutson, 2015) and crowdfunding outcomes (Genevsky, Yoon, & Knutson, 2017), which found a negative coefficient for the aI in their regression analyses. Our results indicate that the aI might play a more important role for predicting sales than previous papers have suggested.

Robustness Checks

We tested whether our findings regarding sales forecasts were robust to alternative model specifications. To this end, we performed four robustness checks. First, we moved the threshold date—which assigned products to the estimation and prediction sets—to four weeks later. Second, we moved the threshold date to four weeks earlier, which led to two more (fewer) products in the prediction (estimation) set. With these two checks, we tested whether our results are robust to the chosen timing for the prediction exercise. Third, we excluded from the estimation set a widely popular and well-known product, which can be considered an outlier in sales per retailer. Fourth, we excluded the two products with the lowest sales per retailer from the estimation set. This approach allowed us to test whether possibly niche items caused a bias in the forecasts and drove our main results.

Table 7 shows the results across the four robustness checks, across all possible combinations of data sources. Across the four sets of robustness checks, the model combining all data continues to provide the best out-of-sample predictions, matching findings obtained using our main specification.

Table 7: Robustness checks for sales forecasts

	Out-of-sample prediction error (MAPE)			
Models with different data sources	Threshold date 4 weeks later	Threshold date 4 weeks earlier	Excluding best-selling product	Excluding two least-selling products
Baseline (constant only)	0.60	0.85	0.81	0.82
Single data type				
Behavioral data (incent.-comp. purchase Q.)	0.61	0.83	0.79	0.81
Market variables only (price, promotion)	0.55	0.69	0.64	0.68
Survey variables (success, desirability, purch. intention)	0.61	0.82	0.86	0.79
fMRI variables (NAC, AI, VMPFC)	**0.37**	**0.60**	**0.58**	**0.59**
Combination of two data types				
Behavioral data + market data	0.55	0.68	0.63	0.68
Behavioral data + survey data	0.52	0.70	0.77	0.69
Behavioral data + fMRI data	**0.36**	0.59	0.57	0.58
Market data + fMRI data	0.40	**0.55**	**0.56**	**0.54**
Market data + survey data	0.54	0.74	0.76	0.70
Survey vars + fMRI data	0.59	0.87	0.95	0.86
Combination of three data types				
Behavioral data + market data + fMRI data	**0.37**	**0.54**	**0.55**	**0.53**
Behavioral data + market data + survey data	0.41	0.63	0.65	0.60
Behavioral data + survey data + fMRI data	0.56	0.76	0.86	0.75
Market data + survey data + fMRI Data	0.41	0.56	0.68	0.55
Combination of four data types (all data)	**0.35**	**0.51**	0.63	**0.51**

We highlight two considerations: First, the MAPE reduces significantly across specifications when the threshold date is moved ahead by four weeks. This is driven mostly by the fact that there are now fewer products in the prediction set, with less variation in sales per retailer, which leads to a more accurate prediction. Second, we note that the MAPE for the full data specification is no longer the best when the best product is removed. This is justified mostly by

the fact that the survey data becomes worse at prediction without that product. In other words, the accuracy of survey respondents is better when that product is considered, most likely because it is a product informative about the popularity of new products. Overall, however, our robustness checks demonstrate that our main findings regarding out-of-sample predictions of sales are robust to a variety of alternative specifications.

Managerial Implications

We evaluated the impact of the different data sets on the profits of a retailer by estimating benefits from using the different data to obtain better predictions, and by obtaining estimates of the costs of acquiring the data. We evaluated the magnitude of the data value using the 17 products that were kept in the prediction set.

To ascertain the costs of acquiring the survey data, we reached out to three suppliers of survey services and were quoted €10,000, €12,000, and €22,000 for surveys of similar sample sizes as the ones used in our study, in terms of both number of products and participants. For the fMRI study, which also includes the incentivized-purchase task, we obtained quotes of €29,000 and €35,000, again with similar conditions to our study. We assume that the market data is free of charge, as the retailer must keep records of prices and promotions, and they know the type of products sold.

To evaluate the benefits of each data set, we computed back-of-envelope values based on the difference in predictions, with and without the different data sets. Our objective was to have estimates of the benefits of having more accurate sales predictions per retailer, which could translate into reductions in stock-outs or in the cost of holding excess stock.[7] For simplicity, we describe the approach when the prediction of sales with additional data reduces the overestimation

[7] Another possibility would be to consider not launching some of the products, the ones for which predicted sales were too low, but given that we do not have data on the launch expenses, we did not attempt this computation.

of sales and assume that the overall value of margins of lost sales (when the prediction is underestimated) is similar. In practice, these values can be different, and a category manager should know both (i.e., the cost of holding excess inventory and the loss in margins from stock-outs).

To ascertain the cost of holding inventory, we obtained information from financial statements of retailers and industry reports: The cost of goods is 75% of the retailer price of the products, similar to other retailers in Germany,[8] while the cost of holding inventory is assumed to be 20% of the value of the inventory.[9] With the information about product retail price, this allowed us to compute the cost of holding inventory, per unit of product, which multiplied by the unit sales per week of each of the products in the prediction set gave us the weekly holding costs. Based on discussions with managers, we assumed that the impact of a different prediction of sales lasts for τ weeks, and after that, the retailer can observe the actual level of sales and correct the inventory levels, no matter what the initial prediction of sales was. We tested $\tau = 4$ and $\tau = 8$.

Given that we obtained the out-of-sample MAPE from our estimations (i.e., the value of the prediction error incurred based on different data sets, measured as a percentage of sales), we could quantify the value of each data set, multiplying the difference in the MAPE of the baseline model and each data MAPE by the weekly sales, valued at the cost of goods. We then multiplied this value – the difference in excess inventory between predictions – by the average weekly holding cost and the number of weeks we assumed to be necessary for the manager to adjust the level of inventory based on actual sales. Hence, the benefit a data set provides to the firm, compared to the naïve model, is given by:

[8] For example, the Rewe Group shows 76.4% of cost of goods sold for 2020 (see https://www.rewe-group-geschaeftsbericht.de/fileadmin/media/pdf/RZF_FinancialStatements_20201231.pdf).
[9] See, for example, https://retailowner.com/Inventory/Costs-of-Excess-Inventory .

$$Value_{Data\ Set} = (MAPE_0 - MAPE_{Data\ Set}) \times Average\ Weekly\ Sales\ (valued\ at\ COGS) \times Weekly\ Holding\ Cost\ \times\ \tau\ weeks\ adjustment. \quad (4)$$

The results for each single-data model and the full model are presented in Table 8, with values aggregated across all 17 products used in the out-of-sample prediction set. The average weekly sales of these products were €188,691, which leads to a holding cost per week of €28,304, given our aforementioned assumptions.

Table 8: Evaluation of each data set

	Difference in MAPE to the baseline model	Reduction in holding cost for 4/ 8 weeks	Benefits of each data set as a % of sales	Benefits of each data set per product	Average cost of acquiring data
Behavioral data	0.02	€2.264/€4.528	0.30%	€133/€266	Part of fMRI cost
Market data	0.15	€16,982/€33,964	2.25%	€998/€1,997	-
Survey data	0.03	€3.396/€6.792	0.45%	€200	€10,000 – €22,000
fMRI data	0.24	€27,171/€54,343	3.60%	€1,598	€29,000 – €35,000
Full model	0.33	€37.360/€74.721	4.95%	€2,198	-

Our estimates show that the fMRI is the most valuable data set to collect, given that it is the one that provides the best increase, on its own, in the out-of-sample MAPE. The overall benefits can range from €27,171 to €54,343, depending on the speed of adjustment to the sales. This justifies its costs of about €30,000 in most situations, considering that in our application, the incentive-compatible purchase decisions were also part of the fMRI and especially because we limited the benefits to only 17 products. The benefits of collecting the data scale linearly with the number of products and with the number of weeks needed to adjust the inventory to market conditions. A full data set that combines all data types would lead to benefits of €37,360 to €74,721, about 5% of sales. Overall, these results highlight the advantage of collecting fMRI data sets to improve the prediction of sales of new products.

Our work also has important implications for neuromarketing vendors. Most of the companies that are offering neuromarketing services are using different EEG-based metrics to tweak ads, making them shorter and thus saving media expenses for their clients (for a list of these companies and their services, see Plassmann & Ling, 2020). Our results indicate that another business opportunity for such companies is to offer neuroforecasting services using approaches such as the one described here to help their clients drive revenue in addition to saving costs.

Conclusion

In this paper, we studied the added value of different data types to the forecasting accuracy of market-level sales of new products. Using data provided by a large German retailer on more recently launched grocery products and on similar products that were previously available on the market, we estimated the contribution of market data (price, promotions, and product type), representative surveys (purchase intention, perceived desirability, and success of the products), fMRI data (in three brain regions involved in purchase decisions: vStr, aI, vmPFC), and incentive-compatible purchase decisions to improving forecast accuracy.

We estimated a regression model and used its estimated coefficients to predict the success of new products. We used the weekly average number of units sold by a retailer as the dependent variable of the regression analysis. Our approach mimics the managerial challenge of obtaining a forecast of not-yet-launched products at a given point in time. We found that using fMRI data to predict the sales of new products significantly increased forecast accuracy. Using only fMRI data, we reduced the prediction error by close to 29% compared to a naïve model (i.e., a model using the average historical sales of previously launched items as an intercept). Such improvement was not possible with any other data type. In addition, we found that although all data types can

improve predictions, some are worth more than others, and given the acquisition cost of data, it is likely that in practice some data is not worth collecting or buying.

We performed various robustness checks, in terms of both data and method of prediction. Through these supplemental tests, we confirmed that brain data of a small number of participants are indeed a robust predictor of sales in the marketplace. In fact, we find that fMRI data are better predictors than traditional customer surveys.

Our findings also contribute to academic research on the predictive utility of fMRI data in a variety of cases and settings (Boksem & Smidts, 2015; Genevsky, Yoon, & Knutson, 2017; Genevsky & Knutson, 2015; Kühn, Strelow, & Gallinat, 2016; Scholz et al., 2017; Venkatraman et al., 2015). We extend prior work in at least three important ways: First, we predicted real-world sales of new products. Second, we integrated product attitudes of a large representative customer sample, market variables, and historical sales data on related products, allowing for a comparison with other information sources that marketers would typically use. And third, we have given managers and researchers an indication of the monetary added value of collecting fMRI data. Taken together, our paper has important novel implications for both marketing research and practice.

References

Ariely D, Berns GS (2010) Neuromarketing: The hope and hype of neuroimaging in business. *Nat. Rev. Neurosci.* 11(4):284–292.

Armstrong JS, Green KC, Graefe A (2015) Golden rule of forecasting: Be conservative. *J. Bus. Res.* 68(8):1717–1731.

ArunKumar KE, Kalaga DV, Kumar CMS, Chilkoor G, Kawaji M, Brenza TM (2021) Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-regressive integrated moving average (ARIMA) and seasonal auto-regressive integrated moving average (SARIMA). *Appl. Soft. Comput.* 103:107161.

Barnett SB, Cerf M (2017) A ticket for your thoughts: Method for predicting content recall and sales using neural similarity of moviegoers. *J. Consum. Res.* 44(1):160–181.

Bass FM (1969) A new product growth for model consumer durables. *Manage. Sci.* 15(5):215–227.

Beard C, Easingwood C (1996) New product launch: Marketing action and launch tactics for high-technology products. *Ind. Market. Manag.* 25(2):87–103.

Berns GS, Moore SE (2012) A neural predictor of cultural popularity. *J. Consum. Psychol.* 22(1):154–160.

Berrar D (2019) Cross-validation. Ranganathan S, Gribskov M, Nakai K, Schönbach C, eds. *Encyclopedia of Bioinformatics and Computational Biology* (Academic Press, Cambridge, MA), 542–545.

Bhaskaran SR, Krishnan V (2009) Effort, revenue, and cost sharing mechanisms for collaborative new product development. *Manage. Sci.* 55(7):1152–1169.

Biyalogorsky E, Boulding W, Staelin R (2006) Stuck in the past: Why managers persist with new product failures. *J. Marketing* 70(2):108–121.

Boksem MA, Smidts A (2015) Brain responses to movie trailers predict individual preferences for movies and their population-wide commercial success. *J. Marketing Res.* 52(4):482–492.

Bruno HA, Plassmann H (2014) Michel et Augustin cookies: Culinary adventurers competing against food industry giants. Case study, INSEAD Publishing, Paris, https://publishing.insead.edu/case/michel-et-augustin.

Cha KC, Suh M, Kwon G, Yang S, Lee EJ (2019) Young consumers' brain responses to pop music on YouTube. *Asia Pac. J. Mark. Logist.* 32(5):1132–1148.

Chandrasekaran D, Tellis GJ (2007) A critical review of marketing research on diffusion of new products. *Rev. Mark. Res.* 3:39–80.

Chib, VS, Rangel A, Shimojo S, O'Doherty JP (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *J. Neurosci.* 29(39):12315–12320.

Cho CH, Lee JG, Tharp M (2001) Different forced-exposure levels to banner advertisements. *J. Advertising Res.* 41(4):45–56.

Chong E, Han C, Park FC (2017) Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Expert Syst. Appl.* 83:187–205.

Cooper RG (1979) The dimensions of industrial new product success and failure. *J. Marketing* 43(3):93–103.

Cooper RG (2011) *Winning at New Products: Creating Value Through Innovation*, 4th ed.

(Basic Books, New York).

Cooper RG, Edgett SJ, Kleinschmidt EJ (2004) Benchmarking best NPD practices—I. *Res. Technol. Manage.* 47(1):31–43.

Cooper RG, Kleinschmidt EJ (1987) New products: What separates winners from losers? *J. Prod. Innovat. Manage.* 4(3):169–184.

Cooper RG, Kleinschmidt EJ (1995) New product performance: Keys to success, profitability & cycle time reduction. *J. Mark. Manag.* 11(4):315–337.

Divakar S, Ratchford BT, Shankar V (2005) Practice prize article—CHAN4CAST: A multichannel, multiregion sales forecasting model and decision support system for consumer packaged goods. *Market. Sci.* 24(3):334–350.

Dmochowski JP, Bezdek MA, Abelson BP, Johnson JS, Schumacher EH, Parra LC (2014) Audience preferences are predicted by temporal reliability of neural processing. *Nat. Commun.* 5(1):1–9.

Falk EB, Berkman ET, Lieberman MD (2012) From neural responses to population behavior: Neural focus group predicts population-level media effects. *Psychol. Sci.* 23(5):439–445.

Falk EB, O'Donnell MB, Tompson S, Gonzalez R, Cin SD, Strecher V, Cummings KM, An L (2016) Functional brain imaging predicts public health campaign success. *Soc. Cogn. Affect. Neur.* 11(2):204–214.

Fan ZP, Che YJ, Chen ZY (2017) Product sales forecasting using online reviews and historical sales data: A method combining the Bass model and sentiment analysis. *J. Bus. Res.* 74:90–100.

Feit EM, Beltramo MA, Feinberg FM (2010) Reality check: Combining choice experiments with market data to estimate the importance of product attributes. *Manage. Sci.* 56(5):785–

800.

Fourt LA, Woodlock JW (1960) Early prediction of market success for new grocery products. *J. Marketing* 25(2):31–38.

Genevsky A, Knutson B (2015) Neural affective mechanisms predict market-level microlending. *Psychol. Sci.* 26(9):1411–1422.

Genevsky A, Tong L, Knutson B (under review) Generalizability of brain activity in forecasting market choice. *J. Consum. Res*.

Genevsky A, Yoon C, Knutson B (2017) When brain beats behavior: Neuroforecasting crowdfunding outcomes. *J. Neurosci.* 37(36):8625–8634.

Hakim A, Levy DJ (2019) A gateway to consumers' minds: Achievements, caveats, and prospects of electroencephalography-based prediction in neuromarketing. *WIRES Cogn. Sci.* 10(2):e1485.

Hardie BGS, Fader PS, Wisniewski M (1998) An empirical comparison of new product trial forecasting models. *J. Forecasting* 17(3-4):209–229.

Harris KM, Keane MP (1998) A model of health plan choice: Inferring preferences and perceptions from a combination of revealed preference and attitudinal data. *J. Econometrics* 89(1-2):131–157.

Hawkins DM (2004) The problem of overfitting. *J. Chem. Inf. Comp. Sci.* 44(1):1–12.

Hultink EJ, Hart S, Robben HSJ, Griffin A (2000) Launch decisions and new product success: An empirical comparison of consumer and industrial products. *J. Prod. Innovat. Manag.* 17(1):5–23.

Jadhav V, Chinnappa Reddy BV, Gaddi GM (2017) Application of ARIMA model for forecasting agricultural prices. *J. Agr. Sci. Tech.* 19:981–992.

Kaewtapee C, Khetchaturat C, Nukreaw R, Krutthai N, Bunchasak C (2021) Artificial neural networks approach for predicting methionine requirement in broiler chickens. *Thai J. Vet. Med.* 51(1):161–168.

Kahn KB (2002) An exploratory investigation of new product forecasting practices. *J. Prod. Innovat. Manag.* 19(2):133–143.

Kahn KB, Chase CW (2018) The state of new-product forecasting. *Foresight (Colch.)* 51:24–31.

Kappe E, Venkataraman S, Stremersch S (2017) Predicting the consequences of marketing policy changes: A new data enrichment method with competitive reactions. *J. Marketing Res.* 54(5):720–736.

Knutson B, Genevsky A (2018) Neuroforecasting aggregate choice. *Curr. Dir. Psychol. Sci.* 27(2):110–115.

Knutson B, Rick S, Wimmer GE, Prelec D, Loewenstein G (2007) Neural predictors of purchases. *Neuron* 53(1):147–156.

Kühn S, Strelow E, Gallinat J (2016) Multiple "buy buttons" in the brain: Forecasting chocolate sales at point-of-sale based on functional brain activation using fMRI. *NeuroImage* 136:122–128.

Kulkarni S, Nikumbh PJ, Anuradha G, Nikam S (2012) S-ANFIS: Sentiment aware adaptive network-based fuzzy inference system for predicting sales performance using blogs/reviews. *Int. J. Multidiscip. Cryptol. Inf. Secur.* 1(2):22–32.

Lee WI, Chen CW, Chen KH, Chen TH, Liu CC (2012) Comparative study on the forecast of fresh food sales using logistic regression, moving average and BPNN methods. *J. Mar. Sci. Technol.* 20(2):142–152.

Litt A, Plassmann H, Shiv B, Rangel A (2011) Dissociating valuation and saliency signals during decision-making. *Cereb. Cortex* 21(1):95–102.

Makridakis S (1993) Accuracy measures: Theoretical and practical concerns. *Int. J. Forecasting* 9(4):527–529.

Morikawa T, Ben-Akiva M, McFadden D (2002) Discrete choice models incorporating revealed preferences and psychometric data. *Adv. Econom.* 16:29–55.

Motoki K, Suzuki S, Kawashima R, Sugiura M (2020) A combination of self-reported data and social-related neural measures forecasts viral marketing success on social media. *J. Interact. Mark.* 52:99–117.

Mueller S, Osidacz P, Francis IL, Lockshin L (2010) Combining discrete choice and informed sensory testing in a two-stage process: Can it predict wine market share? *Food Qual. Prefer.* 21(7):741–754.

Oppenheimer DM, Meyvis T, Davidenko N (2009) Instructional manipulation checks: Detecting satisficing to increase statistical power. *J. Exp. Soc. Psychol.* 45(4):867–872.

Petersen KJ, Handfield RB, Ragatz GL (2005) Supplier integration into new product development: Coordinating product, process and supply chain design. *J. Oper. Manage.* 23(3-4):371–388.

Phaneuf DJ, Taylor LO, Braden JB (2013) Combining revealed and stated preference data to estimate preferences for residential amenities: A GMM approach. *Land Econ.* 89(1):30–52.

Plassmann H, Ling A (2020) The neuromarketing toolkit. Case study, INSEAD Publishing, Paris, https://publishing.insead.edu/case/neuromarketing-toolkit.

Plassmann H, Venkatraman V, Huettel S, Yoon C (2015) Consumer neuroscience: Applications,

challenges, and possible solutions. *J. Marketing Res.* 52(4):427–435.

Poldrack RA (2011) Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding. *Neuron* 72(5):692–697.

Prayudani S, Hizriadi A, Lase YY, Fatmi Y, Al-Khowarizmi (2019) Analysis accuracy of forecasting measurement technique on random k-nearest neighbor (RKNN) using MAPE and MSE. *J. Phys. Conf. Ser.* 1361(1):012089.

Rossi PE, McCulloch RE, Allenby GM (1996) The value of purchase history data in target marketing. *Market. Sci.* 15(4):321–340.

Rothwell R (1974) Factors for success in industrial innovation. *J. Gen. Manage.* 2(2):57–65.

Rothwell R, Freeman C, Horlsey A, Jervis VTP, Robertson AB, Townsend J (1974) SAPPHO updated: Project SAPPHO phase II. *Res. Policy* 3(3):258–291.

Ryans AB (1988) Strategic market entry factors and market share achievement in Japan. *J. Int. Bus. Stud.* 19(3):389–409.

Schneider MJ, Gupta S (2016) Forecasting sales of new and existing products using consumer reviews: A random projections approach. *Int. J. Forecasting* 32(2):243–256.

Scholz C, Baek EC, O'Donnell MB, Kim HS, Cappella JN, Falk EB (2017) A neural model of valuation and information virality. *P. Natl. Acad. Sci. USA* 114(11):2881–2886.

Shestyuk AY, Kasinathan K, Karapoondinott V, Knight RT, Gurumoorthy R (2019) Individual EEG measures of attention, memory, and motivation predict population level TV viewership and Twitter engagement. *PLOS One* 14(3):e0214507.

Tong LC, Acikalin MY, Genevsky A, Shiv B, Knutson B (2020) Brain activity forecasts video engagement in an internet attention market. *P. Natl. Acad. Sci. USA* 117(12):6936–6941.

Tusche A, Bode S, Haynes JD (2010) Neural responses to unattended products predict later

consumer choices. *J. Neurosci.* 30(23):8024–8031.

Urban GL, Weinberg BD, Hauser JR (1996) Premarket forecasting of really-new products. *J. Marketing* 60(1):47–60.

Venkatraman V, Dimoka A, Pavlou PA, Vo K, Hampton W, Bollinger B, Hershfield HE, Ishihara M, Winer RS (2015) Predicting advertising success beyond traditional measures: New insights from neurophysiological methods and market response modeling. *J. Marketing Res.* 52(4):436–452.

Wickramasinghe L, Weliwatta R, Ekanayake P, Jayasinghe J (2021) Modeling the relationship between rice yield and climate variables using statistical and machine learning techniques. *J. Math.* 2021:1–9.

Yarkoni T (2009) Big correlations in little studies: Inflated fMRI correlations reflect low statistical power—commentary on Vul et al. (2009). *Perspect. Psychol. Sci.* 4(3):294–298.

Zhang S, Schmitt BH (2001) Creating local brands in multilingual international markets. *J. Marketing Res.* 38(3):313–325.

Predicting Sales of New Consumer Packaged Products with fMRI, Survey, and Market Data: Supporting Online Appendix

Additional materials and methods: fMRI experiment

Participants. We recruited 53 healthy, right-handed participants using the participant pool of a German university and standard fMRI inclusion criteria. They received a total of €50 for their participation. Of this amount, €40 was used as a budget to spend (or not) on the purchasing task. Participants were told that the goal of the experiment was to study the neural correlates of consumer decision-making. We excluded nine participants from the fMRI sample due to excessive head movement beyond 3 mm/degrees during scanning. Thus, a total of 44 participants were included in the analyses (52.3% female, aged 20–39, M = 27.27 years, SD = 4.86 years; see Table S2 for a sample description).

Procedure. The experiment consisted of three parts. Part one involved a computerized task that took place outside of the fMRI scanner. Participants completed a valuation task to determine their willingness to pay (WTP) for the 56 products using a Becker-deGroot-Marchak (BDM) second price auction mechanism (Becker, DeGroot, & Marschak, 1964). In part two, participants completed an incentive-compatible purchasing task while their brain activation was measured using fMRI. In part three, participants went through the same questionnaire as the participants from the representative survey sample did, except that they evaluated all 56 products in randomized order (outside of the fMRI scanner).

The main fMRI purchasing task was adapted from the SHOP task from Knutson et al. (2007) and displayed using the Scenario Designer software (for the timing and procedure of a sample trial, see Figure S1). In each of the 168 trials, participants were presented with an image of a product (4 sec, product consideration phase), followed by a fixation cross (1-5 seconds), the presentation of the price together with the product (4 seconds, product and price consideration and decision phase), and the response screen (2 seconds, response phase). Inter-trial intervals (fixation

cross) varied from 2 to 6 seconds. Participants were instructed to make their purchase decisions during the product and price consideration phase. They indicated their purchase decisions on a 4-point scale by pressing the respective button on an MRI-compatible response box during the response phase. The mapping of the purchase decision and button press was consistent across participants (strong yes = left index finger, weak yes = left thumb, weak no = right thumb, strong no = right index finger). All participants underwent a training phase to ensure that they understood the meaning of the response buttons.

All 56 products were presented three times across the three runs. Each run included every product once, shown at one of three different price levels. The price levels varied as follows: In every run, one-third of the products were offered for the actual recommended retail price, one-third for a price that was marked up by 20% of the participant's WTP, and one-third for a 20% discount of the participant's individual WTP. This was done to ensure enough variation in the purchasing decision variable since in the original study by Knutson et al. (2007), many nonpurchase trials had to be removed even though the original retail value of the products was discounted by 75%. The mapping of a product to the three price levels was pseudo-randomized across runs. The order of products varied across functional runs. Together, participants made 168 purchase decisions in the fMRI task (56 products x 3 presentations at a different price level each).

At the end of the study, one decision from either the BDM auction or the fMRI purchasing task was implemented by the computer. The total time of the fMRI experiment, including preparation and debriefing time, was 2 hours.

fMRI Data Acquisition. Gradient echo T2*-weighted echo-planar (EPI) images with BOLD contrast were acquired using a 3-Tesla Magnetom Trio scanner (Siemens, Erlangen, Germany) and an eight-channel head coil. Thirty-seven slices were scanned in ascending inter-

leaved order, each 3 mm thick with an interslice gap of 0.3 mm (voxel size: 2 x 2 x 3mm). The flip angle was 90. Other imaging parameters were 2.5 s repetition time (TR) and 45 ms echo time (TE). We also acquired whole-brain high-resolution T1-weighted structural scans using an MP-RAGE sequence resulting in 160 slices (voxel size: 1 x 1 x 1 mm) (TR = 1.3 s, TE = 3.97 ms) to permit anatomical localization of the functional activations at the individual level. Diffusion-weighted imaging data was acquired immediately following the acquisition of T1-weighted structural images for purposes not relevant to this paper.

fMRI Data Preprocessing. Functional images were analyzed using the statistical parametric mapping software SPM12 (http://www.fil.ion.ucl.ac.uk/spm) implemented in MATLAB. Before statistical analysis, functional imaging data were subjected to the following preprocessing steps: (1) slice-timing correction was applied; (2) the realign procedure was used to perform motion correction; (3) the participants' T1 structural volume was co-registered to the mean of the corrected EPI volumes; (4) the group-wise DARTEL registration method included in SPM12 was used to normalize the T1 structural volume to a common group-specific space, with subsequent affine registration to Montreal Neurological Institute (MNI) space; (5) all EPI volumes were normalized to MNI space using the deformation flow fields generated in the previous step, which simultaneously resampled volumes to 2 mm isotropic, (6) and the EPI volumes were smoothed using a Gaussian kernel of 6 mm isotropic, full width at half maximum (FWHM).

fMRI Data Analyses. For each participant, a general linear model (GLM) estimated regressors of interest for each of the 168 trials in the fMRI task (56 products x 3 presentations), separately for each phase of the purchasing task (product consideration, product and price consideration and decision, response phase). The trial-specific regressors of interest of a particular task phase served as input for the theory-driven, region-of-interest (ROI) analyses (see details

below). Trials in the product and product and price consideration phases were defined by the onset and offset of the relevant information on-screen (i.e., product presentation and price display, respectively; see Figure S2). Trials in the response phase were defined by the onset of the decision prompt and participants' execution of a purchase decision (button press) in the trial. Note that we cannot reliably distinguish BOLD responses in the response phase from the previous price phase (due to the lack of variable inter-stimulus intervals between both phases and the "sluggishness" of the BOLD response). The GLMs included as covariates of no interest the six motion parameters estimated from image realignment. Neural activation was modeled by distinct regressors convolved with a canonic hemodynamic response function (hrf). A 128s high-pass cutoff filter was applied to eliminate low-frequency drifts in the data.

Selection of ROIs. We extracted data of three a priori defined regions of interest. The ROIs of the bilateral aI and vStr were created using the Desai atlas in AFNI (https://afni.nimh.nih.gov/AFNIAtlases). The vmPFC ROI was based on Neurosynth, a platform for large-scale, automated synthesis of fMRI data. Thus, ROIs were independently defined with regard to our key analyses—the neural prediction of market-level success of our 56 products—and with regard to our subject sample, reducing the risk of producing false positive results and of circular analysis (i.e. double dipping) (Kriegeskorte et al., 2009). The masks we used for all three ROIs are available on OSF.

ROI-specific activation was calculated by averaging across estimated regressor values of all voxels with the specified mask (separately for each of the three brain regions). For each ROI, we extracted and averaged product-specific data across the three regressors estimated for each product (per task phase), corresponding to the three product presentations in the fMRI purchase task. Data were extracted from two task phases: 1) the product consideration period and 2) the

product and price consideration period during the purchasing decision for a particular product. In line with the prior literature, we extracted from the vStr (Knutson et al., 2007) and aI (Tusche, Bode, & Haynes, 2010) mask during the first period and from the aI (Genevsky & Knutson, 2015; Knutson et al., 2007; Litt et al., 2011) and vmPFC (Genevsky, Yoon, & Knutson, 2017; Hare et al., 2008; Knutson et al., 2007; Litt et al., 2011) during the second period, yielding four values for each of the 56 products for every participant.

Sanity check and supplemental whole-brain analyses. We performed several post hoc analyses at the whole-brain level to further validate the selection of our regions of interest. We aimed to identify brain areas in which measured BOLD signals are systematically modulated by the participants' purchase decisions and perceived product desirability.

To this end, for each participant, we estimated additional GLMs (separately for our behavioral variables of interest listed above). Below, we describe the GLMs using the example of participants' purchase decision value (DV) on each trial. DVs are based on participants' button presses on each trial and coded so that higher values represent a positive purchase decision (1 = strong no, 2 = weak no, 3 = weak yes, 4 = strong yes). During the product and price consideration phase, participants had access to all the information necessary to make a purchase decision (i.e., product and price information) and were instructed to decide whether or not they would want to purchase the product for real. Thus, we hypothesized that DVs are encoded in the brain during the product and price phase of the task (i.e., before the subsequent response phase; see Figure S1). To test this idea, for each participant, we estimated a GLM with the following regressors:

-R1) a boxcar function for the *product consideration* phase on all trials (duration = 4 sec);

-R2) R1 modulated by the subject's stated DV on each trial;

- R3) a boxcar function for the *product and price consideration and decision* phase on all trials (duration = 4 sec);
- R4) R3 modulated by the participant's stated DV on each trial;
- R5) a boxcar function for the *response* period on all trials (duration = reaction time);
- R6) R5 modulated by the participant's stated DV on each trial;
- R7–R9) A boxcar function specifying missed trials, separately for each choice period (durations of 4 sec for product and price periods, respectively; duration of the response period = reaction time on that trial); and
- R10–R15) regressors of non-interest included six motion regressors as well as a session constant.

To examine whether brain responses obtained in the product and price consideration phase are modulated by the DVs, the regressor of interest (R4) was contrasted against an implicit baseline. Subject-specific contrast images were then used in a one-sample *t*-test at the group level (as implemented in SPM). We found that the bilateral vmPFC and the left aI were positively correlated with participants' purchase decision values during the price phase (Figure S3, Table S4) ($p<.001$, family-wise error, FWE, corrected at $p<.05$ at the cluster level). The right aI showed a similar response profile at $p<.001$ (uncorrected, whole-brain) but did not survive statistical correction for multiple comparisons (Figure S3). Next, we examined whether participants' perceived product desirability modulated brain responses during the product consideration phase. To this end, for each participant, we estimated a GLM that was like the one described above, with one exception: DV values in R2, R4, and R6 were replaced with participants' stated desirability of the product shown on a trial. Product desirability scores were based on participants' survey responses completed after the fMRI purchase task. Participants rated liking, attractiveness,

hypothetical purchase intention, and package liking (all on a scale from 1 = very much to 7 = not at all). We reversed the directionality of the scales such that higher values represented more positive product attitudes. We then integrated these positively correlated product attitudes into one averaged desirability index. To test whether neural signals during product consideration are modulated by products' perceived desirability, we contrasted R2 against an implicit baseline and subjected these contrasts to a one-sample *t*-test at the group level. We found that the left vStr and the right aI positively covaried with individuals' perceived product desirability (p <.001, FWE corrected at p <.05 at the cluster level) (Figure S2, Table S3). The same was true for the right vStr and the left aI at a slightly more lenient threshold (p<.001, uncorrected). Overall, these supplemental analyses provide strong support for the functional role of our a priori regions of interest during the choice-relevant periods in the fMRI purchasing task.

Additional References

Becker GM, DeGroot MH, Marschak J (1964) Measuring utility by a single-response sequential method. *Behav. Sci*. 9(3):226–232.

Genevsky A, Knutson B (2015) Neural affective mechanisms predict market-level microlending. *Psychol. Sci.* 26(9):1411–1422.

Genevsky A, Yoon C, Knutson B (2017) When brain beats behavior: Neuroforecasting crowdfunding outcomes. *J. Neurosci.* 37(36):8625–8634.

Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A (2008) Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors. *J. Neurosci.* 28(22):5623–5630.

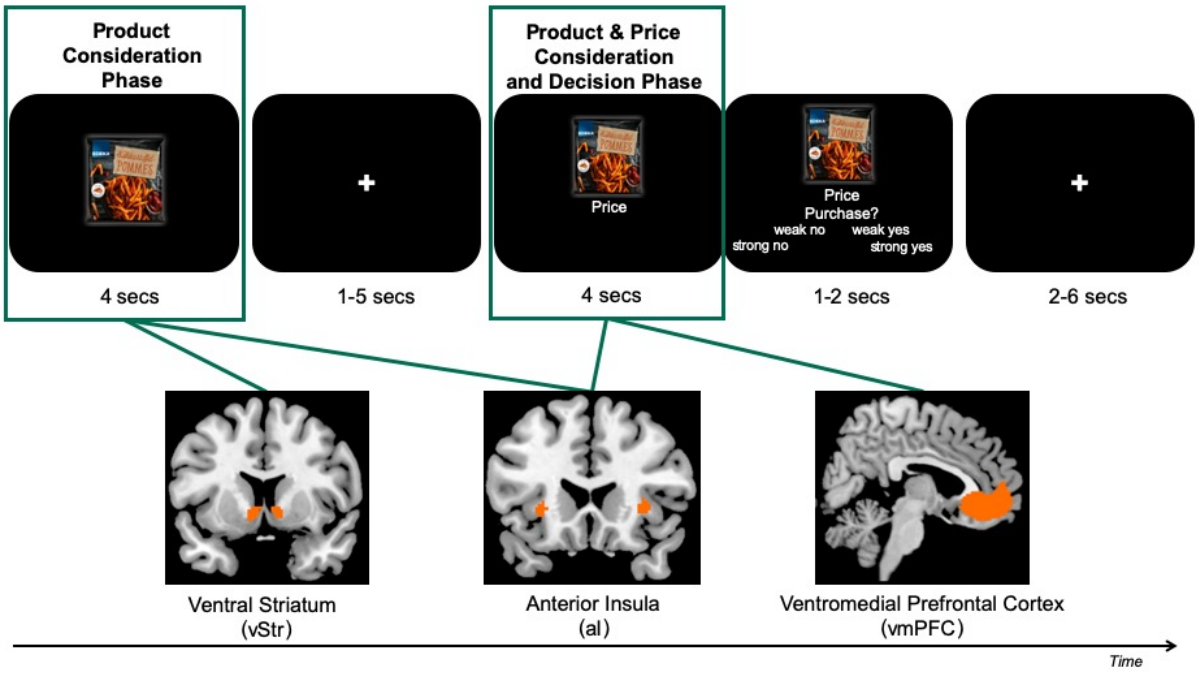
Knutson B, Rick S, Wimmer GE, Prelec D, Loewenstein G (2007) Neural predictors of purchases. *Neuron*, 53(1):147–156.

Kriegeskorte N, Simmons WK, Bellgowan PSF, Baker CI (2009) Circular analysis in systems neuroscience: The dangers of double dipping. *Nat. Neurosci*. 12(5):535–540.

Litt A, Plassmann H, Shiv B, Rangel A (2011) Dissociating valuation and saliency signals during decision-making. *Cereb. Cortex* 21(1):95–102.

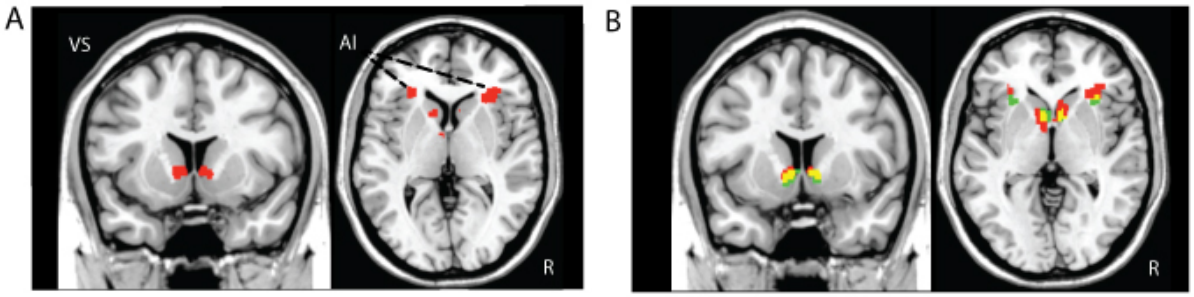
Tusche A, Bode S, Haynes J-D (2010) Neural responses to unattended products predict later consumer choices. *J. Neurosci.* 30(23):8024–8031.

Figure S1: fMRI task, adapted from the shop task of Knutson et al., 2007



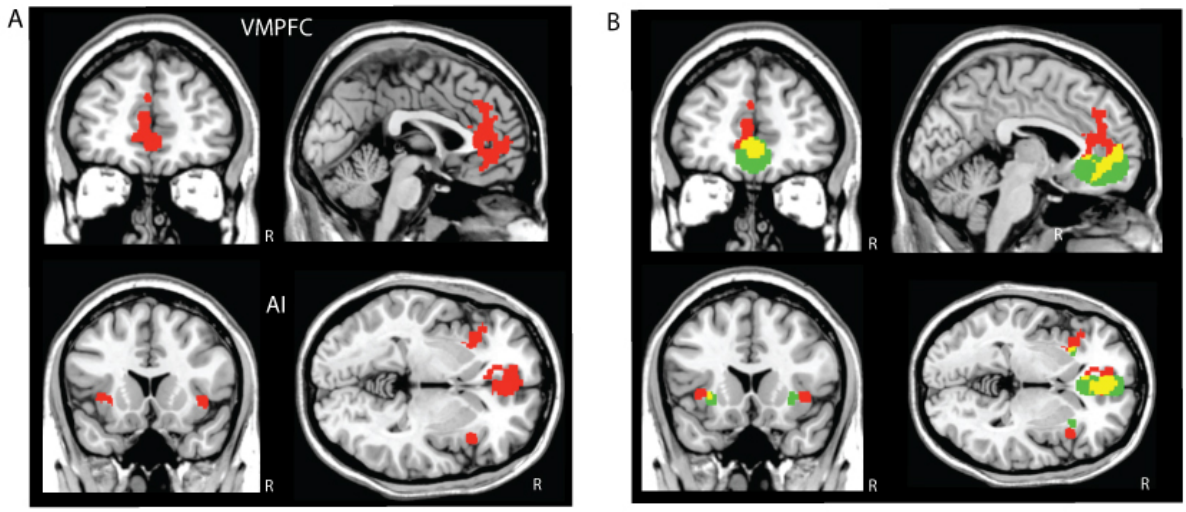
Note: Participants saw the product for 4 secs and then they also saw the price for which they could purchase the product for another 4 secs (separated by a fixation cross shown for a randomized length of 1–5 secs, mean 3 secs, as an inter-stimulus interval). Then, the decision period followed and lasted until the participant indicated the first response or, if the participant did not respond, lasted at least 1 sec with a maximum duration randomly chosen between 1 and 2 secs (mean time of decision period 0.97 secs). To separate different purchasing decisions, participants saw a fixation cross for a randomized length of 2–6 secs, mean 4 secs, as an inter-trial interval. The average time for a purchasing decision trial was 15.59 secs.

Figure S2: Neural correlates of product desirability during product consideration phase



Notes: **A.** Bilateral clusters in the ventral striatum (vStr) and the anterior insula (aI) encoding the desirability of products during the product consideration phase. For illustrative purposes, results are displayed at $p<.001$ uncorrected at the whole-brain level (the cluster in the left anterior insula the right ventral striatum did not survive FWE correction at $p<.05$. **B.** The panel illustrates the overlap (yellow) of the a priori ROIs in the vStr and aI (green) and the clusters identified in the fMRI subject sample (red).

Figure S3. Neural correlates of purchase decision (decision value) during the price phase



A. The figure illustrates the clusters in the ventromedial prefrontal cortex (vmPFC) and the bilateral anterior insula (aI) that covaried with participants' decision value (strong no, no, yes, strong yes) after all choice-relevant information was available (price phase). For illustrative purposes, results are displayed at $p<.001$ uncorrected at the whole-brain level. The cluster in the right anterior insula did not survive FWE correction at $p<.05$ (for details of the clusters in the left aI and the vmPFC see Table S4). **B.** The figure illustrates the overlap (yellow) of the a priori ROIs in the vmPFC and aI (green) and the clusters identified in our fMRI subject sample (red).

Table S1a: Overview of the products in the estimation set

Product category Products model estimation	Launch date product 1	Launch date product 2	Launch date product 3
Champignons	01.2014	01.2014	–
Chicken Nuggets	01.2014	–	–
Chili sauce	01.2014	37.2016	–
Currywurst	01.2014	01.2014	–
Potato fritter	18.2015	–	–
Gin	01.2014	09.2015	–
Gin 2	01.2014	34.2015	–
Ginger Beer	29.2014	40.2015	15.2016
Cucumber lemonade	01.2014	51.2014	16.2015
Oat bread	01.2014	50.2015	–
Licorice liquor	01.2014	01.2014	22.2015
Muesli	01.2014	–	–
Matcha	29.2014	11.2016	37.2016
Bar savory	42.2015	14.2016	–
Protein bar	07.2014	22.2016	–
Smoothie	28.2016	44.2016	–
Tuna	01.2014	17.2015	–

Table S1b: Overview of the products in the out-of-sample prediction set

Product category Products out-of-sample prediction	Launch date product 1	Launch date product 2	Launch date product 3
Champignons	–	–	02.2017
Chicken Nuggets	–	02.2017	02.2017
Chili sauce	–	–	02.2017
Currywurst	–	–	02.2017
Potato fritter	–	02.2017	06.2018
Gin	–	–	48.2016
Oat bread	–	–	14.2018
Muesli	–	49.2016	50.2016
Fries	49.2016	02.2017	02.2017
Bar savory	–	–	26.2017
Protein bar	–	–	52.2016
Juice	02.2017	02.2017	05.2018
Smoothie	–	–	14.2018
Tuna	–	–	02.2017

Table S2. Descriptive statistics (socio-demographic characteristics)

		Representative survey		fMRI experiment	
Socio-demographic characteristics		%	*N*	%	*N*
Gender	Male	40_a	*580*	47.7_a	*21*
	Female	60_a	*871*	52.3_a	*23*
Age	18–29	21_a	*305*	74.9_b	*33*
	30–39	23_b	*334*	24.9_b	*11*
	40–49	28_a	*406*	/	/
	50–59	28_a	*406*	/	/
Federal state	Schleswig-Holstein	3.8	*55*	/	/
	Hamburg	3.3	*48*	/	/
	Niedersachsen	8.5	*123*	/	/
	Bremen	0.6	*8*	/	/
	Nordrhein-Westfalen	22.7	*330*	95.5	*42*
	Hessen	6.2	*90*	/	/
	Rheinland-Pfalz	4.5	*65*	4.5	*2*
	Baden-Württemberg	11.2	*162*	/	/
	Bayern	17	*246*	/	/
	Saarland	1.3	*19*	/	/
	Berlin	6.1	*88*	/	/
	Brandenburg	2.1	*30*	/	/
	Mecklenburg-Vorpommern	1.9	*28*	/	/
	Sachsen	4.2	*61*	/	/
	Sachsen-Anhalt	4	*58*	/	/
	Thüringen	2.8	*40*	/	/
Employment	Public service	12.9	*187*	29.5	*13*
	Health	10.5	*153*	11.4	*5*
	Finance	4.2	*61*	11.4	*5*
	Media	/	/	6.8	*3*
	Public relations	/	/	6.8	*3*
	Marketing	/	/	6.8	*3*
	Retailing	/	/	4.5	*2*
	Other	73.9	*1072*	43.2	*19*
Education	Middle school	43.5_b	*631*	2.3_a	*1*
	High school	28.3_b	*411*	54.5_a	*24*
	University	28.2_b	*409*	43.2_a	*19*
Household lead	Mainly participant	62.6_b	*908*	45.5_a	*20*
	Participant with someone else	37.4_b	*543*	54.5_a	*24*
	Mainly someone else	/	/	/	/
Household size	One person	24.8_a	*360*	31.8_a	*14*
	Two persons	58_a	*842*	47.7_a	*21*
	Three or more persons	17.2_a	*249*	20.5_a	*9*
Income	No indication	11.2	*162*	4.5	*2*
	< €3.000	47	*682*	81.8	*36*
	> €3.000	41.9	*607*	13.6	*6*

Note: Means sharing the same subscript are not statistically different at the 5% level. Percentages sharing the same subscript are not statistically different on their proportions at the 5% level. Each subscript letter denotes a subset of sample categories whose column proportions do not differ significantly from each other at the 5% level.

Table S3. Neural correlates of the perceived desirability of consumer items during the product consideration phase in the fMRI purchase task.

Brain region	Side	k	t	MNI		
				x	y	z
Positive						
Ventral striatum (vStr)	L	158	4.70	-8	6	-2
Anterior insula (aI)	R	377	4.95	32	30	4
Subgenual cortex/anterior cingulate cortex	L	146	4.96	-10	28	-8
Posterior cingulate cortex	L/R	379	5.23	0	-36	36
Motor cortex	R	1256	5.95	40	-18	44
Parietal cortex/visual cortex	L	435	4.86	-28	-78	40
Parietal cortex/visual cortex	R	575	4.79	28	-64	52
Cerebellum	L	764	5.64	-22	-54	-20
Negative						
Motor cortex	L	641	6.49	-46	-20	58

Results are reported at a statistical threshold of $p < 0.001$, FWE corrected at $p < 0.05$ at the cluster level; only peak activations of clusters are reported. L = left hemisphere, R = right hemisphere, k = cluster size in voxels, MNI = Montreal Neurological Institute.

Table S4. Neural correlates of participants' decision values (strong no to strong yes) during the product and price consideration and decision phase of the fMRI purchase task.

Brain region	Side	k	t	MNI		
				x	y	z
Positive						
Superior frontal gyrus	L	192	5.16	-16	30	50
Inferior frontal gyrus	L	266	4.62	-38	40	2
vmPFC/anterior cingulate cortex	L/R	1226	5.61	2	36	12
Anterior insula (aI)	L	139	4.32	-36	22	-2
Midcingulate cortex	L/R	129	4.64	2	-34	36
Precuneus	L	181	4.47	-4	-62	32
Posterior insula/putamen	R	828	6.00	34	-4	-2
Motor cortex	R	146	4.88	18	-22	78
Precentral gyrus	R	180	5.42	28	-18	38
Inferior parietal cortex/angular gyrus	L	139	4.24	-34	-70	48
Cerebellum/visual cortex	L	2884	12.79	-20	-54	-18
Negative						
Motor cortex	L	1394	7.67	-56	-20	38
Motor cortex	L	482	5.20	-8	-16	78
Visual cortex	L	531	6.14	-8	-96	14
Cerebellum/visual cortex	R	2770	9.08	14	-50	-18

Results are reported at a statistical threshold of $p < 0.001$, FWE corrected at $p < 0.05$ at the cluster level; only peak activations of clusters are reported. L = left hemisphere, R = right hemisphere, k = cluster size in voxels, MNI = Montreal Neurological Institute.

Table S5. Parameter estimates of selected combinations of data types within the estimation set when adding vStr during product and price consideration phase

Variables	fMRI only	Combination of two data types	Combination of three data types	All data
Constant	4.81	4.43	1.73	2.73
	(4.32)	(3.29)	(4.49)	(9.19)
Behavioral data				
Incentivized purchases			1.09	0.22
			(1.38)	(2.37)
Market data				
Promotional activities		0.09	0.09	0.08**
		(0.10)	(0.10)	(0.06)
Market price		-0.13***	-0.13***	-0.11***
		(0.04)	(0.04)	(0.04)
Food product		-2.31***	-2.05***	-1.88**
		(0.86)	(0.83)	(1.15)
Survey data				
Success				-0.70
				(4.08)
Desirability				0.12
				(2.57)
Purchase intention				1.28
				(1.27)
fMRI Data				
vStr at product consideration	-10.43***	-7.65**	-7.72**	-8.41**
	(4.48)	(3.58)	(3.37)	(3.83)
vStr at product and price consideration				
	10.86*	8.22***	8.19***	9.08***
	(3.90)	(3.09)	(3.07)	(3.15)
aI at product consideration	7.09	7.57	7.46	10.03*
	(6.10)	(5.05)	(4.50)	(5.55)
aI at product and price consideration	-5.87	0.99	1.28	2.54
	(7.50)	(5.42)	(5.47)	(5.62)
vmPFC at product and price consideration	-1.36	-1.11	-0.62	-0.38
	(5.03)	(3.16)	(3.26)	(3.65)

***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Robust standard errors clustered at the product level in parentheses.

Table S6: Fit and prediction accuracy when adding vStr during product and price consideration phase

	In-sample prediction error			Out-of-sample prediction error	
	Adj. R-square	MAPE	% improv.	MAPE	% improv.
Baseline (constant only)	0.00	0.72		0.84	
Single data type					
Behavioral data (incentivized purchases)	0.08	0.63	12.7%	0.82	1.7%
Market data variables only (price, promotion, food product)	**0.28**	**0.52**	28.3%	**0.69**	**18.0%**
Survey data variables (success, desirability, purch. intention)	0.04	0.66	9.4%	0.81	3.3%
fMRI data variables (vStr, aI, vmPFC)	0.23	0.60	16.6%	0.71	15.5%
Combination of two data types					
Behavioral data + market data	0.28	**0.51**	29.6%	0.68	18.2%
Behavioral data + survey data	0.16	0.56	23.1%	0.70	16.5%
Behavioral data + fMRI data	0.29	0.53	26.4%	0.70	16.7%
Market data + fMRI data	**0.43**	**0.51**	29.2%	**0.59**	**30.0%**
Market data + survey data	0.30	0.54	25.9%	0.71	14.7%
Survey data + fMRI data	0.38	0.46	36.1%	0.98	-16.7%
Combination of three data types					
Behavioral data + market data + fMRI data	0.44	0.48	33.1%	**0.58**	**35.5%**
Behavioral data + market data + Survey data	**0.33**	**0.46**	30.3%	0.67	27.2%
Behavioral data + survey data + fMRI data	0.38	**0.46**	36.7%	0.94	-11.9%
Market data + survey data + fMRI data	0.45	**0.46**	36.3%	0.55	33.8%
Combination of four data types (all data)	0.45	0.48	34.1%	**0.51**	**38.6%**