Attribution Modeling

Evert de Haan

Contents

Abstract

Marketing attribution is the process of allocating appropriate credit to each marketing touchpoint a customer has encountered before conducting the desired customer action, e.g., a purchase. Ideally, this credit should be capturing the incremental effect of the touchpoint on the customer action. Finding this incremental effect is relevant for marketers to decide on budget allocations and to

E. de Haan (✉)
Departement of Marketing, Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands
e-mail: evert.de.haan@rug.nl

C. Homburg et al. (eds), *Handbook of Market Research*,
https://doi.org/10.1007/978-3-319-05542-8_39-1

decide how, when, and where to target which customer. This chapter introduces and discusses various marketing attribution techniques. The techniques range from basic attribution techniques, like touch-based attribution and Shapley values, to advanced attribution techniques, like randomized field experiments and Markov chains. The chapter discusses the up- and downsides of each attribution technique, discusses alternative methods if one method is inappropriate, and links this to the concept of incrementality and causality, i.e., to which degree the technique gives proper credits to the different channels or touchpoints the customer has encountered. This chapter is accompanied by the necessary R-scripts to generate the datasets and estimate the attribution techniques, which can also be downloaded at http://www.evertdehaan.com.

Keywords

Marketing attribution · Advertising effectiveness · Credit allocation · Attribution modeling · Last click · First click · Last touch · First touch · Shapley values · Causality · Incremental effects · Field experiments · Propensity scores · Matching · Markov models · R-code · R-script · Model estimation

Introduction

Due to digitalization and the rise of the Internet, it has become easier to track individual consumers in their (online) customer journey. For marketers, this means that they can get insights into which touchpoints a consumer has encountered. For example, how many times did the consumer see a banner advertisement, what did (s)he search for on a search engine, if (s)he used a price comparison site, and when and how (s)he has visited a firm's website. Furthermore, with online retailing, it is easily possible to observe which consumers have purchased and link the touchpoints to this purchase. The touchpoints and the moment of purchase together form a customer's *path to purchase*. A question that can arise when looking at such a path to purchase is which touchpoints have influenced the purchase decision. This question is the core of (online) marketing attribution.

Attribution is defined as the process to "allocate appropriate credit for a desired customer action to each marketing touchpoint across all online and off-line channels" (Kannan et al. 2016). The desired customer action is typically a conversion or a purchase. In other words, with attribution, one wants to find out:

- To what extent a (combination of) touchpoint(s) has/have impacted the likelihood to purchase (or the likelihood of another desired outcome) for an individual customer. This is individual-level attribution.
- How a (combination of) marketing channel(s) influence(s) the overall sales (or another desired outcome) for the firm. This is aggregate-level attribution.

As Hanssens (2021) has put it, "the key challenge in digital attribution is to estimate the *incremental* purchase probability achieved by a certain media intervention." The word "incremental" is crucial here, i.e., the difference in the purchase probability, because of a specific media intervention (i.e., touchpoint or marketing outing). Finding this incremental effect and thus conducting attribution is relevant for marketers to decide on budget allocations and to decide how, when, and where to target which customer. The importance of this is clear from some practical examples. When eBay stopped using search engine advertising (SEA), they saw in many cases no change in traffic to eBay's website because of the substitution coming from organic (i.e., nonpaid) search engine traffic (Blake et al. 2015). Procter and Gamble cut $200 million in digital ad spend and reallocated this to other channels, including television and radio, which increased its reach by 10% (Johnson 2018). When Uber cut two-third of their ad spending, which saved $100 million in costs, they saw almost no change in app installations (WARC 2021). Knowing which touchpoints contribute to the desired outcome is thus crucial for a firm's bottom-line performance and can help to improve marketing effectiveness.

This chapter will discuss techniques for both individual- and aggregate-level attribution. The chapter furthermore provides R-scripts for the attribution techniques and to generate the datasets. The datasets are easy to adjust, e.g., the scripts can easily be changed to include additional consumers and additional marketing channels or to adjust the model's assumptions and effect sizes. With the script to estimate the attribution models, it is also easy to apply the models to other, e.g., real-life datasets. All of this makes the datasets and R-scripts, and hence this chapter, useful for teaching purposes and marketing practitioners.

The next section of this chapter introduces the dataset for the individual-level attribution. Hereafter, some basic attribution methods are discussed, including touch-based attribution, regression-based models, and Shapley values. Section "Attribution Modeling Process with Experimental Data," introduces a kind of golden standard for attribution, namely a randomized field experiment at the individual customer level. In section "Additional Topics for Individual-Level Attribution," attribution methods are discussed when the ideal data is not available or when additional insights are needed, including propensity score matching and Markov models. In section six, some methods for aggregate-level attribution are discussed. The final section concludes this chapter.

Dataset

We will use the same dataset throughout the next three sections, covering individual consumer-level attribution. The dataset can be generated with the R-script or be downloaded directly, together with all other datasets and R-scripts used in this chapter, at http://www.evertdehaan.com. The dataset contains 50,000 customer journeys, with ~25% of the journeys resulting in a purchase and each journey having between 1 and 50 touchpoints in total. We have in total eight unique touchpoints, ranging from banner impressions and clicks to SEA for brand- and product-related keywords.

The first step in attribution is thus to have the right data available to conduct attribution modeling. For managers and researchers, who want to apply the techniques from this chapter on their own data, we recommend to collect the right data and structure it in the right way. The right data would (at an individual customer level) include the different touchpoints a customer has encountered over time, and can be further enriched with customer-specific information (e.g., demographic and customer relationship data) and other browsing behavior (e.g., clickstream data which provides information on what a customer does on the website, see for instance also Moe 2003). The structure of these data can be similar to the example data, discussed in this section and used in the following three sections. If individual-level data is not available (in general, or for some channels), another option would be to use time series data; more on the structure of this type of data and how to model attribution with these data are discussed in the section "Attribution with Aggregate-Level (Quasi-)Experimental Data."

We can get some descriptive statistics of the journeys with the following R-script.

```
# Get descriptive statistics
library(psych)
describe(consumers)
```

Table 1 shows some of the descriptive results, together with a description of what each variable measures. The 7th ("Banner_no_click") up until the 14th ("Direct_visit") variable measure the number of occurrences for the eight different touchpoints in each path to purchase. With attribution, we want to see how these touchpoints influence the purchase, i.e., the 19th (last) variable. We furthermore have information on the length of the customer relationship and the customer lifetime value. The dataset contains data from two field experiments, captured with the "Firm_banner" and "Flyer_region" variables. In the section "Attribution Modeling Process with Experimental Data," we will explain these experiments further.

An example of a path to purchase, which will come back throughout this chapter, is shown in Fig. 1. This path of purchase is from customer #134 from the dataset. We see that this customer is first coming into contact with a banner by the firm but does not click on it. After this, the customer uses a search engine where (s)he uses a product-related keyword and clicks on a sponsored search link to visit the website. The two subsequent visits are direct visits to the website, which occur by typing in the website's URL. Hereafter are again two banner impressions; the customer clicks on the second banner impression, leading to another website visit. Next are another direct visit and a banner impression. Finally, the customer uses a search engine to search for the company (i.e., branded search), clicks on the sponsored search link, and conducts a purchase.

When looking at Fig. 1, we might ask which of the nine touchpoints has made sure that this customer has conducted the purchase. Was this because of the branded SEA in the end? Or did the banner impression which in the beginning started this

Table 1 Variable descriptions and descriptive statistics

Variable name	Description	Mean	sd	Median	Min	Max
1. Consumer_ID	A unique customer ID				1	50,000
2. Existing_customer	Dummy indicating if the customer made a purchase before	0.50			0	1
3. Relation_length	Amount of months active at the firm (measured on January 1)	15.25	19.53	1	0	60
4. CLV	Customer lifetime value in dollars (measured on January 1)	904.43	1197.41	55.64	0	7341
5. Firm_banner	Dummy variable indicating if the customer was in the firm's banner (1) or charity banner (0) group	0.80			0	1
6. Email_group	Dummy variable indicating if the customer signed up for the email newsletter (1) or not (0) (introduced on January 1)	0.27			0	1
7. Banner_no_click	Count variable, indicating the number of banner impressions without a click (from the firm or charity)	2.04	2.49	1	0	29
8. Banner_click	Count variable, indicating the number of banner clicks (for the firm or charity banner)	0.12	0.39	0	0	6
9. SEA_product_click	Count variable for the amount of website visits through sponsored search results using product-related keywords	0.50	0.89	0	0	10

(continued)

Table 1 (continued)

Variable name	Description	Mean	sd	Median	Min	Max
10. SEA_brand_click	Count variable for the amount of website visits through sponsored search results using firm−/brand-related keywords	0.81	1.21	0	0	14
11. Price_comp_click	Count variable, indicating the number of website visits through a price comparison site	0.44	0.84	0	0	12
12. Email_no_click	Count variable, indicating the number of emails received without clicking on a link and visiting the website	0.10	0.45	0	0	7
13. Email_click	Count variable, indicating the amount of email links clicked on and, i.e., visiting the website	0.05	0.26	0	0	7
14. Direct_visit	Count variable, indicating the number of direct website visits (e.g., by typing in the URL)	2.49	2.82	2	0	32
15. First_channel*	String variable, naming the first channel in the path to purchase					
16. Last_channel*	String variable, naming the last channel in the path to purchase					
17. Amount_touchpoints	Count variable, summing up all touchpoints in the path to purchase	6.53	5.99	5	1	50

(continued)

Table 1 (continued)

Variable name	Description	Mean	sd	Median	Min	Max
18. Flyer_region	Dummy variable, indicating if the consumer lives in a region where the firm distributed their flyers (1) or not (0)	0.50			0	1
19. Purchase	Dummy variable, indicating if the path to purchase ended with a purchase (1) or not (0)	0.26			0	1

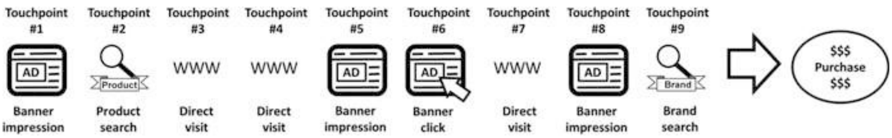

Fig. 1 Example customer path to purchase

path to purchase cause the end result? Or should all touchpoints get (equal or different) credits for the purchase? Or would the customer also have conducted the purchase without any of the (advertisement-based) touchpoints?

We want to determine what would have happened if certain (combinations of) touchpoints were not there. If the conversion still happens without a specific touchpoint, then that touchpoint should not get any credit for the conversion. If the conversion would not happen without the touchpoint, that touchpoint should get credit. A challenge is that we can only observe the path to purchase as it was, i.e., we do not observe if the outcome would be different if the path to purchase would be different. Luckily, attribution techniques can provide insights into this, as discussed in the following sections of this chapter.

Problems with Basic Attribution Methods

This section discusses some basic attribution methods, which have been used a lot in practice, but have the downside that they do not answer the attribution question, namely what the incremental effect of a channel is. We discuss these methods, to show how they work, what insights they bring, and to what extent they relate to the attribution question and are thus useful for practice. We will start with touch- or click-based attribution methods, followed by correlations and regression models, and finally, we discuss the more advanced Shapley value–based attribution method.

Touch-Based Attribution

An advantage of online attribution is that we can observe all consumers' online touchpoints, website visits, and conversions. Based on these data, we can observe the customers' path to purchase and perform *touch-based attribution*, sometimes also called *click-based attribution*. Historically, the most popular form of touch-based attribution is last-touch attribution, i.e., the last touchpoint a customer comes into contact with before a purchase gets all credit for that purchase. In the example from Fig. 1, last-touch attribution would thus give branded SEA full credits for the purchase. There are also other touch-based attribution methods, e.g., first-touch-based attribution gives full credit to the first touchpoint, average (or linear) touch attribution gives all touchpoints equal credit, and time decay attribution gives the lowest credit to the first touchpoint and the highest credit to the last touchpoint. Figure 2 visualizes the different touch-based attribution methods and also indicates what this means for the journey from Fig. 1.

To see to what extent touch-based attribution can come to different conclusions, let us use the dataset introduced in the previous section. The following R-script conducts three different touch-based attribution techniques. Table 2 presents the outcomes of the R-script.

```
# Number of occurrences of touchpoint
consumers_journey$n <- 1
aggregate(consumers_journey$n, by=list(consumers_journey
$Channel_name), FUN=sum)
consumers_journey$n <- NULL
# Last-touch, total amount conversions
aggregate(consumers$Purchase, by=list(consumers$Last_channel),
FUN=sum)
# First-touch, total amount conversions
aggregate(consumers$Purchase, by=list(consumers$First_channel),
FUN=sum)
# Average-touch, total amount conversions
sum(consumers$Banner_click/consumers$Amount_touchpoints*consumers
$Purchase)
```

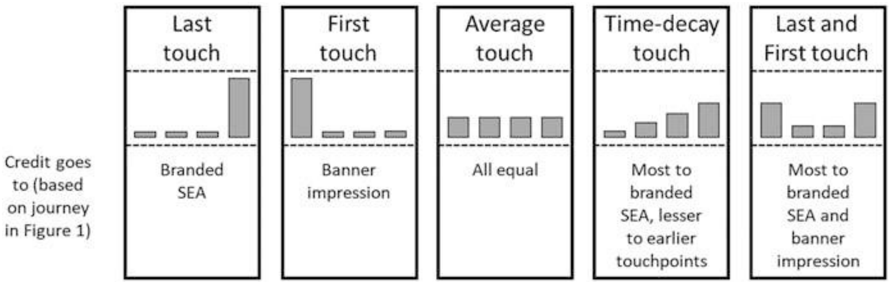

Fig. 2 Touch-based attribution examples

Table 2 Credit allocation using touch-based attribution

		Conversions attributed to channel			Conversion percent per channel		
	n	Last touch	First touch	Average touch	Last touch	First touch	Average touch
Banner click	6,029	154	185	217.6	2.55%	3.07%	3.61%
Banner impression	101,930	0	3,979	3,309.4	0.00%	3.90%	3.25%
Direct	124,486	7,838	4,716	5,426.5	6.30%	3.79%	4.36%
Email click	2,265	166	159	157.1	7.33%	7.02%	6.94%
Email received	5,141	0	354	253.8	0.00%	6.89%	4.94%
Price comparison	21,872	1,468	1,212	992.4	6.71%	5.54%	4.54%
SEA branded	40,253	2,473	1,166	1,633.4	6.14%	2.90%	4.06%
SEA product	24,773	692	1,020	800.8	2.79%	4.12%	3.23%
Total	**326,749**	**12,791**	**12,791**	**12,791**	**3.91%**	**3.91%**	**3.91%**

```
sum(consumers$Banner_no_click/consumers
$Amount_touchpoints*consumers$Purchase)
sum(consumers$Direct_visit/consumers$Amount_touchpoints*consumers
$Purchase)
sum(consumers$Email_click/consumers$Amount_touchpoints*consumers
$Purchase)
sum(consumers$Email_no_click/consumers
$Amount_touchpoints*consumers$Purchase)
sum(consumers$Price_comp_click/consumers
$Amount_touchpoints*consumers$Purchase)
sum(consumers$SEA_brand_click/consumers
$Amount_touchpoints*consumers$Purchase)
sum(consumers$SEA_product_click/consumers
$Amount_touchpoints*consumers$Purchase)
```

As shown in Table 2, the three different touch-based attribution methods come to different conclusions. First of all, with last touch, a sale can only be attributed to a channel that leads directly to a website visit; this is why a mere banner impression or a received email that does not receive a click do not get credit for a conversion. With first touch and average touch, also mere exposures can get credit. For these latter two attribution methods, banner impressions get a relatively high amount of credit. Sometimes for first-touch attribution and average-touch attribution also only actually clicks are being used (i.e., click-based instead of touch-based attribution), which would change the numbers since the banner impression and received email would then drop out.

Direct visits and branded SEA get much credit with last-touch attribution but less with the other two attribution methods. This difference might be because at the end of a (longer) customer journey, the consumer already knows where to buy and goes directly to the website (or types in the brand name at a search engine). Product-related SEA, i.e., people clicking on a sponsored search link when using a product-related keyword, gets relatively more credit with first-touch and average-touch attribution. This difference might also be due to the stage in which the customer uses product-related SEA. Product-related SEA is mainly used at the start when the customer looks for broader information and is not sure yet where to buy the product. These differences are also in line with findings from Rutz and Bucklin (2011).

A downside of all the touch-based attribution methods is that none of these methods answer the attribution question, namely "to what an extent a (combination of) touchpoint(s) has/have impacted the likelihood to purchase." We do, namely, not observe if the conversion would not have happened without a specific (combination of) touchpoint(s). The touch-based attribution methods do thus not work well for attribution purposes, as scientific studies have also shown (e.g., De Haan et al. 2016; Li and Kannan 2014). Instead, we have to use alternative methods that tell us if the effects are causal, i.e., without the touchpoint the outcome (e.g., purchase) would have been different. The criteria for causality, and hence suitable attribution, are:

1. **Covariation:** A shock in the independent variable (i.e., the exposure to a touchpoint) correlates with a shock in the dependent variable (i.e., the desired customer outcome).
2. **Order in time:** The shock in the independent variable has to occur before the shock in the dependent variable.
3. **No third variable:** There are no other variables or reasons that explain the effect (e.g., confounds like seasonality or self-selection bias). In other words, the change in the dependent variable is due to the change in the independent variable (i.e., the touchpoint occurring in the path to purchase).

Touch-based attribution methods do not meet the third criterion since the exposure to a touchpoint and purchase might be driven by, for instance, seasonality, e.g., during peak seasons, the number of purchases is higher, and firms spend more on marketing activities. Another explanation might be the self-selection by consumers, e.g., consumers who sign up for a newsletter are already more likely to purchase in the future, even without receiving the newsletter. We thus have to find alternative methods that can give more certainty about the causality to conduct accurate attribution, especially in terms of excluding all potential third variables.

More information on causality is also provided in other chapters in this book, e.g., Artz and Doering (2021), Bornemann and Hattula (2018), Ebbes et al. (2016), and Valli et al. (2017).

Correlations and Regression Models

Another, more algorithmic, way of doing basic attribution is looking at correlations or estimating a regression model. In such a way, one can relate the touchpoints a customer has come into contact with to the dependent variables of interest, e.g., a purchase. Furthermore, it is possible to see how the touchpoints correlate to each other and other variables like the consumer's characteristics. Let us investigate with the following R-script some correlations in our dataset.

```
# Make correlation plot
library(corrplot)
my_data <- consumers[, c(2,3,4,5,6,7,8,9,10,11,12,13,14,17,18,19)]
res <- cor(my_data)
corrplot(res, type = "upper", tl.col = "black", tl.srt =
45, sig.level = 0.01, insig = "blank")
rm(my_data)
rm(res)
```

Figure 3 visualizes the output of the correlation analysis. We can observe that the "purchase" variable correlates positively, but weakly, with most other variables. "Purchase" is strongest correlated with the "direct visit" and "amount touchpoints" variables. These correlations mean that the direct visits and the path to purchase length are the strongest indicators of a purchase. This could make sense since consumers who visit the website directly might be more likely to know the online retailer already. Indeed, we can observe that direct visits also positively correlate with the "existing customer," "relation length," and "CLV" (customer lifetime value, i.e., how much the customer's future transactions are worth in terms of net present value) variables, indicating that these consumers might already be loyal to the online

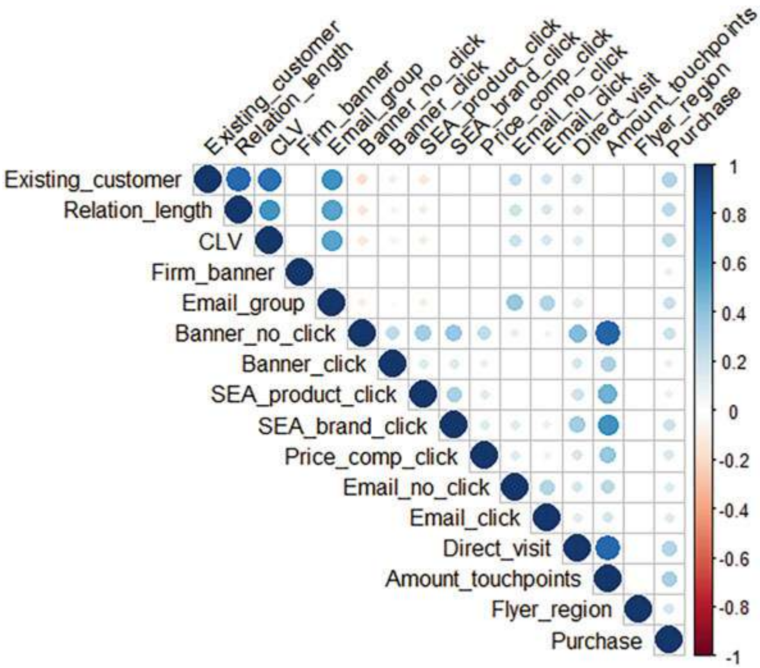

Fig. 3 Correlation matrix

retailer. In this case, it might thus not be the channel that is driving the purchase, but the type of consumer that uses the channel determines the chance of a conversion.

Furthermore, longer paths to purchase might relate to more informed and interested consumers and are more likely to convert. This indicates that the amount of touchpoints is not per se driving purchase, but it is the consumer's underlying "degree of interest." Correlations thus have to be interpreted with caution since they only tell us something about the association between variables but do not give us causal information.

When looking at the correlations from Fig. 3, we can furthermore see that the banner variables and the SEA product variable are positively related to each other, which might relate to the phases in the path to purchase; consumers who first see a banner might later be interested in clicking on it, and later search for product information. Correlation tables can thus be convenient to understand patterns in the data, which could be input for more complicated methods to find out causal effects.

A more elaborate form of investigating the associations between variables is by estimating a regression model (see for more details Skiera et al. [2018]). An advantage of a regression model over correlations is that we can control for third variables and thus can get a step closer to finding causal relations. Since the dependent variable "purchase" is binary, a logistic regression model is most appropriate in our case (see for more details Tillmanns and Krafft [2017]). With the following R-script, we estimate some logistic regression models.

```
# Load package to make an output table of all models
library(sjPlot)
library(sjmisc)
library(sjlabelled)
# Logistic regression models to predict the purchase likelihood
model1 <- glm(Purchase ~ as.factor(Last_channel), data=consumers,
family=binomial)
model2 <- glm(Purchase ~ as.factor(Last_channel) +
Existing_customer + Relation_length + log(CLV+1) +
Amount_touchpoints + Flyer_region, data=consumers,
family=binomial)
tab_model(model1, model2, transform = NULL, collapse.ci = TRUE,
p.style = "stars")
```

Table 3 shows the output of the models. Model 1 only looks at the last channel used and is thus very similar to the last-touch attribution procedure. "Banner_click" is here the reference category, meaning that the interpretation of the parameters is relative to this touchpoint. The variables "Banner_impression" and "Email_received" have a strong negative parameter estimate, which makes sense because these variables do not directly relate to a conversion, as discussed in the previous section. The parameter for "Direct_visit" is positive and significant, meaning that there is a significantly higher chance of conversion when this channel is the last touchpoint than when "Banner_click" is the last touchpoint. These parameters are all in line with what we saw in Table 2. When we control for some customer-

Table 3 Logistic regression output using touch-based attribution

	Model 1	Model 2
	Purchase	Purchase
Predictors	*Log-Odds*	*Log-Odds*
(Intercept)	−1.05 *** (−1.24 – −0.87)	−3.47 *** (−3.70 – −3.25)
Last_channel [Banner_impression]	−17.52(−99.01 – −124.06)	−18.43(−87.53 – −111.02)
Last_channel [Direct_visit]	0.28 ** (0.10–0.47)	0.08(−0.13–0.30)
Last_channel [Email_click]	2.12 *** (1.77–2.48)	1.76 *** (1.37–2.15)
Last_channel [Email_received]	−17.52(−498.46 – −646.32)	−18.71(−566.97 – −613.56)
Last_channel [Price_comp_click]	0.22 * (0.03–0.42)	0.26 * (0.04–0.48)
Last_channel [SEA_brand_click]	0.24 * (0.05–0.43)	0.13(−0.09–0.35)
Last_channel [SEA_product_click]	−0.25 *(−0.45 – −0.05)	−0.20(−0.43–0.04)
Existing_customer		−0.42 **(−0.72 – −0.12)
Relation_length		0.01 *** (0.01–0.01)
CLV + 1 [log]		0.23 *** (0.19–0.27)
Amount_touchpoints		0.15 *** (0.15–0.16)
Flyer_region		1.05 *** (1.00–1.10)
Observations	50,000	50,000
R^2	0.077	0.294

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

specific variables, as is done in Model 2 in Table 3, we can see that "Direct_visit" becomes statistically insignificant. This change in significance might be because the direct channel is used more by existing and loyal customers, who also have a higher chance of conversion, i.e., these variables are confounding variables. Indeed, if the consumer is an existing customer of the firm and has a long relationship and a high CLV, the likelihood of conversion is higher, i.e., the type of customer using the channel (partly) explains the effectiveness of the channel. With this, we demonstrate the advantage of a regression model compared to merely looking at correlations, namely that we can correct for some confounding variables.

Although the correlation table nicely shows how the different channels relate to each other, which can give us an idea about what touchpoints belong together to target specific customers, and how the channels relate to purchase, it does not give us information on the causal effects. As was the case with touch-based attribution, potential third variables are, however, not excluded. The logistic regression model can be more appropriate for attribution than the touch-based attribution since we can incorporate third variables as control variables, like the strength of the relation with the customer. It still does not provide us with causal information since we cannot say with certainty that we control for all other factors and what would have happened if a particular channel or touchpoint did not occur. Hence, also correlations and regression models do not address the underlying question of attribution.

Shapley Value–Based Attribution

As discussed, with attribution, we want to determine what would have happened to the outcome variable of interest provided that a channel was not there. Both the touch-based attribution and attribution based on correlations and a regression model do not answer the attribution question. A method that comes closer, and is also more popular in recent years, is Shapley value–based attribution, which has its roots in cooperative game theory (Shapley 1953). This method compares similar paths to purchase, with the only difference that some paths contain a specific touchpoint, but the other paths do not contain this touchpoint.

To give an example of this method, let us imagine a simplified path to purchase of Fig. 1. Assume we have consumers who first comes into contact with a banner, and the consumer does not click on it. After this, these consumers search for information using a product-specific keyword and visit the website by clicking on a SEA link, and finally the consumers search for information using a brand-specific keyword and revisit the website by clicking on a SEA link. If we have a large enough group of these consumers, we can calculate the conversion probability of such a path. Similarly, we can investigate another group of consumers, but with the difference that the first touch point (i.e., the banner impression) is not included in the path to purchase. If we compare the first path to purchase's conversion probability with the second path to purchase, we can calculate the increase in conversion probability if an extra touchpoint was there. This procedure is what we do with the following R-script.

```
# Get all paths to purchase that are "Banner impression" -> "SEAR
product click" ->  "SEA brand click"
consumers_shapley <- consumers[consumers$Amount_touchpoints==3,]
consumers_shapley <- consumers_shapley[consumers_shapley
$First_channel=="Banner_impression",]
consumers_shapley <- consumers_shapley[consumers_shapley
$Last_channel=="SEA_brand_click",]
consumers_shapley <- consumers_shapley[consumers_shapley
$SEA_product_click==1,]
mean(consumers_shapley$Purchase)
# Get similar paths to purchase excluding "Banner impression"
consumers_shapley_ex1 <- consumers[consumers
$Amount_touchpoints==2,]
consumers_shapley_ex1 <- consumers_shapley_ex1
[consumers_shapley_ex1$First_channel=="SEA_product_click",]
consumers_shapley_ex1 <- consumers_shapley_ex1
[consumers_shapley_ex1$Last_channel=="SEA_brand_click",]
mean(consumers_shapley_ex1$Purchase)
# Get similar paths to purchase excluding "SEAR product click"
consumers_shapley_ex2 <- consumers[consumers
$Amount_touchpoints==2,]
consumers_shapley_ex2 <- consumers_shapley_ex2
[consumers_shapley_ex2$First_channel=="Banner_impression",]
consumers_shapley_ex2 <- consumers_shapley_ex2
[consumers_shapley_ex2$Last_channel=="SEA_brand_click",]
```

```
mean(consumers_shapley_ex2$Purchase)
# Get similar paths to purchase excluding "SEA brand click"
consumers_shapley_ex3 <- consumers[consumers
$Amount_touchpoints==2,]
consumers_shapley_ex3 <- consumers_shapley_ex3
[consumers_shapley_ex3$First_channel=="Banner_impression",]
consumers_shapley_ex3 <- consumers_shapley_ex3
[consumers_shapley_ex3$Last_channel=="SEA_product_click",]
mean(consumers_shapley_ex3$Purchase)
# Impact channel 1
mean(consumers_shapley$Purchase) - mean(consumers_shapley_ex1
$Purchase)
# Impact channel 2
mean(consumers_shapley$Purchase) - mean(consumers_shapley_ex2
$Purchase)
# Impact channel 3
mean(consumers_shapley$Purchase) - mean(consumers_shapley_ex3
$Purchase)
```

Figure 4 visualizes the output of the R-script above. We can see that the full path to purchase discussed above occurs 86 times in our dataset, and in 16.28% of the cases result in a conversion. When we leave out the banner impression, the reduced path to purchase occurs 267 times in our dataset, and in 7.49% of the cases result in a conversion, meaning that when the banner impression is left out, the conversion

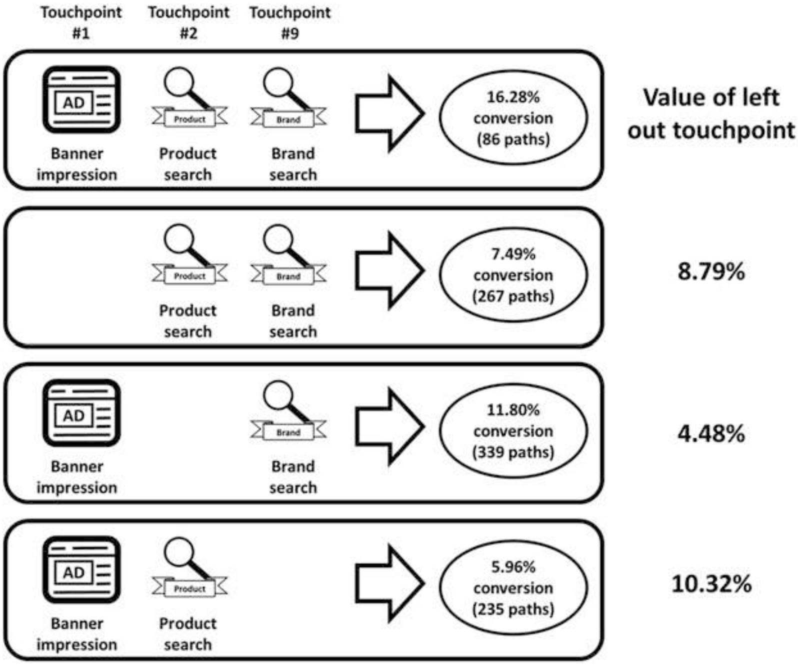

Fig. 4 Shapley value–based attribution example

probability drops 8.79% points. This difference is the credit the banner impression should get. Similarly, Fig. 4 provides the value of the other two touchpoints, with branded SEA getting the highest credit of a 10.32% points increase in the conversion probability. It is important to note that we get a higher conversion rate when adding up the value of each touchpoint than the observed conversion of 16.28%. There can be multiple reasons for this, e.g., there could be a synergy effect, meaning that the combination of touch points is responsible for the purchase, and dropping one thus has a substantial effect. There could also be a time effect, e.g., a path to purchase with three touchpoints generally has a higher conversion since a longer path to purchase could indicate more interested consumers.

Of course, the example given can easily be applied to all kinds of paths to purchase with different combinations of touchpoints. Doing so can provide information in what situation which touchpoint has the largest impact on conversion. Furthermore, the Shapley value procedure is appropriate to investigate the impact of the order of channels, achievable by comparing similar paths to purchase, with a difference that the order of the touch points differs. A challenge is that specific paths to purchase might only occur a limited number of times in the dataset, making the calculated conversion probability less reliable.

In general, the Shapley value approach is a more appropriate way of attribution than touch-based or regression-based attribution. This Shapley value procedure comes close to answer the attribution question "what would have happened if certain (combinations of) touchpoints were not there," i.e., "estimate the *incremental* purchase probability achieved by a certain media intervention" from Hanssens (2021). It does, however, not fully answer this question. First of all, the paths to purchase we are comparing come from different consumers, and the characteristics of these consumers might also (partly) explain the differences in the conversion probabilities. Secondly, touchpoints do not occur at random, e.g., the branded SEA only occurs if the consumer is actively looking for a brand using a search engine. If branded SEA would have been off, the consumer would probably still have visited the website but instead would have clicked on an organic search link to do so or would visit the website directly. Shapely values thus still have some limitations, and we cannot be sure the findings are causal. Because of that, we have to use more advanced techniques and make use of certain features in the data (e.g., a field experiment or other form of randomization), which can provide us with causal insights. The following two sections discusses some of these techniques.

Attribution Modeling Process with Experimental Data

An ideal form of attribution is when there is data available, coming from a field experiment where one (or multiple) channels are turned off for some groups of customers. This "golden standard" is especially useful for individual-level attribution. To find the impact of a channel, consumers are randomly placed into two groups: one group that can encounter the channel of interest (i.e., the treatment group) and one group that cannot encounter this channel (i.e., the control group).

Due to the random allocation of consumers in these two groups, we can assume that consumers of the two groups only differ in terms of being targeted or not being targeted via one specific channel (see for more details Landwehr (2019)). The dataset generated in the "Dataset" section of this chapter includes a randomized field experiment, similar to the experiments by Hoban and Bucklin (2015) and Li et al. (2021). In this experiment, 80% of the consumers are in the group which can encounter the banner ad from the firm (i.e., our treatment group). The other 20% of the consumers are in the group which can encounter the banner ad from an unrelated charity organization (i.e., our control group). The control group uses an unrelated charity ad since this makes it possible to observe how many ads from the firm the consumer would have seen if the consumer was in the treatment group. The actual treatment (i.e., the firm ad) was thus not provided to the control group, but the charity ad captures for the consumers in the control group if they would have seen the firm ad (and also, how many times) or if they would have been in the treatment group. This method assumes that the targeting of the charity ad in the control group is set up the same as the targeting of the firm's ad in the treatment group. For more details on this method, see Hoban and Bucklin (2015).

Since the charity ad is unrelated to the firm, the causal effect of this ad on firm performance and customer behavior should be zero. The number of exposures to the charity ad might, however, correlate with the conversion probability. This correlation might occur since someone who sees the ad more often might (1) be online more often and (2) visit the website where the campaign is running more often, which both relate to the conversion probability. The charity ad captures these confounding effects, and the difference between the treatment group (containing both the causal and confounding effects) and the control group (containing only the confounding effects) thus captures the causal effect of the firm's ad.

Next to banner advertising, the dataset also contains a second random experiment. The firm distributed a flyer in some randomly selected regions (e.g., based on postal code), while this was not the case in other regions. This setup leads to a random 50% of the consumers in the dataset receiving a flyer from the firm. Since we have information in which region the consumers live (e.g., based on sign-up information from the customer or their IP address information), we know which individual consumers did get the flyer. This setup thus allows us to investigate the effectiveness of this offline advertising form at the individual consumer level. We can furthermore investigate if there are synergy effects between the banner ads and the flyer, i.e., if being exposed to both advertising forms increases the purchase likelihood beyond the two individual effects (i.e., positive synergy) or if they weaken each other since they might be substitutes (i.e., negative synergy). This setup of distributing flyers is somewhat in line with the study by Wiesel et al. (2011), although they have conducted and analyzed their experiment at a higher level of aggregation, namely at the regional level instead of the individual consumer level.

A good thing to inspect first is if consumers in the firm's banner group and who have received a flyer indeed have a higher likelihood of purchasing. Furthermore, we can explore if there is a synergy effect between the two channels.

```
# Plot of banner ad and purchase likelihood
library(ggplot2)
myData <- aggregate(consumers$Purchase,
                    by = list(Firm_banner = consumers$Firm_banner),
                    FUN = function(x) c(mean = mean(x), sd = sd(x),
                                        n = length(x)))
myData <- do.call(data.frame, myData)
myData$se <- myData$x.sd / sqrt(myData$x.n)
colnames(myData) <- c("Firm_banner", "mean", "sd", "n", "se")
myData$names <- c(paste(myData$Firm_banner, "Firm_banner"))
p <- ggplot(data = myData, aes(x = factor(Firm_banner), y = mean))
p + geom_bar(stat = "identity",
             position = position_dodge(0.9), fill="steelblue") +
  geom_errorbar(aes(ymax = mean + 2*se,
                    ymin = mean - 2*se), position = position_dodge
(0.9),
                width = 0.25) +
  labs(x = "Firm_banner", y = "Conversion rate") +
  ggtitle("Conversion rate by firm_banner") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  geom_text(position = position_dodge(width= .9), aes(y = mean,
label = paste(format(mean*100, digits = 2, nsmall = 2), "%",
sep = "")), size = 4, vjust = 5)
# Plot of flyer and purchase likelihood
myData <- aggregate(consumers$Purchase,
                    by = list(Flyer_region = consumers
$Flyer_region),
                    FUN = function(x) c(mean = mean(x), sd = sd(x),
                                        n = length(x)))
myData <- do.call(data.frame, myData)
myData$se <- myData$x.sd / sqrt(myData$x.n)
colnames(myData) <- c("Flyer_region", "mean", "sd", "n", "se")
myData$names <- c(paste(myData$Flyer_region, "Flyer_region"))
p <- ggplot(data = myData, aes(x = factor(Flyer_region), y = mean))
p + geom_bar(stat = "identity",
             position = position_dodge(0.9), fill="steelblue") +
  geom_errorbar(aes(ymax = mean + 2*se,
                    ymin = mean - 2*se), position = position_dodge
(0.9),
                width = 0.25) +
 labs(x = "Flyer_region", y = "Conversion rate") +
  ggtitle("Conversion rate by Flyer_region") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  geom_text(position = position_dodge(width= .9), aes(y = mean,
label = paste(format(mean*100, digits = 2, nsmall = 2), "%",
sep = "")), size = 4, vjust = 5)
# Plot of banner ad + flyer and purchase likelihood
myData <- aggregate(consumers$Purchase,
                    by = list(Firm_banner = consumers$Firm_banner,
Flyer_region = consumers$Flyer_region),
                     FUN = function(x) c(mean = mean(x), sd = sd(x),
                                         n = length(x)))
myData <- do.call(data.frame, myData)
```

```
myData$se <- myData$x.sd / sqrt(myData$x.n)
colnames(myData) <- c("Firm_banner", "Flyer_region", "mean", "sd",
"n", "se")
myData$names <- c(paste(myData$Firm_banner, "Firm_banner /",
                        myData$Flyer_region, " Flyer_region"))
p <- ggplot(data = myData, aes(x = factor(Firm_banner), y = mean,
                               fill = factor(Flyer_region)))
p + geom_bar(stat = "identity",
             position = position_dodge(0.9)) +
  geom_errorbar(aes(ymax = mean + 2*se,
                    ymin = mean - 2*se), position = position_dodge
(0.9),
              width = 0.25) +
 labs(x = "Firm_banner", y = "Conversion rate") +
  ggtitle("Conversion rate by firm_banner and flyer_region") +
  scale_fill_discrete(name = "Flyer region") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  geom_text(position = position_dodge(width= .9), aes(y = mean,
label = paste(format(mean*100, digits = 2, nsmall = 2), "%", sep =
"")), size = 4, vjust = 5)
```

The left plot of Fig. 5 shows for banner advertising the difference between consumers in the treatment group (i.e., those in the firm ad group) versus the control group (i.e., those in the charity ad group). We can see that the conversion rate is 27.77% for the treatment group, which is significantly higher than the control group's 16.88% conversion rate. Banner advertising does thus seem to be very effective, increasing the purchase likelihood with 10.89% points. This value can be used as an input to perform ROI calculations for banner advertising. If we assume the average consumer gets 2.16 banner impressions (based on the mean number of banner impressions with and without a click, see Table 1), the cost per mile (CPM, i.e., the costs of 1,000 banner impressions) is $10, and the value of a conversion (i.e., profit before marketing costs) is $1, then the ROI would be:

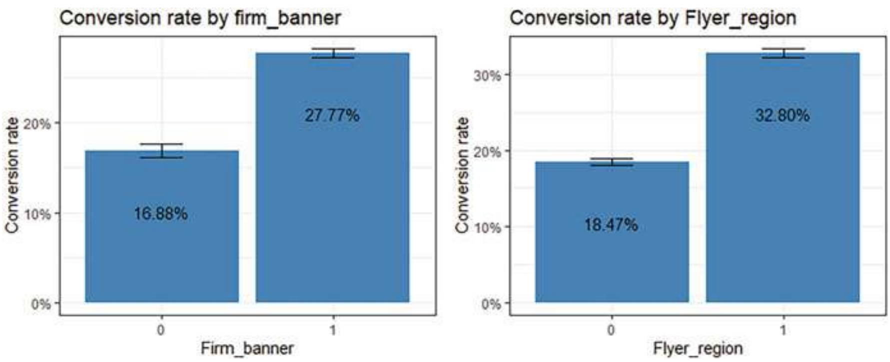

Fig. 5 Banner and flyer effectiveness visualized

$$\boldsymbol{ROI} = \frac{\boldsymbol{\Delta coversion{\cdot}profit} - \boldsymbol{impressions{\cdot}}\frac{\boldsymbol{CPM}}{\mathbf{1000}}}{\boldsymbol{impressions{\cdot}}\frac{\boldsymbol{CPM}}{\mathbf{1000}}}\cdot\mathbf{100}\%$$

$$= \mathbf{0.1089{\cdot}1} - \mathbf{2.16}\cdot\frac{\frac{\mathbf{10}}{\mathbf{1000}}}{\mathbf{2.16}\cdot\frac{\mathbf{10}}{\mathbf{1000{\cdot}100}\%=\mathbf{404.17}\%}} \qquad (1)$$

Where "$impressions{\cdot}\frac{CPM}{1000}$" are the costs per consumer for targeting him or her with banner advertising, while "$\Delta coversion \cdot profit$" is the improvement in profitability before taking the marketing costs into account. In this case, we can see that the ROI is 404.17%, i.e., banner advertising is very profitable on average. Note that this is a generated dataset and that in reality the effectiveness of banner advertising is usually much smaller compared to this example. Hoban and Bucklin (2015) have found an uplift of between 0.065 and 0.985% points in the purchase likelihood, depending on the user segment.

The right-hand graph of Fig. 5 shows that receiving a flyer also significantly increases the purchase likelihood from 18.47% to 32.80%. The effect of receiving a flyer is thus even stronger compared to the effectiveness of banner advertising, although a flyer might also be much more expensive than banner impressions. When we know the cost of distributing a flyer, we can again use this to calculate the ROI of flyers and decide to invest in it.

Figure 6 shows the synergy effect of banner advertising and the flyer. When the firm does not distribute a flyer, being in the firm's banner group increases the purchase likelihood from 13.61–20.26%, i.e., an increase in conversion of 6.65% points. When the firm distributes a flyer, being in the firm's banner group increases the purchase likelihood from 19.70–35.91%, i.e., an increase in conversion of 16.21% points. In other words, when the firm distributes a flyer, the banner becomes more effective. Such positive synergy between on- and offline advertising is in line with some studies, e.g., Lesscher et al. (2021) and Pauwels et al. (2016a).

To formally test the direct and synergy effects, we can also use a logistic regression model, as done with the following R-script.

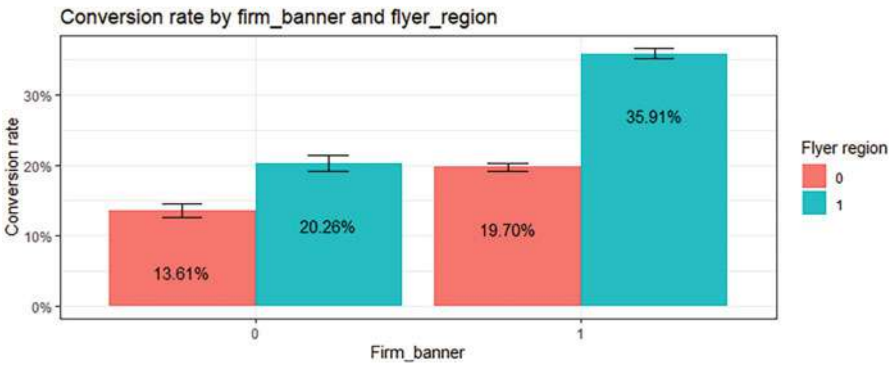

Fig. 6 Banner and flyer synergy visualized

```
# Load package to make an output table of all models
library(sjPlot)
library(sjmisc)
library(sjlabelled)
# Logistic regression model of banner ad treatment + flyer and
purchase likelihood
model1 <- glm(Purchase ~ Firm_banner + Flyer_region,
data=consumers, family=binomial)
model2 <- glm(Purchase ~ Firm_banner*Flyer_region, data=consumers,
family=binomial)
tab_model(model1, model2, transform = NULL, collapse.ci = TRUE,
p.style = "stars")
```

The first two columns of Table 4, i.e., Model 3 and 4, show the output of these models. Model 3 shows that being in the firm's banner group and living in a flyer region significantly impact the purchase likelihood. We can also see that the parameter for "flyer region" is larger in value, i.e., we can conclude that the flyer is most effective in increasing the purchase likelihood. These findings are all in line with what we could see in Fig. 5.

When we look at Model 4 from Table 4, we can also see the interaction effect. The interaction, i.e., the parameter for "Firm_banner * Flyer_region," is statistically significant and positive, meaning that there is indeed a positive synergy, i.e., the combined effect of the two channels is larger than the sum of the two channels' individual effects. This finding is in line with what we have observed in Fig. 6.

To interpret these parameters and the synergy effect, let us take an example. By looking at Model 4 in Table 4, we can observe that being in the firm banner (i.e., treatment) group positively impacts purchase. The parameter of 0.44 indicates that the odds of purchasing are ([exp(0.44)-1]*100%) 55.3% higher for consumers in the firm ad group than consumers in the charity ad group, provided that they did not receive a flyer. If they did receive a flyer, then the log-odds increases with 0.35 (i.e., the parameter for the interaction effect), meaning that the odds of purchasing are ([exp(0.44 + 0.35)-1]*100%) 120.3% higher for consumers in the firm ad group than consumers in the charity ad group, provided that they do receive a flyer. For more details on interpreting the parameters of a logistic regression model and how to recalculate this to probabilities, please check Chapter 8 of Leeflang et al. (2015).

We can easily extend the estimated models, for instance, to find out in which situations a marketing instrument is more or less effective. Let us, for this part, only focus on banner advertising. One assumption would be that the higher purchase likelihood due to the banner ad will only occur for consumers who have actually seen the banner, i.e., the number of impressions should be above zero to have an effect. For consumers who have not seen the banner, being randomly allocated in the firm ad group or charity ad group should not make a difference in the purchase likelihood. We can test this assumption with the following R-script.

```
# Create some new variables
consumers$Banner_exposures <- consumers$Banner_no_click +
consumers$Banner_click
```

Table 4 Logistic regression output based on randomized field experiments

	Model 3	Model 4	Model 5	Model 6	Model 7
	Purchase	Purchase	Purchase	Purchase	Purchase
Predictors	*Log-Odds*	*Log-Odds*	*Log-Odds*	*Log-Odds*	*Log-Odds*
(Intercept)	−2.02 ***(−2.08 – −1.97)	−1.85 ***(−1.93 – −1.77)	−1.54 ***(−1.64 – −1.45)	−3.79 ***(−4.20 – −3.42)	−1.65 ***(−1.73 – −1.57)
Firm_banner	0.65 *** (0.59–0.71)	0.44 *** (0.36–0.53)	−0.01(−0.12–0.09)	−0.01(−0.43–0.44)	−0.04(−0.13–0.05)
Flyer_region	0.77 *** (0.73–0.82)	0.48 *** (0.37–0.58)			
Firm_banner * Flyer_region		0.35 *** (0.23–0.46)			
Banner_seen			−0.07(−0.19–0.04)	0.29(−0.13–0.74)	
Firm_banner * Banner_seen			0.90 *** (0.77–1.02)	2.12 *** (1.62–2.58)	
Existing_customer				2.80 *** (2.42–3.23)	
Firm_banner * Existing_customer				−0.03(−0.50–0.40)	
Banner_seen * Existing_customer				0.04(−0.43–0.47)	
(Firm_banner * Banner_seen) * Existing_customer				−1.49 ***(−1.97 – −0.98)	
Banner_exposures +1 [log]					0.06(−0.01–0.13)
Firm_banner * Banner_exposures +1 [log]					0.70 *** (0.62–0.78)
Observations	50,000	50,000	50,000	50,000	50,000
R^2 Tjur	0.038	0.039	0.030	0.128	0.062

* *$p < 0.05$* ** *$p < 0.01$* *** *$p < 0.001$*

```
consumers$Banner_seen <- ifelse(consumers$Banner_exposures==0,0,1)
# Plot of banner ad group + banner ad seen and purchase likelihood
myData <- aggregate(consumers$Purchase,
                    by = list(Banner_seen = consumers$Banner_seen,
Firm_banner = consumers$Firm_banner),
                    FUN = function(x) c(mean = mean(x), sd = sd(x),
                                       n = length(x)))
myData <- do.call(data.frame, myData)
myData$se <- myData$x.sd / sqrt(myData$x.n)
colnames(myData) <- c("Banner_seen", "Firm_banner", "mean", "sd",
"n", "se")
myData$names <- c(paste(myData$Banner_seen, "Banner_seen /",
                        myData$Firm_banner, " Firm_banner"))
p <- ggplot(data = myData, aes(x = factor(Banner_seen), y = mean,
                               fill = factor(Firm_banner)))
p + geom_bar(stat = "identity",
             position = position_dodge(0.9)) +
  geom_errorbar(aes(ymax = mean + 2*se,
                    ymin = mean - 2*se), position = position_dodge
(0.9),
                width = 0.25) +
  labs(x = "Banner_seen", y = "Conversion rate") +
  ggtitle("Conversion rate by Banner_seen and Firm_banner") +
  scale_fill_discrete(name = "Firm_banner") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  geom_text(position = position_dodge(width= .9), aes(y = mean,
label = paste(format(mean*100, digits = 2, nsmall = 2), "%", sep =
"")), size = 4, vjust = 5)
# Logistic regression model of banner ad treatment + banner ad seen
and purchase likelihood
model3 <- glm(Purchase ~ Firm_banner + Banner_seen +
Firm_banner*Banner_seen, data=consumers, family=binomial)
tab_model(model1, model2, model3, transform = NULL, collapse.ci =
TRUE, p.style = "stars")
```

As the two bars on the left in Fig. 7 show, there is no substantial difference between the control and treatment groups if there were no banner ad impressions. In both cases, the purchase likelihood is just over 17%. This nonsignificant difference is indeed in line with what we would expect; if there is no ad impression, there should be no difference between the firm and charity ad groups. Investigating this is also a good test if the randomization of the experiment worked; if the difference was significant, this would indicate that something might have gone wrong with the randomization.

The two bars on the right of Fig. 7 show a substantial difference between the two groups, provided that there was at least one banner ad exposure. Consumers in the control group who saw the charity banner at least once have a purchase likelihood of 16.56%. Consumers in the treatment group who saw the firm's banner at least once have a purchase likelihood of 32.46%. Being exposed to the banner at least once does have a positive effect since it increases the purchase likelihood by 15.90% points.

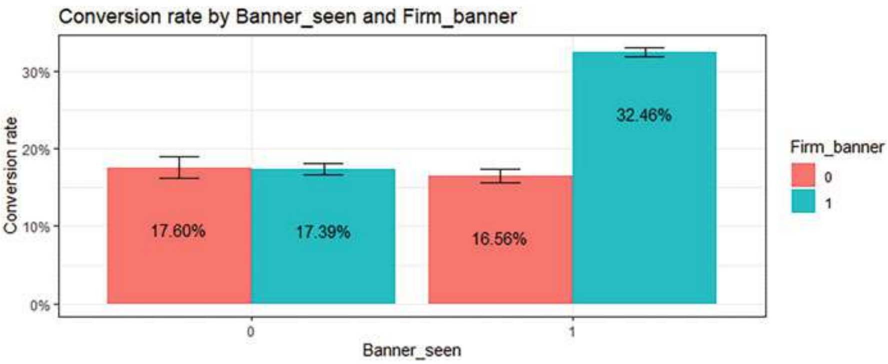

Fig. 7 Banner effectiveness when the banner is seen visualized

The regression output, presented in Model 5 of Table 4, confirms what we see in Fig. 7. The parameter "Firm_banner" is insignificant, which indicates that when the number of ad exposures is zero, the treatment and control groups do not differ in their purchase likelihood. The parameter for "Banner_seen" is also insignificant, indicating no difference in the purchase likelihood between the consumers who did or did not see the charity ad for the control group. This insignificance aligns with what we might expect since the charity ad is irrelevant and should not cause sales at the firm. However, this parameter could, in theory, be significant since it does capture the difference between being more active online (on websites where the campaign is running), i.e., it captures the confounding effects of being exposed to a banner ad.

The parameter for the interaction effect between "Firm_banner" and "Banner_seen" captures the causal effect of seeing the banner ad of the firm. In Model 5 of Table 4, we can see that this interaction is significant and positive, indicating that consumers who saw the firm's banner ad are more likely to purchase due to the exposure to the firm's ad.

Now that we know that the banner ad is, on average, effective in generating purchases, we might want to find out how the effects differ between consumers. For this, we can add additional interaction effects to our regression equation. To illustrate this, we can investigate if the banner works better for potential or existing customers by running the following R-script.

```
# Plot of banner ad treatment + banner ad seen for new/potential
customers
myData <- consumers[consumers$Existing_customer==0,]
myData <- aggregate(myData$Purchase,
                    by = list(Banner_seen = myData$Banner_seen,
Firm_banner = myData$Firm_banner),
                   FUN = function(x) c(mean = mean(x), sd = sd(x),
                                        n = length(x)))
myData <- do.call(data.frame, myData)
myData$se <- myData$x.sd / sqrt(myData$x.n)
```

```
colnames(myData) <- c("Banner_seen", "Firm_banner", "mean", "sd",
"n", "se")
myData$names <- c(paste(myData$Banner_seen, "Banner_seen /",
                        myData$Firm_banner, " Firm_banner"))
p <- ggplot(data = myData, aes(x = factor(Banner_seen), y = mean,
                               fill = factor(Firm_banner)))
p + geom_bar(stat = "identity",
             position = position_dodge(0.9)) +
  geom_errorbar(aes(ymax = mean + 2*se,
                    ymin = mean - 2*se), position = position_dodge
(0.9),
                    width = 0.25) +
  labs(x = "Banner_seen", y = "Conversion rate") +
  ggtitle("Conversion rate by Banner_seen and Firm_banner for
potential/new customers") +
  scale_fill_discrete(name = "Firm_banner") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  geom_text(position = position_dodge(width= .9), aes(y = mean,
label = paste(format(mean*100, digits = 2, nsmall = 2), "%", sep =
"")), size = 4, vjust = 2)
# Plot of banner ad treatment + banner ad seen for existing
customers
myData <- consumers[consumers$Existing_customer==1,]
myData <- aggregate(myData$Purchase,
                    by = list(Banner_seen = myData$Banner_seen,
Firm_banner = myData$Firm_banner),
                    FUN = function(x) c(mean = mean(x), sd = sd(x),
                                        n = length(x)))
myData <- do.call(data.frame, myData)
myData$se <- myData$x.sd / sqrt(myData$x.n)
colnames(myData) <- c("Banner_seen", "Firm_banner", "mean", "sd",
"n", "se")
myData$names <- c(paste(myData$Banner_seen, "Banner_seen /",
                        myData$Firm_banner, " Firm_banner"))
p <- ggplot(data = myData, aes(x = factor(Banner_seen), y = mean,
                               fill = factor(Firm_banner)))
p + geom_bar(stat = "identity",
            position = position_dodge(0.9)) +
  geom_errorbar(aes(ymax = mean + 2*se,
                    ymin = mean - 2*se), position = position_dodge(0.9),
                width = 0.25) +
  labs(x = "Banner_seen", y = "Conversion rate") +
  ggtitle("Conversion rate by Banner_seen and Firm_banner for
existing customers") +
  scale_fill_discrete(name = "Firm_banner") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  geom_text(position = position_dodge(width= .9), aes(y = mean,
label = paste(format(mean*100, digits = 2, nsmall = 2), "%", sep =
"")), size = 4, vjust = 5)
# Logistic regression model of banner ad treatment + banner ad seen
and purchase likelihood
model4 <- glm(Purchase ~ Firm_banner*Banner_seen*Existing_customer,
data=consumers, family=binomial)
```

```
tab_model(model1, model2, model3, model4, transform = NULL,
collapse.ci = TRUE, p.style = "stars")
```

The top graph of Fig. 8 shows the impact of banner advertising for new (i.e., potential) customers. What can be seen is that when there are no ad impressions (i.e., no treatment, the two left bars), the control and treatment group do not differ. This makes sense, since the groups do not differ from each other. When the ad is seen, i.e., there is at least one ad impression, the two groups differ from each other. When the new/potential customer sees the charity banner, the purchase likelihood is 2.93%, and with the firm banner this is 19.85%. The causal impact of banner advertising is thus 16.92% points increase in purchase likelihood.

The bottom graph of Fig. 8 shows the effects for existing customers. We can see that they have a higher purchase likelihood than new customers, even when they do not see the banner. This makes sense, since existing customers have already made a purchase before and are more likely to come back compared to someone who has not made a purchase before. We also can see that there is again no significant difference between the control group and the treatment group when the banner ad is not seen, which again is what we would expect. If the control group sees the charity banner, the purchase likelihood goes up from 27.16% to 34.11%. This significant increase

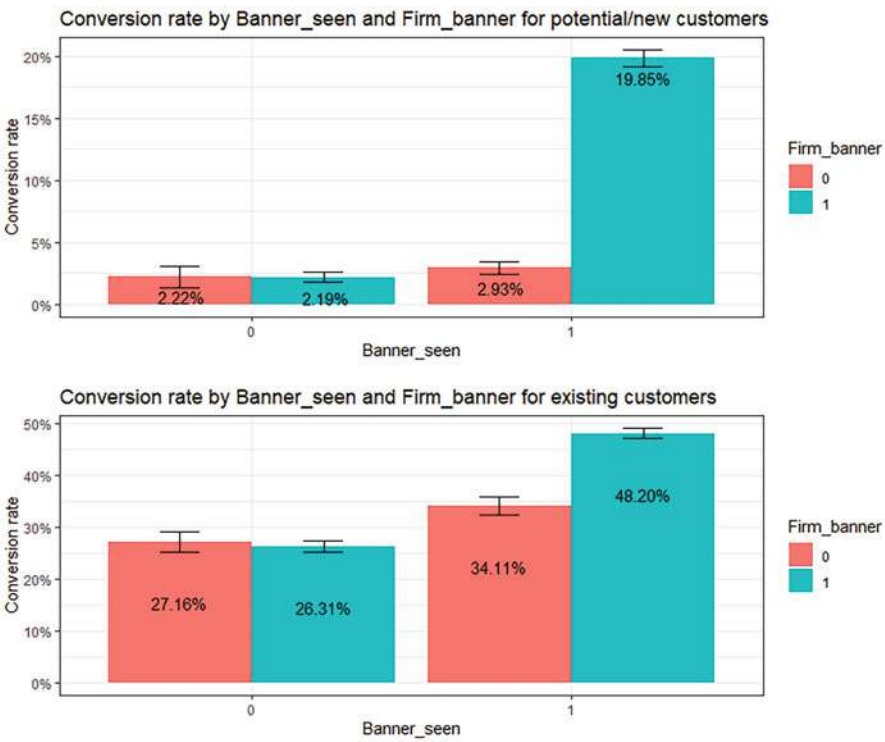

Fig. 8 Banner effectiveness difference between potential and existing customers visualized

can be explained by the fact that the ones seeing the charity ad are more online and they visit the website where the campaign is running, i.e., because they are more active, they might be more likely to purchase. This part of the effect is however not driven by the ad itself. For the treatment group, we do however see that the growth in the purchase likelihood is significantly larger when they see the banner ad at least once. This extra increase of ([48.20–26.31%] – [34.11–27.16%]) 14.94% points is the causal effect of the banner ad for existing customers. We can see that this effect of the banner for existing customers is smaller (a bit smaller in percentage points, and a lot smaller in percentages) than for new customers, where the purchase likelihood jumps up when exposed to the firm's banner.

To test the significance of the effects, we can look at Model 6 in Table 4. The parameter for "Firm_banner" and its interaction with "Existing_customer" are both insignificant, meaning that the randomization has no significant effect on the purchase likelihood provided that there are no ad impressions. The parameter for "Banner_seen" is insignificant, which indicates that seeing the charity banner does not increase the purchase likelihood. The parameter for "Existing_customer" is statistically significant, indicating that existing customers are more likely to conduct a purchase.

The interaction between "Firm_banner" and "Banner_seen" is significant, as expected, which in this case means that for potential customers, seeing the firm's banner at least once has a positive impact on the purchase likelihood. The three-way interaction, which also includes "Existing_customer," is significant and negative; this means that, for existing customers, exposures to the firm's banner are significantly *less* effective than for potential customers. The banner is still effective for existing customers since the two-way interaction minus the three-way interaction remains positive (i.e., 2.12–1.49 = 0.63) (One can also test the significance of this.63, by changing the "Existing_customer" dummy in a "New_customer" dummy in Model 6 (i.e., this dummy is the opposite of the "Existing_customer" dummy).). This positive but smaller impact of banner advertising for existing customers aligns with what we saw in Fig. 8. If the firm wants to target banners, it is thus more effective to target the potential customers instead of the existing customers.

Finally, with the following R-script, we can also test the impact of having more ad exposures.

```
# Plot of banner ad treatment + number of exposures and purchase
likelihood
myData <- consumers
myData$Banner_exposures <- ifelse(myData$Banner_exposures > 6, 6,
myData$Banner_exposures)
myData$Firm_banner <- as.factor(myData$Firm_banner)
myData <- aggregate(myData$Purchase,
                    by = list(Firm_banner = myData$Firm_banner,
Banner_exposures = myData$Banner_exposures),
                    FUN = function(x) c(mean = mean(x), sd = sd(x),
                                        n = length(x)))
myData <- do.call(data.frame, myData)
```

```
myData$se <- myData$x.sd / sqrt(myData$x.n)
colnames(myData) <- c("Firm_banner", "Banner_exposures", "mean",
"sd", "n", "se")
myData$names <- c(paste(myData$Firm_banner, "Firm_banner /",
                        myData$Banner_exposures, " Banner_exposures"))
p <- ggplot(data = myData, aes(x = Banner_exposures, y = mean,
                               group = as.factor(Firm_banner), color =
Firm_banner))
p + geom_line() +
  geom_point() +
  geom_errorbar(aes(ymin=mean - 2*se, ymax=mean + 2*se), width=.2,
                position=position_dodge(0.05)) +
  labs(x = "Number of banner exposures", y = "Conversion rate") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  ggtitle("Conversion rate by number of ad exposures") +
geom_text(position = position_dodge(width= .9), aes(y = mean, label
= paste(format(mean*100, digits = 2, nsmall = 2), "%", sep = "")), size =
4, vjust = -1)
# Logistic regression model of banner ad treatment + number of
exposures and purchase likelihood
model5 <- glm(Purchase ~ Firm_banner*log(Banner_exposures+1),
data=consumers, family=binomial)
tab_model(model1, model2, model3, model4, model5, transform = NULL,
collapse.ci = TRUE, p.style = "stars")
```

The bottom (red) line in Fig. 9 shows the impact of zero (first bar) up to six-plus (seventh bar, six and more ad impressions are aggregated due to the size of the graph and the small amount of observations with many ad impressions) ad impressions for the control group (i.e., those in the charity ad group). We do not see much change in purchase likelihood here, e.g., the purchase likelihood is 17.60% with zero ad impressions, and this increases to 19.00% with six or more impressions, all falling within the same confidence interval.

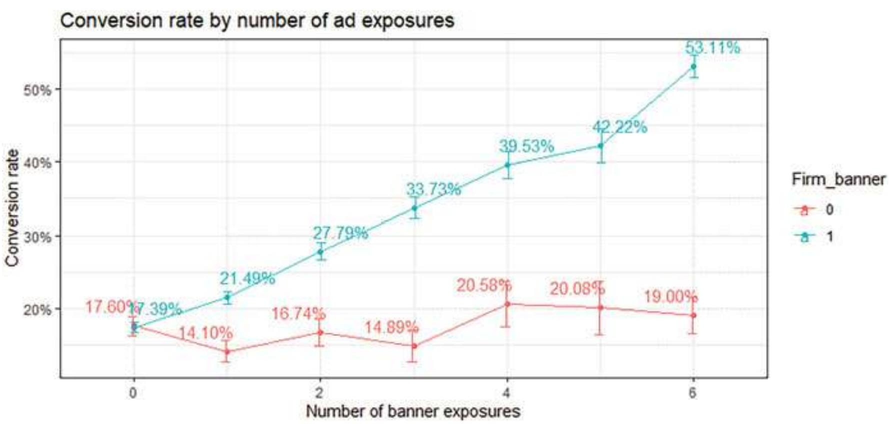

Fig. 9 Banner effectiveness depending on the impression amount visualized

The top (blue) line in Fig. 9 shows the impact of having more banner ad exposures for consumers in the treatment group, i.e., those exposed to the firm's ad. Here we can see a clear positive impact of having more ad exposures on the purchase likelihood. The difference between the treatment group and the control group is the *incremental effect* of the firm's banner ad. The incremental effect seems to increase with the number of banners, indicating that a higher number of banner ads increases conversion further. When the incremental effect would stabilize, this would signal that an additional banner impression is not worthwhile. This kind of analysis can help to investigate the optimal amount of banner impressions. For more details, also see the study by Hoban and Bucklin (2015). The timing of advertising can be determined similarly, as Braun and Moe (2013) demonstrate in their study. Försch and De Haan (2018) study a combination of frequency and timing of advertising.

Model 7 in Table 4 shows the parameter estimates of the frequency effect. Again, the parameter for "Firm_banner" is insignificant, indicating that the treatment and control groups do not differ in their purchase likelihood when there are zero ad exposures, which aligns with what we expect. Furthermore, "log(Banner_exposures + 1)" is insignificant, indicating that a higher number of exposures to the charity ad does not significantly impact the purchase likelihood. This parameter captures the confounding effects. The interaction effect captures the causal impact of the firm's banner ad and is, therefore, the parameter of primary interest. This interaction effect is statistically significant and positive, meaning that the higher number of exposures to the firm's banner causes an increase in the likelihood to purchase. These results are again in line with Fig. 9.

The models discussed in this section are easily adaptable. It is, for instance, possible to test if receiving a flyer also influences the impact of the number of ad exposures or if the effects of the number of ad impressions are nonlinear. If there is information on the websites where the campaign is running, it is also possible to investigate if an ad exposure from website A has a different impact on the purchase likelihood than an ad exposure from website B. Similarly, it is possible to test the impact of different ad creatives by having different groups of people exposed to different ad creatives. Such a setup is also called *A/B testing*. It is furthermore possible to test other channels with a randomized experiment.

An advantage of a randomized field experiment is that we can be sure of the causality, provided there is true randomization, and there are thus no other differences between the two groups. A downside is that a randomized field experiment at the consumer level is not possible for all channels. Luckily this can be overcome by conducting the experiment at a more aggregate level, e.g., varying between regions or conducting changes over time, as explained in the section "Attribution with Aggregate-Level (Quasi-)Experimental Data" of this chapter. Randomization is impossible for some other channels since consumers decide which channel to use, i.e., the channel exposure cannot be randomized. In such cases, other techniques have to be applied, as discussed in the next section. Another downside of conducting field experiments is that it can be challenging and time consuming to do this for all channels.

Furthermore, when investigating the interaction effects between channels, e.g., if online and offline advertising might make each other stronger, a more complicated experimental design is needed. Such a design is doable for two channels (e.g., via a two-by-two experimental design, as done in our example in this section), but this will be practically impossible for a large number of channels. In such cases, different methods are appropriate, as explained in the "Exploring Paths in Conversion (Markov Chain)" section of this chapter.

Additional Topics for Individual-Level Attribution

This section discusses some more advanced attribution methods suitable for individual-level data. In some cases, conducting field experiments is not possible or one is interested in the customer journey that is affected by a certain channel, without all potential channels being the subject of an experiment. In these cases, it is still possible to conduct attribution and find the causal impacts of specific (combinations of) channels, although the methods become more complex and there might be less certainty that all criteria of causality do indeed hold. We will start with propensity score matching, which is suitable when the channel is not subject to an experiment but one still wants to explore the impact as a quasi-experiment. The second method uses Markov chains to explore how different channels affect the journey, and the outcome of a journey, which provides more detailed insights compared to looking at (combinations of) channels in isolation. We will end this section with a discussion of other methods and further procedures for individual-level attribution.

Attribution with Individual-Level Quasi-Experimental Data (Propensity Score Matching)

In some cases, a random experiment is not possible, feasible, or desirable, but we still want to draw causal conclusions. If we, for instance, observe that customers who signed up for a loyalty program are buying more frequently, can we then conclude that this is due to the loyalty program? Or are customers who purchase more also more likely to join a loyalty program? Or is it a combination of the two? With a loyalty program, it is impossible to randomly sign-up people since customers have to give their consent to sign up, and it is undesirable to deny random customers' membership to such a program. A random experiment is thus impossible, infeasible, or undesirable to do in such a situation. For other channels, e.g., email, it is possible to not send out an email to a random group of customers to test the effectiveness of this channel, but this might be undesirable, for instance, because the customers expect to receive the email for which they have signed up.

In the cases described above, a feasible alternative is to conduct a quasi-experiment, which is possible via *propensity score matching* (PSM). With PSM, we link consumers to each other who are equally likely to get exposed to a channel, but due to random chance, one did get exposed, and the other one did not get exposed. If we do this for many pairs of consumers, we get two groups of which both had an equal chance of exposure, but one had the exposure (i.e., the treatment group) and the other did not (i.e., the control group). If we have these two groups, we can then investigate the channel the same way as with a randomized field experiment, as described in the previous section. We can thus test if the two groups differ in terms of their outcome (e.g., purchase likelihood), which would provide us the causal impact of the channel.

Multiple studies use PSM to conduct attribution. One example is the study by De Haan et al. (2018), who have instigated what impacts device switching in an online customer journey (e.g., a person starts looking for information on a smartphone and then switches to a laptop). Since device usage and device switching is not a random decision and is not randomizable with an experiment, De Haan et al. (2018) matched sessions that were equally likely to have a specific device switch, but in some cases, the switch did occur while in other cases this switch did not occur. De Haan et al. (2018) find that switching from a mobile device to a nonmobile device increases the conversion probability, especially if the product the customer is investigating is riskier (e.g., the product category has a higher perceived risk, the customer has less experience buying in the category, or the product is relatively expensive).

To conduct PSM, we have to go through the following five steps (see Chapter 1 of Pan and Bai 2015 for more details):

1. Estimate the likelihood of being treated (e.g., subscribing to an email newsletter)
2. Check for the overlap in the propensity scores (e.g., likelihood of subscribing) of treatment and control group
3. Match observations from the treatment group with observations from the control group who have a similar propensity score
4. Verify the balance of covariates to check if the matching was successful
5. Conduct multivariate analysis based on the matched sample, similar as we have done in the "Attribution Modeling Process with Experimental Data" section

A logistic regression model can estimate the treatment likelihood, with treatment (e.g., receiving an email newsletter) as a dependent variable and the drivers of treatment as independent variables. The independent variables should be exogenous, which means that the treatment does not influence these variables. The number of website visits might be a good predictor for email subscription, but the email subscription might also drive website visits. Rubin (2001) recommends selecting the independent variables based on theory and prior research.

Demographic variables might work in our case since this might be related to the probability of signing up for the email service, and the email does not influence people's demographics. Unfortunately, however, we do not have demographic variables in our dataset. Since the email service started on the first of January (see Table 1), and we have the relationship length and the CLV up until the first of January, we can use these latter two variables as independent variables; since these two variables provide information before the email service's introduction, the email cannot have caused changes in the relationship length or CLV.

For the PSM, we will be using the "MatchIt" package in R (Ho et al. 2021). The following R-script estimates the propensity scores and matches the consumers who have a similar propensity to sign up for the email service.

```
# Load packages
library(MatchIt)
library(cobalt)
# Only include existing customers, since they are the only ones
signing up to the email (for potential customers, email sign up is
zero in our dataset)
existing_customers <- consumers[consumers$Existing_customer==1,]
# Log transform CLV, since it is highly skewed
existing_customers$Log_CLV <- log(existing_customers$CLV)
# Conduct matching of consumers who did and did not subscribe to the
email (i.e. step 1 and 3 of the PSM process). In this example, the
Email_group membership is predicted using the Relation_length and
the log of CLV. The matching is automatically performed with the
code.
m.out <- matchit(Email_group ~ Relation_length + Log_CLV,
                 data=existing_customers, caliper=0.05)
# Investigate the PSM outcome
summary(m.out, standardize = TRUE)
love.plot(m.out)
```

In the R-script, the term "caliper" indicates the number of standard deviations the matched consumers' propensity score can maximum be apart. If we set this value higher, this would result in consumers who being less similar to each other can be matched. This higher value does result in more matched consumers, but the difference between the matched consumers can get larger, which reduces the appropriateness of PSM since we want to create comparable groups. If we set the value of "caliper" to zero, only consumers with identical propensity scores are matched, resulting in more similar samples, but this can substantially reduce the sample size since there might be few (or sometimes even non) exact matches. Determining the appropriate value for "caliper" can be achieved by trial and error, e.g., by checking what happens to the outcome of step 4; if the samples are unbalanced, the value of "caliper" should decrease; if the sample is very well balanced, but the sample sizes are too small, one can test if a higher value of "caliper" also still works.

Figure 10 shows the output of the results of the matching. As shown at the bottom of this Figure, we have 11,727 observations in the control group (i.e., those who do not receive an email) and 13,304 consumers in the treatment group (i.e., those who do receive an email). After matching, we have 10,257 observations in both groups. At the top of Fig. 10, we see some descriptive statistics. The mean relationship lengths before matching are 32.61 months and 28.04 months for the treatment and control group, respectively, which is equal to a 0.27 standard deviation difference. After matching, this difference drops to 0.02 standard deviations. A similar pattern is visible for the logarithmic of CLV and the "distance" (i.e., the polarity score). As the

```
Call:
matchit(formula = Email_group ~ Relation_length + Log_CLV, data = existing_customers,
    caliper = 0.05)

Summary of Balance for All Data:
                Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean eCDF Max
distance               0.5528        0.5074          0.4538     0.8578    0.1210   0.1822
Relation_length       32.6097       28.0416          0.2669     1.0013    0.0761   0.1192
Log_CLV                7.3907        7.1379          0.3684     0.7584    0.0956   0.1366


Summary of Balance for Matched Data:
                Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean eCDF Max
distance               0.5309        0.5273          0.0356     1.0395    0.0110   0.0235
Relation_length       29.9863       29.5953          0.0228     0.9991    0.0065   0.0169
Log_CLV                7.2835        7.2649          0.0272     1.0224    0.0096   0.0246
                Std. Pair Dist.
distance                 0.0356
Relation_length          0.8579
Log_CLV                  0.7120

Percent Balance Improvement:
                Std. Mean Diff. Var. Ratio eCDF Mean eCDF Max
distance                   92.2       74.7      90.9     87.1
Relation_length            91.4       34.1      91.4     85.9
Log_CLV                    92.6       92.0      90.0     82.0

Sample Sizes:
          Control Treated
All         11727   13304
Matched     10257   10257
Unmatched    1470    3047
Discarded       0       0
```

Fig. 10 R-output of "MatchIt"

"Percent Balance Improvement" section shows, the differences between the means for distance, relation length, and the log of CLV of the two groups drop by 92.2%, 91.4%, and 92.6%, respectively. In other words, after matching, the groups are much more balanced.

Figure 11 visualizes the standard mean differences, as presented in Fig. 10, which clarifies that the matched (adjusted) sample is much more balanced than the unmatched (unadjusted sample). As can be seen, the mean difference between the two groups drops substantially when going from the unmatched sample (red dots on the right-hand side in Fig. 11) to the matched sample (blue dots on the left-hand side in Fig. 11). In the matched sample, the differences are close to zero, which is again what we want to achieve.

Next, we want to investigate steps two and four of the PSM procedure, which we can do with the following R-script.

```
# Plot balance diagnostics (i.e., step 4 of the PSM process)
plot(m.out, type = "jitter", interactive = FALSE)
plot(m.out, type = "qq", interactive = FALSE)
bal.plot(m.out, which = "both")
bal.plot(m.out, var.name = "Relation_length", which = "both")
bal.plot(m.out, var.name = "Log_CLV", which = "both")
```

Figure 12 shows the distribution of the propensity scores for the matched treatment and control groups (i.e., the middle two groups) and the propensity scores of the unmatched consumers (i.e., the top and bottom groups). By looking at this Figure, there seems to be a very similar distribution and a good overlap in the propensity scores of the two matched groups, i.e., we seem to meet the criterion mentioned in the second step of the PSM procedure ("Check for the Overlap in the Propensity Score").

Figure 13 shows the Q-Q plot of the two variables which we used to match the consumers. In this particular plot, we want the observations close to the diagonal line, which indicates that the matched consumers in the treatment and control groups are the same. As can be seen, if we use all consumers (i.e., matched and unmatched), there is quite some deviation from the diagonal line, i.e., the consumers are not very similar. After matching, the observations almost perfectly follow this diagonal line, indicating that the matching worked, and we meet the fourth step of the PSM procedure ("Verify the balance of covariates to check if the matching was successful").

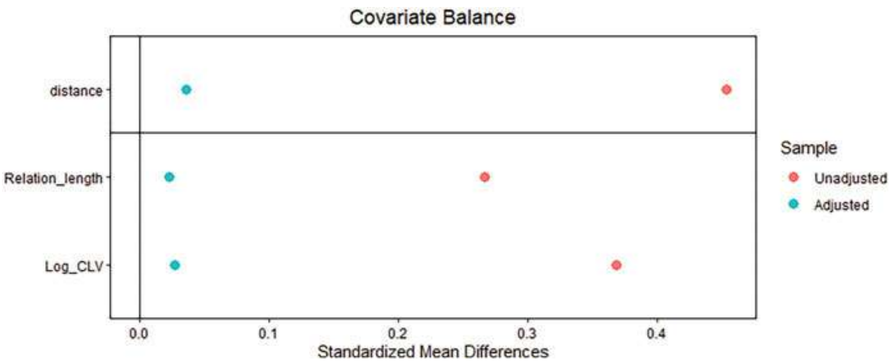

Fig. 11 R-output of covariate balance

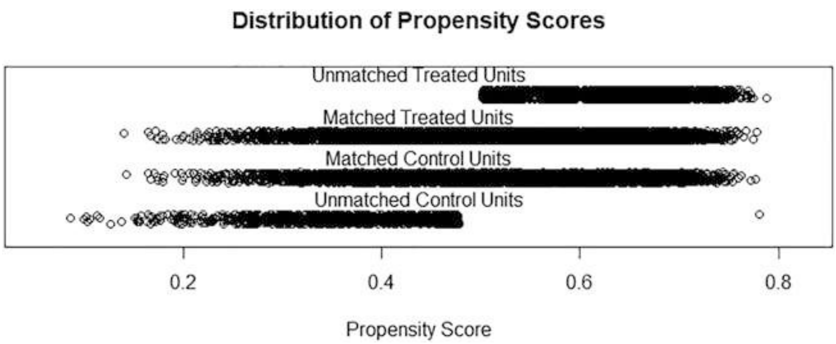

Fig. 12 R-output of propensity scores distribution for matched and unmatched groups

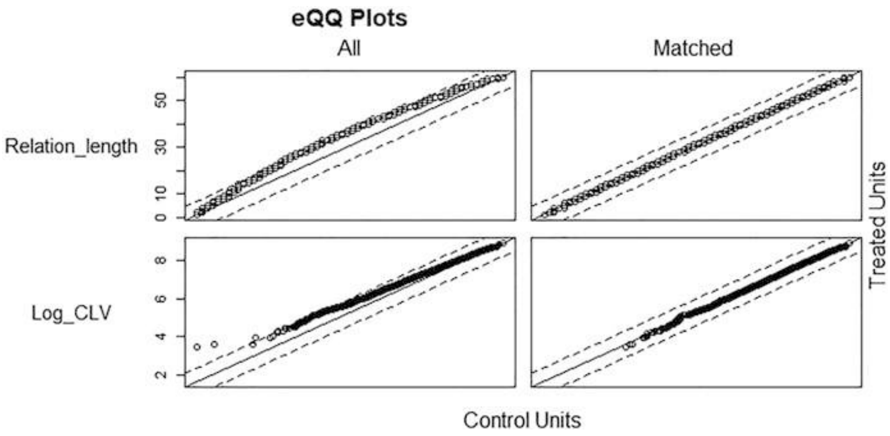

Fig. 13 R-output of QQ-plots for matched and all samples

Figure 14 is somewhat similar to the previous two figures but visualizes the overlap in a density plot. Again, the distributions of the "distance" (i.e., the propensity scores), relation length, and the logarithmic of CLV are not very well overlapping before matching (i.e., the graphs on the left), while after matching, they are very nicely overlapping (i.e., the graphs on the right). Again, we thus seem to meet the criteria of steps two and four of the PSM procedure.

Since the matching seems to have worked, we can use this new matched sample similar to the randomized field experiment discussed in "Attribution Modeling Process with Experimental Data" section of this chapter. For email, we can, for instance, check if (after matching) there is a significant effect on the likelihood to purchase and if this effect differs for different customers, e.g., those who were in the firm's ad group or the charity ad group.

```
# Get a dataframe with the matched data
matched_data <- match.data(m.out)
# Plot of email and purchase likelihood (unmatched)
myData <- aggregate(existing_customers$Purchase,
                   by = list(Email_group = existing_customers
$Email_group),
                  FUN = function(x) c(mean = mean(x), sd = sd(x),
                                      n = length(x)))
myData <- do.call(data.frame, myData)
myData$se <- myData$x.sd / sqrt(myData$x.n)
colnames(myData) <- c("Email_group", "mean", "sd", "n", "se")
myData$names <- c(paste(myData$Email_group, "Email_group"))
p <- ggplot(data = myData, aes(x = factor(Email_group), y = mean))
p + geom_bar(stat = "identity",
            position = position_dodge(0.9), fill="steelblue") +
  geom_errorbar(aes(ymax = mean + 2*se,
                  ymin = mean - 2*se), position = position_dodge
                  (0.9),
```

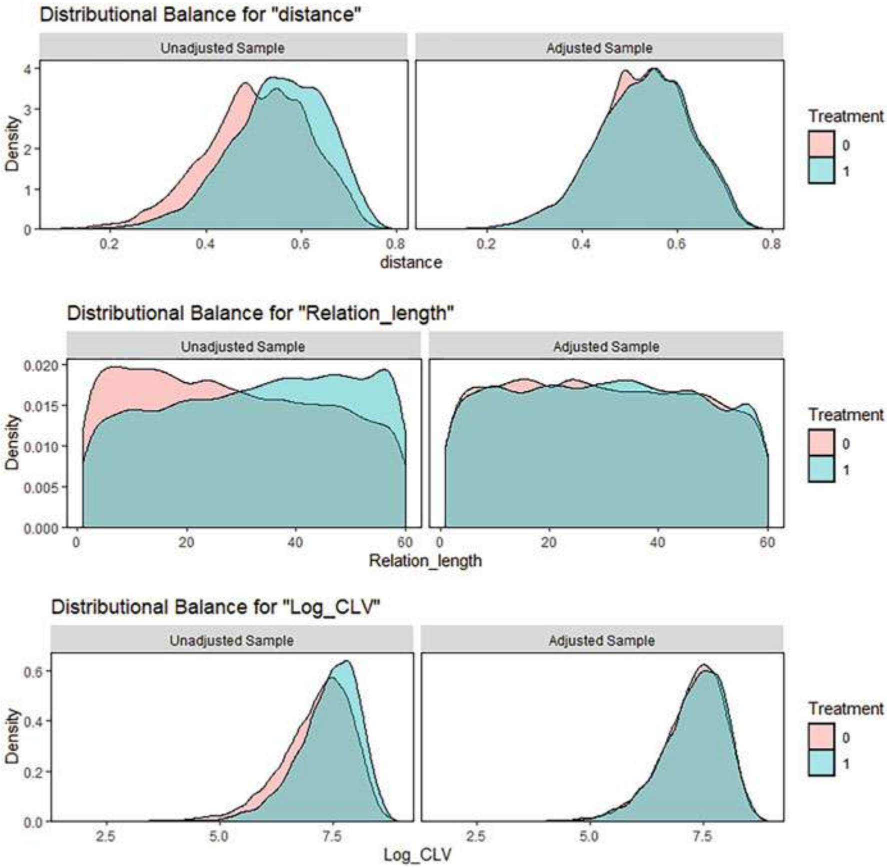

Fig. 14 Distribution balance of "distance" (top), "Relation_length" (middle), and "Log-CLV" (bottom)

```
                    width = 0.25) +
  labs(x = "Email_group", y = "Conversion rate") +
  ggtitle("Conversion rate by Email_group (full sample)") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  geom_text(position = position_dodge(width= .9), aes(y = mean,
label = paste(format(mean*100, digits = 2, nsmall = 2), "%", sep =
"")), size = 4, vjust = 5)
# Plot of email and purchase likelihood (matched)
myData <- aggregate(matched_data$Purchase,
                    by = list(Email_group =
                    matched_data$Email_group),
                    FUN = function(x) c(mean = mean(x), sd = sd(x),
                                        n = length(x)))
myData <- do.call(data.frame, myData)
myData$se <- myData$x.sd / sqrt(myData$x.n)
colnames(myData) <- c("Email_group", "mean", "sd", "n", "se")
```

```
myData$names <- c(paste(myData$Email_group, "Email_group"))
p <- ggplot(data = myData, aes(x = factor(Email_group), y = mean))
p + geom_bar(stat = "identity",
             position = position_dodge(0.9), fill="steelblue") +
  geom_errorbar(aes(ymax = mean + 2*se,
                ymin = mean - 2*se), position = position_dodge(0.9),
                width = 0.25) +
  labs(x = "Email_group", y = "Conversion rate") +
  ggtitle("Conversion rate by Email_group (matched sample)") +
  scale_y_continuous(labels = function(x) paste0(x*100, "%")) +
  geom_text(position = position_dodge(width= .9), aes(y = mean,
label = paste(format(mean*100, digits = 2, nsmall = 2), "%", sep =
"")), size = 4, vjust = 5)
# Load package to make an output table of all models
library(sjPlot)
library(sjmisc)
library(sjlabelled)
# Estimate logistic regression models
model1 <- glm(Purchase ~ Email_group, data=matched_data,
family=binomial)
model2 <- glm(Purchase ~ Email_group, data=existing_customers,
family=binomial)
model3 <- glm(Purchase ~ Email_group + Firm_banner +
Email_group*Firm_banner, data=matched_data, family=binomial)
model4 <- glm(Purchase ~ Email_group + Firm_banner +
Email_group*Firm_banner, data=existing_customers, family=binomial)
tab_model(model1, model2, model3, model4, transform = NULL,
collapse.ci = TRUE, p.style = "stars")
```

Figure 15 shows the purchase likelihood for the full sample (left graph) and the matched sample (right graph). Before matching, the consumers who receive an email have a purchase likelihood of 41.27%, while those who do not receive the email have a purchase likelihood of 34.36%. There is thus a 6.91% point higher purchase likelihood for the consumers who do receive an email. This 6.91% point difference is a combination of the actual causal effect of the email and the confounding effects (e.g., those signing up for an email are already more loyal and have a higher purchase likelihood to begin with). If we look at the matched sample, the difference between

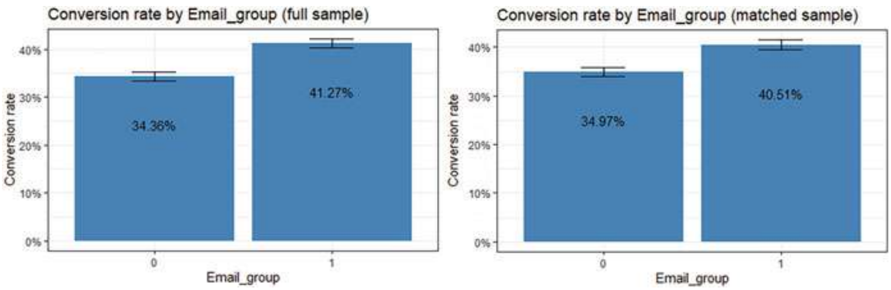

Fig. 15 Email effectiveness for unmatched (left) and matched (right) samples visualized

the consumers who do and those who do not receive the email drops to 5.54% points. This 5.54% point increase in the conversion likelihood is thus the true causal impact of the email.

The output of the four logistic regression models, which replicate Fig. 15, can be seen in Table 5. Model 8 and Model 10 in Table 5 are estimated on the matched sample, while Model 9 and Model 11 use the full sample. Model 9 shows the effect of email is significant and 0.29, meaning that people who receive an email have a ([exp(0.29)-1]*100%) 33.6% higher odds of purchasing than consumers who did not receive an email. Since this second model uses the full sample, it does include some confounds due to the self-selection to subscribe to the email; those who receive an email might be more loyal customers and have a higher likelihood of purchasing. Comparing the two groups is thus a bit comparing apple and oranges, and we cannot interpret the parameter as a causal effect of email on purchase likelihood.

Model 8 in Table 5 uses the matched sample. The two groups are in this sample comparable to each other and only differ in the subscription to the email service. The parameter estimate is now still significant and has a value of 0.24, meaning that people who receive an email have a ([exp(0.24)-1]*100%) 27.1% higher odds of purchasing consumers who did not receive an email; due to the matching we can assume that this effect is causal. Not controlling for the self-selection does thus somewhat overestimate the impact of email on the purchase likelihood, which is in line with what we could see in Fig. 15.

Model 10 and 11 from Table 5 test if there is a synergy effect between email and exposure to the firm banner. Since the interaction effect is insignificant in both models, we cannot conclude that there is a synergy effect.

PSM is a good way to find the effectiveness of a channel in situations where conducting a randomized field experiment is not an option. A challenge here is that the variables used for matching should be exogenous, i.e., the variables should

Table 5 Logistic regression output of matched (columns 1 and 3) and full (columns 2 and 4) samples

	Model 8	Model 9	Model 10	Model 11
	Purchase	Purchase	Purchase	Purchase
Predictors	*Log-Odds*	*Log-Odds*	*Log-Odds*	*Log-Odds*
(Intercept)	−0.62 *** (−0.66 – −0.58)	−0.65 *** (−0.69 – −0.61)	−0.97 *** (−1.06 – −0.87)	−0.99 *** (−1.08 – −0.90)
Email_group	0.24 *** (0.18–0.29)	0.29 *** (0.24–0.35)	0.31 *** (0.18–0.45)	0.38 *** (0.26–0.50)
Firm_banner			0.42 *** (0.32–0.53)	0.42 *** (0.32–0.52)
Email_group * Firm_banner			−0.09 (−0.24–0.06)	−0.10 (−0.24–0.03)
Observations	20,514	25,031	20,514	25,031
R^2 Tjur	0.003	0.005	0.008	0.010

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

correlate with, but not be influenced by, the channel choice. For more details on PSM, including alternative matching procedures, check the book by Pan and Bai (2015).

Exploring Paths in Conversion (Markov Chain)

Another method used for attribution is the Markov chain. With a Markov chain, we can observe through which stages (i.e., channels or touchpoints) a consumer moves in the path to purchase and what the final stage (i.e., conversion or not) is. We can use this Markov chain to estimate the likelihood of encountering a specific touchpoint and the likelihood that a particular path will convert. Anderl et al. (2016) used a Markov approach to map the purchase journey and compare this method to basic attribution methods like last-touch attribution.

Figure 16 provides a simple example of such a graphical Markov chain with three channels. We can see that at the start of the path to purchase, the customer is most likely to use channel 2 or 3, then switches between the three channels, and channel 2 is most likely to lead to a purchase (based on last touch). Such a visualization nicely illustrates how the path to purchase looks like, i.e., how consumers switch between touchpoints and how this leads to an outcome. Instead of visualizing, such a Markov chain can also be represented with a transition matrix, as we will demonstrate later in this section.

For attribution, we can turn off a channel in the Markov chain and simulate what would happen to the path to purchase in terms of touchpoints that the consumer encounters and the outcome of the path to purchase. This is a more sophisticated procedure than touch-based attribution since we consider that the entire path to purchase can change when a channel drops out and that there might be alternative channels that pick up the role of the turned-off channel. Fig. 17 visualizes an example of what happens when we turn off a channel, e.g., the traffic to the other channels changes, and the impact on purchase is also affected.

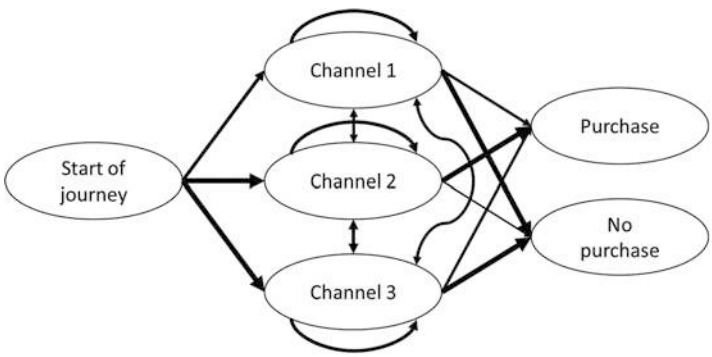

Fig. 16 Simple example Markov chain with three channels

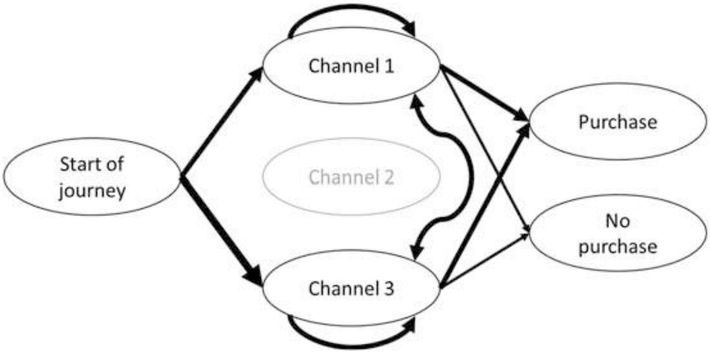

Fig. 17 Simple example Markov chain with one channel turned off

Let us with the following R-script estimate the Markov chain on our dataset. For this, we only use consumers who are in the group that could see the firm's banner. For estimating the Markov chain, we use the R-packages "msm" (Jackson 2019) and "markovchain" (Spedicato 2021).

```
# Add start channel
consumers_journey2 <- consumers_journey[consumers_journey
$Channel_number==1,]
consumers_journey2$Channel_number <- 0
consumers_journey2$Channel_name <- "Start"
consumers_journey2$Channel <- 1
consumers_journey2 <- rbind(consumers_journey, consumers_journey2)
# Add final channel (i.e., conversion or not)
consumers_journey3 <- consumers_journey[consumers_journey
$Channel_number==1,]
touchpoints <- aggregate(consumers_journey2$Channel_number,by=list
(consumers_journey2$Consumer_ID),FUN=max)
consumers_journey3$Channel_number <- touchpoints$x + 1
consumers_journey3$Channel_name <- ifelse(consumers$Purchase==1,
"Purchase", "No Purchase")
consumers_journey3$Channel <- ifelse(consumers$Purchase==1,
10, 11)
consumers_journey2 <- rbind(consumers_journey2,
consumers_journey3)
consumers_journey2 <- consumers_journey2[order(consumers_journey2
$Consumer_ID, consumers_journey2$Channel_number),]
rm(consumers_journey3)
rm(touchpoints)
# Split data up in firm and charity banner group
consumers_journey2_firm <-
 consumers_journey2[consumers_journey2$Firm_banner==1,]
consumers_journey2_charity <-
 consumers_journey2[consumers_journey2$Firm_banner==0,]
# Estimate transition matrix
library(msm)
```

```
markov_chain <- statetable.msm(Channel, Consumer_ID,
data=consumers_journey2_firm)
markov_chain <- prop.table(markov_chain, margin=1)
markov_chain <- as.data.frame(t(markov_chain))
markov_chain <- markov_chain[,3]
markov_chain[100:121] <- c(rep(0,9),1,0,rep(0,10),1)
library(markovchain)
mcjourney <- new("markovchain", states = c
("Start","Banner_impression","Banner_click",
"SEA_product_click","SEA_brand_click","Price_comp_click",
                                 "Email_received", "Email_click",
"Direct_visit",
                                 "Purchase", "No purchase"),
                 transitionMatrix = matrix(data=markov_chain,
                 byrow= TRUE, nrow = 11))
show(mcjourney)
plot(mcjourney, edge.arrow.size=0.25)
```

Figure 18 visualizes the Markov chain, and Fig. 19 provides the transition matrix; both show the same information in a different format. We can observe that there is a 34.72% chance the consumer will first encounter a banner impression. When the consumer in the current state sees a banner impression, there is a 20.74% chance that the next phase will be a direct visit. A direct visit has a 6.80% chance to be followed by a purchase (i.e., conversion) and a 13.04% chance of an unsuccessful end of the path to purchase (and an 80.16% chance of continuing with one of the eight touchpoints).

Such a transition matrix can thus show what the likely next touchpoint the customer will encounter is and how likely the path to purchase will end successfully or unsuccessfully. We can make a simulation based on the transition matrix to investigate what happens after a certain number of stages. The following R-script calculates what the path to purchase looks like after 5 and 50 stages.

```
# Estimate the stage consumers are in after 5 and 50 phases of the
path to purchase, the initial state is the start state
initialstate <- c(1,0,0,0,0,0,0,0,0,0,0)
after5touchpoints <- initialstate * (mcjourney ^ 5)
after5touchpoints
after50touchpoints <- initialstate * (mcjourney ^ 50)
after50touchpoints
```

When running this R-script, we can see that after five stages, i.e., five touchpoints (including the start), there is a 13.28% chance of a conversion, a 35.04% chance of the path to purchase ending with no purchase, and a 51.68% chance that the path to purchase will continue. After 50 stages, there is a 27.76% chance of a conversion, a 72.21% chance of no conversion, and a 0.03% chance of the path to purchase to continue. The 27.76% conversion is similar to the 27.77% conversion we saw for the firm banner group in Fig. 5, i.e., this transition matrix nicely matches reality.

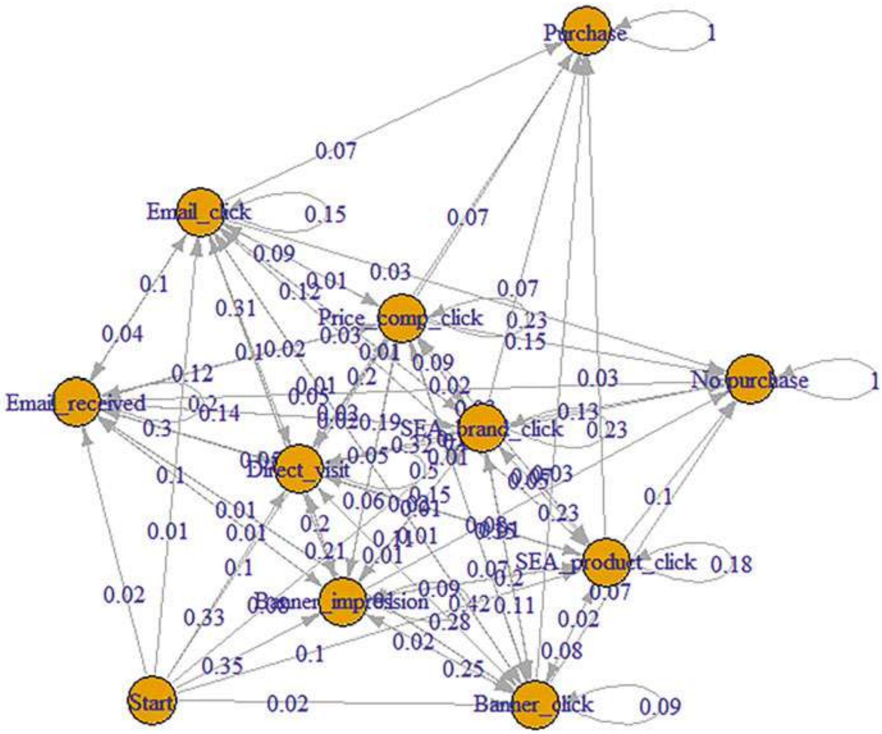

Fig. 18 Markov chain plot path to purchase

We can estimate what would happen if banner advertising is off by setting off this channel in the transition matrix. We can do this with the following R-script.

```
# Estimate transition matrix with banner impressions and clicks
dropped out
markov_chain_banner_off <- statetable.msm(Channel, Consumer_ID,
data=consumers_journey2_firm)
markov_chain_banner_off <- markov_chain_banner_off[c(1,4:9),c
(1,4:11)]
markov_chain_banner_off <- prop.table(markov_chain_banner_off,
margin=1)
markov_chain_banner_off <- as.data.frame(t
(markov_chain_banner_off))
markov_chain_banner_off <- markov_chain_banner_off[,3]
markov_chain_banner_off[64:81] <- c(rep(0,7),1,0,rep(0,8),1)
mcjourney_banner_off <- new("markovchain", states = c("Start",
"SEA_product_click","SEA_brand_click","Price_comp_click",
                                      "Email_received",
"Email_click", "Direct_visit",
                                      "Purchase", "No purchase"),
                       transitionMatrix = matrix
(data=markov_chain_banner_off,
```

	Start	Banner_imp.	Banner_click	SEA_prod._click	SEA_brand_click	Price_comp_click	Email_received	Email_click	Direct_visit	Purchase	No purchase
Start	0.000	0.347	0.015	0.102	0.083	0.099	0.018	0.008	0.328	0.000	0.000
Banner_imp.	0.000	0.415	0.022	0.090	0.110	0.055	0.015	0.007	0.207	0.000	0.080
Banner_click	0.000	0.254	0.091	0.077	0.108	0.067	0.012	0.007	0.285	0.028	0.071
SEA_prod._click	0.000	0.200	0.019	0.182	0.233	0.067	0.012	0.006	0.148	0.032	0.100
SEA_brand_click	0.000	0.148	0.010	0.050	0.229	0.019	0.018	0.008	0.319	0.067	0.131
Price_comp_click	0.000	0.189	0.011	0.032	0.094	0.230	0.018	0.007	0.198	0.072	0.149
Email_received	0.000	0.096	0.009	0.047	0.140	0.122	0.199	0.042	0.300	0.000	0.047
Email_click	0.000	0.097	0.013	0.029	0.123	0.086	0.102	0.148	0.306	0.072	0.025
Direct_visit	0.000	0.200	0.010	0.021	0.050	0.021	0.000	0.000	0.500	0.068	0.130
Purchase	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
No purchase	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000

Fig. 19 Transition matrix path to purchase

```
                                                byrow= TRUE, nrow = 9))
show(mcjourney_banner_off)
plot(mcjourney_banner_off, edge.arrow.size=0.25)
# Estimate the stage consumers are in after 5 and 50 phases of the
path to purchase, the initial state is the start state
initialstate <- c(1,0,0,0,0,0,0,0,0)
after5touchpoints <- initialstate * (mcjourney_banner_off ^ 5)
after5touchpoints
after50touchpoints <- initialstate * (mcjourney_banner_off ^ 50)
after50touchpoints
```

When running this R-script, we can observe that when the banner channel is off, the conversion rate after 50 stages is 33.39%, i.e., a substantial increase compared to the 27.76% we found when banner advertising is on. We do, however, need to remember that 34.72% of the journeys started with a banner impression and 1.51% start with a banner click (see Fig. 19), and these journeys would now not have occurred, leaving us with 63.77% of the journeys. If we take 63.77% of 33.39%, we get to a conversion rate of 21.29%, which is substantially lower than the 27.76% we have found earlier.

We can validate if this 21.29% conversion rate is close to the truth by investigating the transition matrix of the control group, i.e., the group where the firm's banner advertising was off in reality. We do this with the following R-script.

```
# Estimate transition matrix for the charity banner group
markov_chain <- statetable.msm(Channel, Consumer_ID,
data=consumers_journey2_charity)
markov_chain <- prop.table(markov_chain, margin=1)
markov_chain <- as.data.frame(t(markov_chain))
markov_chain <- markov_chain[,3]
markov_chain[100:121] <- c(rep(0,9),1,0,rep(0,10),1)
mcjourney <- new("markovchain", states = c
("Start","Banner_impression","Banner_click",
"SEA_product_click","SEA_brand_click","Price_comp_click",
                                      "Email_received", "Email_click",
"Direct_visit",
                                      "Purchase", "No purchase"),
                         transitionMatrix = matrix(data=markov_chain,
                                      byrow= TRUE, nrow = 11))
show(mcjourney)
plot(mcjourney, edge.arrow.size=0.25)
# Estimate the stage consumers are in after 5 and 50 phases of the
path to purchase, the initial state is the start state
initialstate <- c(1,0,0,0,0,0,0,0,0,0,0)
after5touchpoints <- initialstate * (mcjourney ^ 5)
after5touchpoints
after50touchpoints <- initialstate * (mcjourney ^ 50)
after50touchpoints
```

Running this R-script, we find that the conversion after 50 stages is 16.88%, i.e., still somewhat lower than the 21.29% that we found in our previous analysis, but it is perfectly in line with what we observed in Fig. 5. Turning off one channel in the Markov chain does thus seem to provide us a somewhat accurate estimate of the conversion, but it is far from perfect. This is because turning off one channel will also affect the transition matrix since some channels might be good alternatives. Combining the Markov chain with experimental data, as we have done here, does provide good insights. An advantage of this Markov chain, e.g., by comparing the Markov chain of the firm's banner group and the charity's banner group, is that we can explore how turning off one channel also impacts the usage of other channels in the path to purchase. For more details, also have a look at Spedicato et al. (2016).

Further Methods of Individual-Level Attribution

In this section, we discussed some methods for individual-level attribution. The provided R-scripts can be adopted and adjusted to be suitable for other, e.g., real-life datasets. The discussed models can also be used to investigate other outcome variables. Retention, customer acquisition, and other forms of customer behavior are examples of variables that can be valuable to investigate (e.g., Gupta et al. 2004; Gupta and Zeithaml 2006).

Next to other outcome variables, also other techniques can be used to analyze the data. Alternatives for the logistic regression model are causal trees and causal forests, which are machine learning techniques that try to estimate the treatment effect (i.e., the difference between the treatment and control group) and explain in which situations this treatment effect is larger or smaller. Such techniques can be convenient to determine which consumers should get a treatment (i.e., be targeted). For details on such techniques, see Hitsch and Misra (2018).

More advanced attribution techniques are available to investigate the impact of the order in which touchpoints occur. This can be done by including carryover and spillover effects in the models. Furthermore, the time between touchpoints can be important, which can be captured by including parameters that capture decay and restoration effects. Braun and Moe (2013) and Li and Kannan (2014) provide information on models that incorporate such effects.

Attribution with Aggregate-Level (Quasi-)Experimental Data

In some cases, it is impossible to do attribution at the individual customer level, for instance, when one wants to find out the impact of mass media advertising on conversion, but it is unknown which individual consumers have come in contact with this form of advertising. Furthermore, with stricter privacy laws, consumers blocking and deleting cookies, and consumers switching between different browsers and devices, tracking consumers across their path to purchase can become more complicated and sometimes even impossible, not allowed, or undesired. In such cases, attribution using experimental data is still achievable at a more aggregate level, e.g., instead of at the individual consumer level, one can conduct aggregate-level experiments or look at variations in the data across regions or over time.

Similarly, as with attribution at the individual level, randomized field experiments can be helpful for attribution at a more aggregate level. A challenge, in this case, is that it is not possible to randomly allocate consumers in different (treatment and control) groups. Instead, the randomization can, for instance, take place at the product, category, or regional level, over time, or at a combination of these levels. Two examples of studies conducting such experiments are Blake et al. (2015) and Wiesel et al. (2011). This section discusses three examples of analysis coming from data from such aggregate-level experiments. The first two forms of analysis, namely the *before-after analysis* and the *before-during-after analysis*, focus on variations over time. The third form of analysis focuses on variations over time and regions and is called the *difference-in-differences analysis*. In the end, we will also discuss some further methods when conducting aggregate-level attribution.

Before-After Analysis

To test the impact of a channel on the traffic to a website or a store, or any other outcome variable, and an experiment at the individual level is not possible, not desirable, or just impractical, an alternative solution is to turn off the channel for some time to see what happens to the desired outcome variable. If turning off a channel is too risky, an alternative is to decrease the expenditures for this channel temporally or, if the expectation is that the channel has a positive impact on the outcome variable, the expenditure can temporally be increased.

To give a simple example of this procedure, let us generate one dataset in line with the data used by Blake et al. (2015). In this example, we look at the search engine channel, and a firm decides to stop using SEA after 20 weekly observations. As one might assume, when stopping SEA, the traffic to the website will decrease. However, it might also be possible that some people who go to the website by clicking on a SEA link would still visit the website when the firm does not conduct SEA; instead of visiting the website through SEA, the consumer might in such case use an organic (nonpaid) link on the search engine, or they might have visited the website directly. Indeed, Blake et al. (2015) have shown that for eBay conducting SEA for branded keywords (i.e., keywords or search phrases containing "eBay") is not profitable; when eBay stops bidding for these keywords, they still show up high on the organic search results, and users still visit eBay's website. In such a case, organic search is a perfect alternative for paid search, and paid search should not get credit for the visits and the resulting sale. If after turning off SEA the traffic decreases, i.e., it is not (entirely) substituted by other channels, SEA should get credit for the lost traffic.

To investigate this, let us generate a dataset of 40 weekly observations. The firm turns off SEA after week 20, and we have data on the traffic coming to a website via SEA and organic search and the total traffic coming in via the search engine.

```
# Load required packages -----------------------------------------
library(ggplot2)
library(reshape2)
#Turn off scientific notation
options(scipen = 999)
# Generate dataset before-after analysis ------------------------
n=40 #has to be an even number for the later code to run correctly
set.seed(1234)
search_data <- data.frame(c(1:n), c(rep(1, times=n/2),rep(0,
times=n/2)))
names(search_data)[1] <- paste("Week")
names(search_data)[2] <- paste("SEA_on")
search_data$Paid_volume <- ifelse(search_data$SEA_on==1,
                                 25000+sample(0:5000, n),
                              0)
search_data$Organic_volume <- ifelse(search_data$SEA_on==1,
                              search_data$Paid_volume*1.5 +
10000                         + sample(0:5000, n),
                              67500 + sample(5000))
# Create total traffic variable
search_data$Total_volume <- search_data$Paid_volume + search_data
$Organic_volume
# Create plot of total, organic and paid traffic
dd = search_data[,c(1,3:5)]
dd = melt(dd, id=c("Week"))
colnames(dd)[3] <- "Volume"
ggplot(dd) + geom_line(aes(x=Week, y=Volume, colour=variable)) +
  scale_colour_manual(values=c("red","green","blue"))
rm(dd)
rm(n)
```

After running this R-script, we have a dataset with 40 weekly observations. Figure 20 shows the plotted data. As can be seen, in the first 20 weeks, both the organic traffic and the paid traffic are relatively stable, but after week 20, the paid traffic drops to zero, which makes sense since SEA is off for this period. We can also see that after week 20, there is a substantial increase in organic traffic to the website. The total traffic does go down somewhat, but not as much as the lost SEA traffic, i.e., it seems that organic search is partly capturing the lost traffic of SEA. Assuming that everything else is stable over time, the explanation for the increase in organic search is that some visitors who would otherwise have clicked on a SEA link now click on the organic search link. This example nicely shows that last-click attribution would have overestimated the impact of SEA since there is an alternative channel available which browsers would use to go to the website if SEA is not available.

To see if SEA has a significant impact on the total traffic to the website, we can run the following R-script to estimate some regression models.

```
# Load package to make an output table of all models
library(sjPlot)
library(sjmisc)
library(sjlabelled)
```

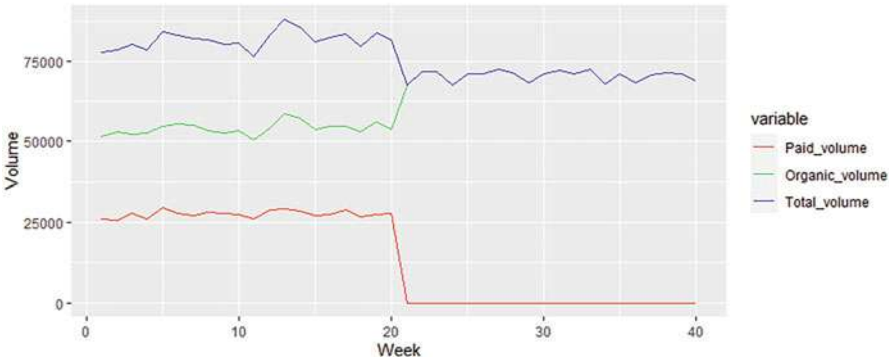

Fig. 20 Data of before-after analysis visualized

```
#regression models before-after analysis
before_after_model1 <- lm(Total_volume ~ SEA_on, data=search_data)
before_after_model2 <- lm(Total_volume ~ SEA_on + Week,
data=search_data)
before_after_model3 <- lm(Organic_volume ~ SEA_on,
data=search_data)
before_after_model4 <- lm(Organic_volume ~ SEA_on + Week,
data=search_data)
tab_model(before_after_model1, before_after_model2,
before_after_model3,
        before_after_model4, collapse.ci = TRUE, p.style =
"stars")
```

As we can see from the output in Table 6, Model 12 shows that when SEA is on, the traffic coming in through the search engine is 11,178.78 visitors higher on average (95% CI [9703.67, 12653.88]), which is highly significant. Controlling for the trend variable "week," which could capture a linear upward or downward trend in the number of visitors over the 40 weeks, does not substantially change the estimate and also does not improve the model, as is shown by Model 13. Model 14 shows that when SEA is on, organic search traffic to the website is a significant 16,432.58 units lower, or the other way round; when turning off SEA, organic search traffic goes up 16,432.58 on average, which is the substation effect of this channel for SEA. Controlling for the week again does not bring much change, as Model 15 shows.

We can use the parameter estimates also as a basis to find the return on investment (ROI) of SEA. Let us assume a 10% conversion rate, a gross profit (before marketing costs) of $20 per conversion, and SEA costs $0.50 per click on average. What we furthermore need is the total traffic from SEA when SEA is on. For this, we can take the mean amount of SEA clicks when SEA is on, which we can get with the following R-script.

```
# Mean amount of Paid clicks (i.e. variable 3 in our dataset) when
SEA is on mean(search_data[search_data$SEA_on==1,3])
```

Table 6 Regression output of before-after analyses

	Model 12	Model 13	Model 14	Model 15
	Total_volume	Total_volume	Organic_volume	Organic_volume
Predictors	*Estimates*	*Estimates*	*Estimates*	*Estimates*
(Intercept)	70444.05 *** (69400.99–71487.11)	67755.39 *** (63765.61–71745.17)	70444.05 *** (69632.73–71255.37)	68747.27 *** (65615.41–71879.13)
SEA_on	11178.78 *** (9703.67–12653.88)	12941.83 *** (10024.28–15859.39)	−16432.58 ***(−17579.95 – −15285.20)	−15319.93 ***(−17610.13 – −13029.74)
Week		88.15(−38.22–214.53)		55.63(−43.57–154.83)
Observations	40	40	40	40
R^2/R^2 adjusted	0.861/0.857	0.868/0.861	0.957/0.956	0.958/0.956

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

We find that there are 27,611.35 paid clicks on average when SEA is on. We can now fill in the following ROI formula, based on the parameter from Model 12 from Table 6 (The assumption here is that the 11,178.78 additional visitors is indeed caused by SEA, i.e., without SEA we would lose these visitors, in line with what the model shows.):

$$\begin{aligned} &\boldsymbol{Incremental\ gross\ profit\ (if\ SEA\ is\ on) = 11,178.78\ additional\ visitors\cdot} \\ &\boldsymbol{10\%conversion\cdot 20/conversion = 22,357.56} \end{aligned} \tag{2}$$

$$\boldsymbol{Incremental\ costs\ (if\ SEA\ is\ on) = 27,611.35\ clicks\cdot 0.50/click = 13,805.68} \tag{3}$$

$$\boldsymbol{ROI = \frac{Gross\ profit - Costs}{Costs}\cdot 100\% = 22,357.56 - \frac{13,805.68}{13,805.68\cdot 100\% = 61.94\%}} \tag{4}$$

So, the ROI of SEA is, in this case, 61.94%. We can also replace the 11,178.78 in the formula with the upper and lower score of the confidence interval (i.e., the 95% CI [9703.67, 12653.88]). If we do so, we find the 95% confidence interval of the ROI to be 40.58% and 83.31%, which indicates SEA is, on average, an excellent investment in this case. Note that this is a generated dataset. In reality, the effectiveness and ROI of SEA can be much different. Blake et al. (2015) have for instance found a very low effectiveness of SEA in their study, which might be explained by the fact that they used data from a very well-known firm, namely eBay.

One can easily use this before-after analysis for other questions when investigating the impact of SEA:

- For search engine advertising, instead of looking at the visitors via SEA and organic search, one can also directly investigate the total visitors of the website or the total purchases per week. The advantage of this is that turning off SEA might also impact other channels (e.g., direct website visit), and conversion rates might differ per channel (e.g., the 10% conversion we used in the ROI calculation might not hold for every channel), so looking at sales as a dependent variable can be more appropriate to calculate the ROI.
- Instead of looking at the overall impact of SEA, one can also use this procedure for different types of keywords; e.g., what happens when turning off SEA for branded keywords? Or what happens when turning off SEA for keywords for a specific product category?
- Furthermore, this procedure is helpful to investigate the impact of bids; e.g., what happens when lowing the bids for a period by x%? Or what happens if the budget is temporarily increased or decreased with y%? These kinds of experiments can help to find more optimal bids and determine the budget allocation.

By conducting such experiments, one can thus find out if SEA is, on average, a profitable channel, but it is also possible to improve the bids per keyword and the overall budget allocation. Next to SEA, before-after analyses are also helpful for other channels, including offline mass media like TV and radio advertising.

A limitation of a before-after analysis is that we assume that all other factors which influence traffic before and after the change (e.g., turning off the channel), which we do not control for, are stable over time. Let us assume that the company conducting the experiment is selling sunglasses, and the experiment starts in October, and we do observe that the sales in the "after" period are lower than the "before" period. Can we then conclude that this decrease in sales is due to turning off SEA? Or could this decrease be (partly) due to seasonality, e.g., the "after" period was less sunny, and the demand for sunglasses was lower than the "before" period? Alternatively, if a company is growing over time, the after period is affected by the experiment and the company's growth, which might hide part of the effect estimated with the model. Adding control variables in the regression model can (partly) overcome these confounding effects; e.g., one can control for temperature and the average hours of sunshine per day and a trend by adding a trend variable to the model.

In practice, we cannot control for everything that might change over time and impact our dependent variable of interest. With that, we cannot guarantee causality since we cannot exclude all potential third variables. However, there are two ways we can improve upon the before-after analysis. The following two sections discuss these alternative analyses.

Before-During-aAfter Analysis

An extension of the before-after analysis is the before-during-after analysis in which the turned-off channel (or changed in any other way) is turned back on after the experimental period. An advantage with this is that we can now inspect if the outcome variable (e.g., visitors or sales) goes back to the level as it was before; this is what we would expect when the found impact is indeed causal and when there is no long-term (positive or negative) impact of turning off or on a channel (e.g., via carryover effects or other dynamic effects). This approach is also used by Blake et al. (2015) when investigating the impact of SEA at eBay.

To conduct the before-during-after analysis, let us first generate some data to make this clearer. We can do this with the following R-script.

```
# Generate dataset before-during-after analysis attribution ------
n=60 #has to be a multiple of 3 for the later code to run correctly
set.seed(1234)
search_data <- data.frame(c(1:n),
                  c(rep(1, times=n/3),rep(0, times=n/3),rep(1,
times=n/3)),
                  c(rep(0, times=n/3),rep(1, times=n/3),rep(0,
times=n/3)),
                  c(rep(0, times=n/3),rep(0, times=n/3),rep(1,
times=n/3)))
```

```
names(search_data)[1] <- paste("Week")
names(search_data)[2] <- paste("SEA_on")
names(search_data)[3] <- paste("During")
names(search_data)[4] <- paste("After")
search_data$Paid_volume <- ifelse(search_data$SEA_on==1,
                            25000+sample(0:5000, n),
                            0)
search_data$Organic_volume <- ifelse(search_data$SEA_on==1,
                        search_data$Paid_volume*1.5 + 10000 +
                        sample(0:5000, n),
                        67500 + sample(5000))
# Create total traffic variable
search_data$Total_volume <- search_data$Paid_volume + search_data
$Organic_volume
# Create plot of total, organic and paid traffic
##Subset the necessary columns
dd = search_data[,c(1,5:7)]
dd = melt(dd, id=c("Week"))
colnames(dd)[3] <- "Volume"
ggplot(dd) + geom_line(aes(x=Week, y=Volume, colour=variable)) +
  scale_colour_manual(values=c("red","green","blue"))
rm(dd)
rm(n)
```

After running this R-script, we have a dataset similar to the before-after study in the previous section, but with 20 additional weekly observations, i.e., the period in which SEA is back on. Figure 21 plots these data. For the first 40 weekly observations, we can see a similar pattern as with the before-after study visualized in Fig. 20, and in the last 20 weeks, when SEA is back on, we can see that everything goes back to a similar situation as in the first 20 weeks. This last period is interesting since if there were other changes over time (e.g., seasonality, an overall change in the website traffic), we would expect that the after period is different from the before period. If the before and after periods are similar, and the period during the experiment is different, we have some more certainty that the change is due to turning off SEA and not (also) due to other factors.

To see if turning off SEA has a significant impact, and if this impact is gone when turning SEA back on, we can run the following R-script to estimate the necessary regression models.

```
#regression models before-during-after analysis
before_during_after_model1 <- lm(Total_volume ~ During + After,
data=search_data)
before_during_after_model2 <- lm(Total_volume ~ During + After +
Week, data=search_data)
tab_model(before_during_after_model1, before_during_after_model2,
          collapse.ci = TRUE, p.style = "stars")
```

As we can see from the output in Table 7, in the "during" period (i.e., when SEA is off), the total traffic coming in through the search engine is significantly decreasing compared to the "before" period. This finding is in line with the before-after analysis. The "after" period parameter is insignificant; since the "before" period is our reference case, this means that there is no significant difference in the number of website visits before the experiment took place and after the experiment has finished. When SEA is back on, the situation does thus goes back to normal. If the "after" period would significantly deviate from the "before" period in the before-during-after analysis, it would signal that there is also something else changing over time, or there might be dynamic effects of turning off and on SEA which the model does not fully capture. As a result, in the case of a significant "after" period, we should be cautious when interpreting the "during" period. If we find the reason for a significant "after" period, e.g., seasonality, we can control for these factors in the model and investigate if controlling for this indeed leads to a nonsignificant "after" period.

An even more robust approach to exclude other factors is the difference-in-differences analysis, discussed in the next section.

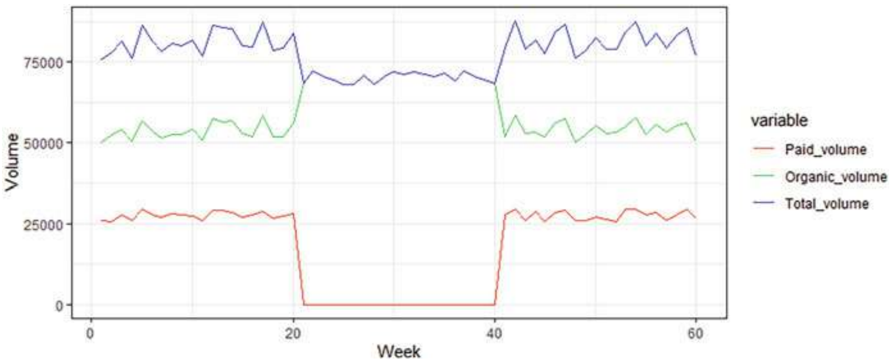

Fig. 21 Data of before-during-after analysis visualized

Table 7 Regression output for before-during-after analyses

	Model 16	Model 17
	Total_volume	Total_volume
Predictors	*Estimates*	*Estimates*
(Intercept)	81233.12 *** (79856.18–82610.07)	80331.76 *** (78343.03–82320.50)
During	−10961.42 ***(−12908.72 – −9014.13)	−12678.30 ***(−16038.48 – −9318.12)
After	471.80(−1475.50–2419.10)	−2961.95(−8783.57–2859.66)
Week		85.84(−51.39–223.08)
Observations	60	60
R^2/R^2 adjusted	0.756/0.748	0.763/0.750

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Difference-in-Differences Analysis

With the before-after analysis, there could be confounding factors that change over time and influence the results. The before-during-after analysis is somewhat more robust since one can compare the before and after periods to see if everything goes back to normal when the experimental period is over. There might still have been some other factors influencing the results, e.g., a holiday during the experimental period or activities by a competitor. To be even more confident that the effects are indeed causal, one can make the change (or treatment) only in some regions (the treatment regions) and not make this change in other (similar) regions (the control regions). If the regions are randomly assigned, and there is no difference observed between the regions before the experiment, then the deviations during the experiment show us the impact of the experiment (e.g., turning off a channel in the treatment regions). Multiple studies use this approach of conducting changes in only some regions; examples include Blake et al. (2015) and Wiesel et al. (2011).

To give an example of a difference-in-differences analysis, let us generate some data with the following R-script.

```
# Generate dataset dif-in-dif analysis attribution --------------
n=40 #has to be an even number for the later code to run correctly
set.seed(1234)
search_data <- data.frame(c(1:n), c(rep(1, times=n/2),rep(0,
times=n/2)))
names(search_data)[1] <- paste("Week")
names(search_data)[2] <- paste("SEA_on")
search_data$Total_traffic_region1 <- ifelse(search_data$SEA_on==1,
                                  75000+sample(0:3000, n),
                                  70000+sample(0:3000, n))
search_data$Total_traffic_region2 <- 75000+sample(0:3000, n)
# Create plot of traffic per region
dd = search_data[,c(1,3:4)]
library(reshape2)
dd = melt(dd, id=c("Week"))
colnames(dd)[3] <- "Volume"
ggplot(dd) + geom_line(aes(x=Week, y=Volume, colour=variable)) +
  scale_colour_manual(values=c("red","blue"))
# Create some additional variables
dd$treatment_region <- c(rep(1, times=n),rep(0, times=n))
dd$SEA_off <- c(rep(0, times=n/2),rep(1, times=n/2),rep(0,
times=n/2),rep(1, times=n/2))
rm(n)
```

As shown in Fig. 22, the traffic to the website is for the control region (region 2) relatively stable over the 40 weeks. For the treatment region (region 1), the traffic volume is similar to the control region in the first 20 weeks (i.e., when in both regions SEA was on), but the traffic volume starts to deviate when SEA is off in the treatment region. Assuming that the regions are indeed similar, and thus also that

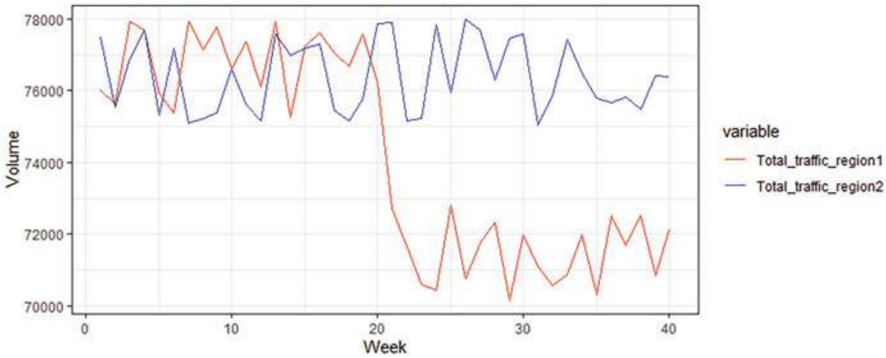

Fig. 22 Data of difference-in-differences analysis visualized

seasonality and trends similarly affect both regions, we can assume that the deviation between the two regions is indeed due to the treatment (i.e., turning off SEA).

To test the significance, we can estimate a difference-in-differences model. This model looks as follows:

$$\boldsymbol{Y_{it} = \beta_0 + \beta_1 \cdot Treatment_i + \beta_2 \cdot After_t + \beta_3 \cdot Treatment_i \cdot After_t + e_{it}} \tag{5}$$

Where $Treatment_i$ indicates if the observation is from the treatment group (1) or control group (0), $After_t$ indicates if the period was after (1) or before (0) the treatment has taken place. The parameter β_1 indicates if the treatment group differs from the control group in the period before the experiment took place; if the treatment region is similar to the control region, this parameter should be insignificant. The parameter β_2 indicates if the control group differs in the after period compared to the before period; if there are no changes over time, this should be insignificant. If there are changes over time, which are not due to the experiment, these changes are captured by β_2. The main parameter of interest is β_3, which captures the deviation of the treatment group from the control group after the treatment; if the two groups are indeed the same over time, with the only difference being the treatment, then β_3 captures the causal effect of the treatment.

To estimate the difference-in-differences model, the following R-script can be used.

```
# Regression model difference-in-differences analysis
dif_dif_model1 <- lm(Volume ~ treatment_region + SEA_off +
treatment_region*SEA_off, data=dd)
dif_dif_model2 <- lm(Volume ~ treatment_region + SEA_off +
treatment_region*SEA_off + Week, data=dd)
tab_model(dif_dif_model1, dif_dif_model2,
          collapse.ci = TRUE, p.style = "stars")
```

As can be seen in Table 8, the parameter for “treatment region” is insignificant in both Models 18 and 19. In line with the explanation of the parameter β_1 of Eq. (5),

Table 8 Regression output for difference-in-differences analyses

	Model 18	Model 19
	Volume	Volume
Predictors	*Estimates*	*Estimates*
(Intercept)	76327.15 *** (75901.70–76752.60)	76381.41 *** (75802.38–76960.43)
treatment_region	532.20(−69.48–1133.88)	532.20(−73.30–1137.70)
SEA_off	150.55(−451.13–752.23)	253.90(−704.20–1211.99)
treatment_region * SEA_off	−5525.40 ***(−6376.30 – −4674.50)	−5525.40 ***(−6381.70 – −4669.10)
Week		−5.17(−42.29–31.96)
Observations	80	80
R^2/R^2 adjusted	0.849/0.843	0.849/0.841

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

this insignificance indicates that the treatment group does not differ significantly from the control group in the period before the experiment took place. If the regions are assigned randomly to the control and treatment condition, an insignificant parameter is thus indeed what we would expect. If the parameter is significant, it signals that there is already a difference between the two groups before the experiment took place. The parameter of "SEA_off," i.e., our treatment, is also insignificant. Given the interaction in the model, this parameter indicates that the traffic volume is not significantly different for the control group in the "after" period. The interaction is highly significant; this parameter is the effect of the treatment (i.e., turning off SEA) in the treatment region and tells us that turning off SEA does significantly decrease the traffic to the website. We can conclude that turning off SEA results in a visitor drop of 5,525.40 (i.e., the difference-in-differences effect).

We can also investigate the difference-in-differences by simply using a cross-table, which can sometimes be easier to understand and communicate the results than a somewhat more complicated regression model.

```
# Create crosstable of difference-in-differences
with(dd, tapply(Volume,
list(treatment_region=treatment_region,SEA_off=SEA_off), mean) )
```

Table 9 shows the cross-table with the difference-in-differences effect, i.e., taking the difference between the before and after periods for the control group and the treatment group and then taking the difference between these two differences (hence "difference-in-differences"), is the same as the parameter estimate from Table 8. An advantage of using a regression model instead of a cross-table is that we can observe if the difference is statistically significant, and we can include control variables in the regression model.

For a difference-in-differences analysis, it is essential that the treatment and the control groups are comparable. In their study, Blake et al. (2015) have assured this

Table 9 Difference-in-differences analyses in a cross-table

	After (treatment period)	Before (control period)	Difference
Treatment group	71,484.50	76,859.35	−5,374.85
Control group	76,477.70	76,327.15	−150.55
Difference	−4,993.20	532.20	**−5,525.40**

by matching regions to each other, which are similar in their historical sales patterns over time. Wiesel et al. (2011) did match regions that were similar in consumer expenditure, the recency, frequency, and monetary value of the purchases, and the number of new and existing customers.

If the control region is different in size compared to the treatment region, but (apart from the treatment) there are no differences over time, this is still not a problem; the size differences will be captured by β_1 of Eq. (5), which would then be statistically significant. Alternatively, weights can make the regions in the before period comparable.

If there are differences in size between the groups, but these differences are not absolute (e.g., region one has on average 1,000 more visitors, which stay similar over time) but relative (e.g., region one has on average 10% more visitors, which stays similar over time) in size, it is better to log transform the dependent variable of interest. By log transforming the dependent variable, the model investigates the relative differences.

To give another example of a difference-in-differences analysis, now with some of the challenges discussed above let us create a new dataset. In this new dataset, the control region is ~75% larger than the treatment region, we have a holiday period in weeks 25 and 26, which causes 10% additional sales in both the treatment and control regions, and the traffic is increasing over time.

```
# Generate second dataset dif-in-dif analysis attribution
n=40 #has to be an even number for the later code to run correctly
set.seed(1234)
search_data <- data.frame(c(1:n), c(rep(1, times=n/2),rep(0,
times=n/2)))
names(search_data)[1] <- paste("Week")
names(search_data)[2] <- paste("SEA_on")
search_data$Holidays <- ifelse(search_data$Week>24 & search_data
$Week<27 , 1, 0)
search_data$Total_traffic_region1 <- (ifelse(search_data$SEA_on==1,
                                   75000 + search_data$Week*300 +
sample(0:3000, n),
                                   70000 + search_data$Week*300) +
sample(0:3000, n))*ifelse(search_data$Holidays==1, 1.1, 1)
search_data$Total_traffic_region2 <- (75000 + search_data$Week*300
+ sample(0:3000, n))* ifelse(search_data$Holidays==1, 1.1, 1)*1.75
# Create plot of traffic per region
dd = search_data[,c(1,4:5)]
library(reshape2)
dd = melt(dd, id=c("Week"))
```

```
colnames(dd)[3] <- "Volume"
ggplot(dd) + geom_line(aes(x=Week, y=Volume, colour=variable)) +
  scale_colour_manual(values=c("red","blue"))
# Create some additional variables
dd$treatment_region <- c(rep(1, times=n),rep(0, times=n))
dd$SEA_off <- c(rep(0, times=n/2),rep(1, times=n/2),rep(0,
times=n/2),rep(1, times=n/2))
dd$Holidays <- ifelse(dd$Week>24 & dd$Week<27 , 1, 0)
rm(n)
```

As shown in Fig. 23, the control region (i.e., region 2, indicated with the blue line) is indeed larger than the treatment region (i.e., region 1, indicated with the red line). Furthermore, we can observe a drop in traffic to the website in the treatment region directly after week 20. This drop is not visible in the treatment region. We do observe in both groups the holiday peak in weeks 25 and 26. If we would just take the mean value of traffic before and after turning off SEA in the treatment region, the number of visitors before and after the treatment would be very similar because of the upward trend and the holiday peak in weeks 25 and 26. Using a simple before-after analysis would, in this case, not be appropriate, and a difference-in-differences analysis is likely to show better the causal effect of turning off SEA.

To demonstrate this, let us estimate a series of difference-in-differences models with the following R-script.

```
# Regression model difference-in-differences analysis
dif_dif_model3 <- lm(log(Volume) ~ treatment_region + SEA_off +
treatment_region*SEA_off, data=dd)
dif_dif_model4 <- lm(log(Volume) ~ treatment_region + SEA_off +
treatment_region*SEA_off + Week, data=dd)
dif_dif_model5 <- lm(log(Volume) ~ treatment_region + SEA_off +
treatment_region*SEA_off + Week + Holidays, data=dd)
tab_model(dif_dif_model3, dif_dif_model4, dif_dif_model5,
          collapse.ci = TRUE, p.style = "stars")
```

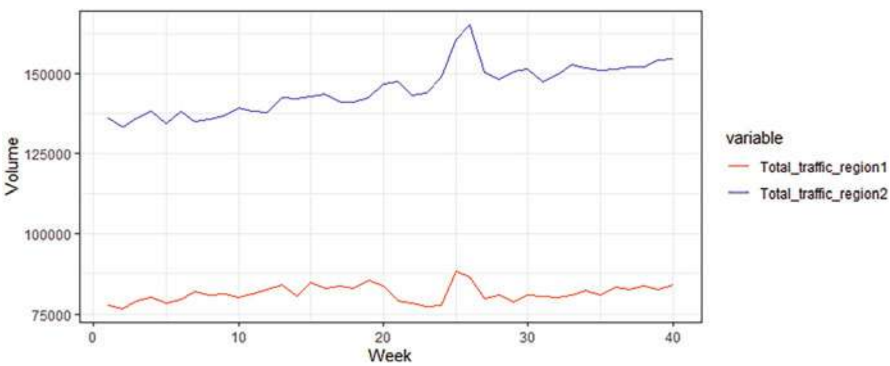

Fig. 23 Data of trending difference-in-differences analysis visualized

Table 10 shows that the parameter for “treatment region” and the difference-in-differences effect (i.e., the interaction parameter) are consistent over the three versions of the model. Since the changes over time (i.e., the trend and the holidays) have the same relative effect size in both regions, the model is robust for these changes. The confidence interval for the difference-in-differences parameter does become smaller if we control for the trend and the holiday weeks because these events result in more noise in the data.

The parameter for “treatment region” is 0.54, which means, due to the log transformation of the dependent variable, the treatment region has about (exp (−0.54)*100% – 1) 41.7% fewer visitors compared to the control region, or in other words, the control region has (exp(0.54)*100% –1) 71.6% more visitors than the treatment regional, comparable to the 75% which we have set it to be. The holiday parameter is significant and has a value of 0.10, meaning that during the holidays, the number of visitors is (exp(0.10)*100% – 1) 10.5% higher, in line with the 10% we have set. Turning SEA off in the treatment region, i.e., the difference-in-differences effect, leads to a (exp(−0.08)*100% – 1) 7.7% decrease in traffic to the website. These figures can be the basis for ROI calculations of the SEA channel. For more information on interpreting the parameters of a regression model when the dependent variable is log transformed, please have a look at Chapter 2 of Leeflang et al. (2015).

In Model 20 of Table 10, the parameter for “SEA_off” is significant. In line with the discussion of eq. (5), a significant β_2 means that there is a difference in the “before” and the “after” period for the control region. This is true since there is an upward trend and a holiday peak in the after period. When including the “week” and “holiday” variables as control variables, as we do in Model 22, this effect disappears, i.e., we have with the full model controlled for the differences over time.

Table 10 Regression output for trending difference-in-differences analyses

	Model 20	Model 21	Model 22
	log(Volume)	log(Volume)	log(Volume)
Predictors	*Estimates*	*Estimates*	*Estimates*
(Intercept)	11.84 *** (11.83–11.86)	11.81 *** (11.80–11.83)	11.80 *** (11.79–11.81)
treatment_region	−0.54 ***(−0.56 – −0.52)	−0.54 ***(−0.55 – −0.52)	−0.54 ***(−0.54 – −0.53)
SEA_off	0.08 *** (0.06–0.10)	0.02(−0.00–0.05)	−0.00(−0.02–0.01)
treatment_region * SEA_off	−0.08 ***(−0.11 – −0.06)	−0.08 ***(−0.11 – −0.06)	−0.08 ***(−0.09 – −0.07)
Week		0.00 *** (0.00–0.00)	0.00 *** (0.00–0.00)
Holidays			0.10 *** (0.09–0.12)
Observations	80	80	80
R^2/R^2 adjusted	0.989/0.989	0.993/0.993	0.998/0.998

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Let us compare the results from Table 10 with what we would find if we would only have information from the treatment region and, based on that, estimate a before-after analysis. We do this with the following R-script.

```
# Regression model before-after analysis on difference-in-
differences data
before_after_model5 <- lm(log(Volume) ~ SEA_off, data=dd,
subset=treatment_region==1)
before_after_model6 <- lm(log(Volume) ~ SEA_off + Week, data=dd,
subset=treatment_region==1)
before_after_model7 <- lm(log(Volume) ~ SEA_off + Week + Holidays,
data=dd, subset=treatment_region==1)
tab_model(before_after_model5, before_after_model6,
before_after_model7,
          collapse.ci = TRUE, p.style = "stars")
```

Indeed, Model 23 in Table 11 shows that when we do a simple before-after analysis, without controlling for the time effects the parameter for turning SEA off is precisely zero and insignificant, i.e., there is no difference between the before and after periods. This is indeed in line with what we see when we look at the total traffic from region 1 in Fig. 23; due to the upward trend, the before and after periods look very similar, although it is clear that the traffic drops immediately when SEA is off, but it recovers due to the (unrelated to SEA) upward trend. When controlling for this upward trend by including a trend variable, as is done in Model 24 in Table 11, we get much closer to the actual effect of SEA as shown by the difference-in-differences analyses in Table 10. We are still somewhat off because the traffic in the after period is higher for two weeks due to the holiday peak, which is unrelated to turning SEA off and thus hides part of the drop in the number of visitors. When also controlling for this holiday peak, as is done in Model 25 in Table 11, we get the same (i.e., accurate) parameter estimate for turning off SEA as in the difference-in-differences models from Table 10. The results from Table 11 do thus show that a before-after analysis is sensitive for other variations over time. When we control for these changes, as we do in Model 25 in Table 11, we get a pretty accurate parameter estimate, but if we do not control for all variables, we might get the wrong estimate (as shown in Model 23 of Table 11). With this, we thus show the superiority of a difference-in-differences analysis over the before-after analysis.

The difference-in-differences analysis still has some drawbacks, namely that it needs an experiment to run this analysis, which is in practice not always possible, or it might be too time consuming or too risky to conduct. Furthermore, with a difference-in-differences analysis and the before(−during)-after analyses, it is hard to capture long-term effects. When, for instance, turning off TV advertising, it might be that this does not directly lead to significantly lower visitors or sales since it might take some time before the effect takes place (i.e., wear-in and wear-out effects). In order to capture this, the model needs to include dynamic effects. Including dynamics is possible with lag terms or stock variables (Hanssens 2021). Other methods

Table 11 Regression output for trending difference-in-differences data using before-after analyses

	Model 23	Model 24	Model 25
	log(Volume)	log(Volume)	log(Volume)
Predictors	*Estimates*	*Estimates*	*Estimates*
(Intercept)	11.31 *** (11.29–11.32)	11.27 *** (11.25–11.29)	11.26 *** (11.25–11.27)
SEA_off	0.00(−0.02–0.02)	-0.07 ***(−0.10 – −0.03)	−0.09 ***(−0.11 – −0.08)
Week		0.00 *** (0.00–0.00)	0.00 *** (0.00–0.00)
Holidays			0.10 *** (0.08–0.12)
Observations	40	40	40
R^2/R^2 adjusted	0.000/−0.026	0.395/0.362	0.851/0.839

* *$p < 0.05$* ** *$p < 0.01$* *** *$p < 0.001$*

which capture dynamic effects, and do not rely on experimental data, will be discussed in the next section.

For more information on exploiting data from field experiments and applied time series analysis, see Artz and Doering (2021) and Wang and Yildirim (2021).

Further Methods of Aggregate-Level Attribution

In some cases, it is not possible, or it might be too time consuming or too risky, to conduct an experiment. Because of this, managers might not be willing to set up an experiment or have the time to wait for the results. Firms typically have historical data on their advertising expenditures per channel and data on other marketing mix components like pricing and distribution and performance data like sales. These data are typically available over time, and the changes over time in the marketing variables can thus be related to changes over time in firm performance.

One challenge is that decisions on changes in the marketing mix are not set in isolation; e.g., (past) firm performance might drive current pricing and advertising expenditure. A drop in sales might make managers decide to lower the price to recover market share, while higher sales in the previous period might lead to higher budgets for advertising in the current period, as the advertising budget is often a function of (previous and expected) revenue. A standard regression model does not take these kinds of reversed causality and endogeneity into account.

Another challenge is that the effects can be dynamic, e.g., spending more on advertising in the current period might not only affect current firm performance, but it might also still have an impact in the next period(s). These carryover effects are not taken into account in a standard regression model unless explicitly included via, for instance, (multiple) lag terms or ad stock variables (see Hanssens [2021] for more details on this).

A model which can take reversed causality, (specific forms of) endogeneity, and dynamics into account is the vector autoregressive (VAR) model. One example of an attribution paper that uses a VAR model is Wiesel et al. (2011), who investigate how online and offline advertising affect the online and offline sales channel usage and the profit of a B2B office furniture seller. Another attribution paper that uses a series of VAR models is De Haan et al. (2016). They investigate how various online and offline advertising channels impact website progression and sales revenue at an online retailer for five different product categories. Both Wiesel et al. (2011) and De Haan et al. (2016) demonstrate how the VAR models can be used to improve advertising budget allocation. Other studies related to attribution which have used VAR models include Trusov et al. (2009), Srinivasan et al. (2010), Pauwels et al. (2016b), and Srinivasan et al. (2016), just to name a few. The book chapter from Srinivasan (2021) discusses the VAR model in detail, and for the interested reader on www.evertdehaan.com there is also a R-script available to conduct a simple VAR model for attribution.

There are also more sophisticated methods capable of performing attribution at an aggregate level. One R-package to highlight is "CausalImpact" (Brodersen and Hauser 2021). This package is helpful to explore experiments, i.e., a random event in which there is a sudden change in one marketing channel. The package tries to estimate the value of the outcome variable if the event did not occur, i.e., the so-called counterfactual scenario. This counterfactual scenario is exactly what we want to have when conducting attribution, namely finding out what would have happened if a channel or touchpoint was not active. For more details, see Brodersen and Hauser (2021) and try out the example code, which can be downloaded at www.evertdehaan.com.

The aggregate-level attribution methods can also be used with other dependent variables similar to the individual-level attribution methods. Examples would be the number of offline (i.e., brick-and-mortar) stores visits over time, revenue, profitability, market share, and stock return. These variables indicate another advantage of aggregate-level attribution, namely that one can use dependent variables not measured at the individual customer level or variables that are hard to link to an individual's online behavior or touchpoints. Also, variables related to perceptual outcomes, e.g., customer mindset metrics like customer satisfaction and brand awareness, are interesting as (intermediate) outcome variables. Srinivasan et al. (2010) and De Haan et al. (2021) demonstrate the advantage of using such perceptual outcomes instead of or next to using more transactional (financial and behavioral) outcome variables.

Furthermore, it is possible to combine different techniques, e.g., start with a VAR model, followed by an aggregate-level experiment, as Wiesel et al. (2011) did. Such a setup can help overcome some of the challenges and risks, e.g., the data for a VAR model is easier to collect and is a good way to get first insights, while an experiment provides more certainty of finding causal effects. Also, other techniques to analyze aggregate-level data are possible, including state-space models, structural models, and Bayesian analysis. See the book by Leeflang et al. (2017) for details on these and further methods.

Conclusion

As introduced at the beginning of this chapter, attribution is the process to "allocate appropriate credit for a desired customer action to each marketing touchpoint across all online and off-line channels" (Kannan et al. 2016). With this it is essential to find the incremental effect a specific touchpoint has on the outcome of a path to purchase. Basic attribution methods, based purely on encountering a specific touchpoint (i.e., touch-based attribution), are not suited for this since they do not provide information about what would have happened if the touchpoint was not there. In order to find this out, Shapley values already provide better insights, but the best insights can be retrieved by conducting (individual- or aggregate-level) field experiments.

In situations where individual-level field experiments are not possible, feasible, or desirable, more elaborate attribution methods are available, including PSM, Markov chain, and aggregate-level field experiments. All of these methods have their advantages and disadvantages, as discussed in this chapter. Furthermore, there is a wide range of other (simple and complicated) procedures and models for attribution. However, the basic idea remains the same; finding out what would have happened without a specific touchpoint gives a certain amount of credit to that touchpoint.

Therefore, this chapter introduces attribution modeling, which can help to critically evaluate current attribution methods used within an organization and give directions on how this can be improved. We should also not consider attribution a goal in itself, but it can help decide on budget allocations and decide how, when, and where to target which customer. With this, attribution can help make marketing more accountable and make better advertising and targeting decisions.

Cross-References

- ▶ Analysis of Variance
- ▶ Applied Time-Series Analysis in Marketing
- ▶ Dealing with Endogeneity: A Nontechnical Guide for Marketing Researchers
- ▶ Experiments in Market Research
- ▶ Exploiting Data from Field Experiments
- ▶ ExploitingData from Field Experiments
- ▶ Field Experiments
- ▶ Logistic Regression and Discriminant Analysis
- ▶ Modeling Marketing Dynamics Using Vector Autoregressive (VAR) Models
- ▶ Panel Data Analysis: A Non-technical Introduction for Marketing Researchers
- ▶ Regression Analysis
- ▶ Return on Media Models

References

Anderl, E., Becker, I., Von Wangenheim, F., & Schumann, J. H. (2016). Mapping the customer journey: Lessons learned from graph-based online attribution modeling. *International Journal of Research in Marketing, 33*(3), 457–474.

Artz, M., & Doering, H. (2021). Exploiting Data from Field Experiments. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.

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

Bornemann, T., & Hattula, S. (2018). Experiments in Market Research. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.

Braun, M., & Moe, W. W. (2013). Online display advertising: Modeling the effects of multiple creatives and individual impression histories. *Marketing Science, 32*(5), 753–767.

Brodersen, K. H., & Hauser, A. (2021). CausalImpact. Available at https://cran.r-project.org/web/packages/CausalImpact/vignettes/CausalImpact.html

De Haan, E., Wiesel, T., & Pauwels, K. (2016). The effectiveness of different forms of online advertising for purchase conversion in a multiple-channel attribution framework. *International Journal of Research in Marketing, 33*(3), 491–507.

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

De Haan, E., Verhoef, P. C., & Wiesel, T. (2021). Customer feedback metrics for marketing accountability. *Review of Marketing Research*, forthcoming.

Ebbes, P., Papies, D., & Van Heerde, H. J. (2016). Dealing with Endogeneity: A Nontechnical Guide for Marketing Researchers. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.

Försch, S., & de Haan, E. (2018). Targeting online display ads: Choosing their frequency and spacing. *International Journal of Research in Marketing, 35*(4), 661–672.

Gupta, S., & Zeithaml, V. (2006). Customer metrics and their impact on financial performance. *Marketing Science, 25*(6), 718–739.

Gupta, S., Lehmann, D. R., & Stuart, J. A. (2004). Valuing customers. *Journal of Marketing Research, 41*(1), 7–18.

Hanssens, D. M. (2021). Return on Media Models. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.

Hitsch, G. J., & Misra, S. (2018). Heterogeneous treatment effects and optimal targeting policy evaluation. Available at SSRN, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3111957

Ho, D., Imai, K., King, G., Stuart, E., Whitworth, A., & Greifer, N. (2021). Package 'MatchIt'. Available at https://cran.r-project.org/web/packages/MatchIt/MatchIt.pdf

Hoban, P. R., & Bucklin, R. E. (2015). Effects of internet display advertising in the purchase funnel: Model-based insights from a randomized field experiment. *Journal of Marketing Research, 52*(3), 375–393.

Jackson, C. (2019). Package 'msm'. Available at https://cran.r-project.org/web/packages/msm/msm.pdf

Johnson, L. (2018). When Procter & Gamble Cut $200 Million in Digital Ad Spend, It Increased Its Reach 10%. Available at https://www.adweek.com/brand-marketing/when-procter-gamble-cut-200-million-in-digital-ad-spend-its-marketing-became-10-more-effective/

Kannan, P. K., Reinartz, W., & Verhoef, P. C. (2016). The path to purchase and attribution modeling: Introduction to special section.

Landwehr, J. R. (2019). Analysis of Variance. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.

Leeflang, P. S. H., Wieringa, J. E., Bijmolt, T. H. A., & Pauwels, K. (2015). *Modeling markets: Analyzing marketing phenomena and improving marketing decision making* (Ser. International series in quantitative marketing). Springer.

Leeflang, P. S. H., Wieringa, J. E., Bijmolt, T. H. A., & Pauwels, K. (2017). *Advanced methods for modeling markets* (Ser. International series in quantitative marketing). Springer.

Lesscher, L., Lobschat, L., & Verhoef, P. C. (2021). Do offline and online go hand in hand? Cross-channel and synergy effects of direct mailing and display advertising. *International Journal of Research in Marketing*, forthcoming.
Li, H., & Kannan, P. K. (2014). Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research, 51*(1), 40–56.
Li, J., Luo, X., Lu, X., & Moriguchi, T. (2021). The double-edged effects of E-commerce cart retargeting: Does retargeting too early backfire? *Journal of Marketing*, forthcoming.
Moe, W. W. (2003). Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of Consumer Psychology, 13*(1–2), 29–39.
Pan, W., & Bai, H. (Eds.). (2015). *Propensity score analysis: Fundamentals and developments*. Guilford Publications.
Pauwels, K., Demirci, C., Yildirim, G., & Srinivasan, S. (2016a). The impact of brand familiarity on online and offline media synergy. *International Journal of Research in Marketing, 33*(4), 739–753.
Pauwels, K., Aksehirli, Z., & Lackman, A. (2016b). Like the ad or the brand? Marketing stimulates different electronic word-of-mouth content to drive online and offline performance. *International Journal of Research in Marketing, 33*(3), 639–655.
Rubin, D. B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology, 2*(3), 169–188.
Rutz, O. J., & Bucklin, R. E. (2011). From generic to branded: A model of spillover in paid search advertising. *Journal of Marketing Research, 48*(1), 87–102.
Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games, 2*(28), 307–317.
Skiera, B., Reiner, J., & Albers, S. (2018). Regression Analysis. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.
Spedicato, G. A. (2021). Package 'markovchain'. Available at https://cran.r-project.org/web/packages/markovchain/markovchain.pdf
Spedicato, G. A., Kang, T. S., Yalamanchi, S. B., Yadav, D., & Cordón, I. (2016). The markovchain package: A package for easily handling Discrete Markov Chains in R. https://cran.r-project.org/web/packages/markovchain/vignettes/an_introduction_to_markovchain_package.pdf
Srinivasan, S. (2021). Modeling Marketing Dynamics Using Vector Autoregressive (VAR) Models. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.
Srinivasan, S., Vanhuele, M., & Pauwels, K. (2010). Mind-set metrics in market response models: An integrative approach. *Journal of Marketing Research, 47*(4), 672–684.
Srinivasan, S., Rutz, O. J., & Pauwels, K. (2016). Paths to and off purchase: Quantifying the impact of traditional marketing and online consumer activity. *Journal of the Academy of Marketing Science, 44*(4), 440–453.
Tillmanns, S., & Krafft, M. (2017). Logistic Regression and Discriminant Analysis. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.
Trusov, M., Bucklin, R. E., & Pauwels, K. (2009). Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of Marketing, 73*(5), 90–102.
WARC. (2021). Uber's former performance chief charts extent of ad fraud. Available at https://www.warc.com/newsandopinion/news/ubers-former-performance-chief-charts-extent-of-ad-fraud/44527
Wiesel, T., Pauwels, K., & Arts, J. (2011). Practice prize paper – Marketing's profit impact: Quantifying online and off-line funnel progression. *Marketing Science, 30*(4), 604–611.
Valli, V., Stahl, F., & Feit, E. M. (2017). Field Experiments. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.
Wang, W., & Yildirim, G. (2021). Applied Time-Series Analysis in Marketing. In C. Homburg, M. Klarmann, A. E. Vomberg (Eds.). Handbook of Market Research. Springer, Cham.

Article

AMERICAN MARKETING ASSOCIATION

Journal of Marketing Research
2021, Vol. 58(6) 1101–1119

DOI: 10.1177/0022243720940693
journals.sagepub.com/home/mrj
SAGE

Construal Matching in Online Search: Applying Text Analysis to Illuminate the Consumer Decision Journey

Ashlee Humphreys, Mathew S. Isaac, and Rebecca Jen-Hui Wang

Abstract

As consumers move through their decision journey, they adopt different goals (e.g., transactional vs. informational). In this research, the authors propose that consumer goals can be detected through textual analysis of online search queries and that both marketers and consumers can benefit when paid search results and advertisements match consumer search–related goals. In bridging construal level theory and textual analysis, the authors show that consumers at different stages of the decision journey tend to assume different levels of mental construal, or mindsets (i.e., abstract vs. concrete). They find evidence of a fluency-driven matching effect in online search such that when consumer mindsets are more abstract (more concrete), consumers generate textual search queries that use more abstract (more concrete) language. Furthermore, they are more likely to click on search engine results and ad content that matches their mindset, thereby experiencing more search satisfaction and perceiving greater goal progress. Six empirical studies, including a pilot study, a survey, three lab experiments, and a field experiment involving over 128,000 ad impressions provide support for this construal matching effect in online search.

Keywords

consumer decision journey, construal level, abstract vs. concrete mindsets, online search, processing fluency, goal progress, search queries

Online supplement: https://doi.org/10.1177/0022243720940693

Since their inception in 1995, search engines have provided a means of communication that allows companies to organize and diffuse information efficiently to customers. Monetization of search engine result pages began with simple yellow-page-style listings online. Today, paid search and advertising follow the cost-per-click auction model, which requires firms to bid on keywords, displaying their ad upon winning the auction. According to eMarketer, companies will spend $385 billion on digital advertising in 2020 (Enberg 2019).

Although digital advertising has become more prominent in a firm's marketing mix, most changes in marketing practice are taking place with respect to keyword bidding and metrics development. The specific ad content that is served following a successful bid, however, has remained static. That is, once a firm wins a keyword auction, it serves the same ad copy regardless of the consumer's goal or stage in the decision process.

Furthermore, practitioners currently tend to buy search advertising and craft ad messages that appeal to users who are near the purchase stage of the consumer decision journey (e.g., Batra and Keller 2016; Lemon and Verhoef 2016) or the bottom of the purchase funnel (e.g., Hoban and Bucklin 2015). This means that ad copy is often created with concrete calls to action (e.g., "buy now!"), and managers set high bid prices for action-oriented words such as "buy." Our contention is that there is valuable and underutilized space at the early stages of the consumer decision journey, a space where many users are conducting queries that are informational or exploratory rather than transactional in nature. If managers are able to accurately assess where consumers are situated in the decision-making process, they should be able to serve content in digital ads and paid search results that matches consumers' specific goals at that stage of the consumer journey. Conversely, if there is a mismatch between consumers' search goals and ad content, consumers may be less inclined to click on the

Ashlee Humphreys is Associate Professor, Medill School of Journalism, Media, and Integrated Marketing Communications, and Kellogg School of Management, Northwestern University, USA (email: a-humphreys@northwestern.edu). Mathew S. Isaac is Genevieve Albers Professor of Marketing, Albers School of Business and Economics, Seattle University, USA (email: isaacm@seattleu.edu). Rebecca Jen-Hui Wang is Assistant Professor of Marketing, College of Business, Lehigh University, USA (email: rwang@lehigh.edu).

advertised content at all, which could cost the brand a chance to engage with potential customers in meaningful and valuable ways.

Building on the premise that both marketers and consumers may be advantaged if firms serve content that matches consumer search–related goals, the present research incorporates construal level theory (Liberman, Trope, and Stephan 2007; Trope and Liberman 2003; Trope, Liberman, and Wakslak 2007) and text analysis (e.g., Berger et al. 2020; Humphreys and Wang 2017) to improve outcomes in an online search context. We find that consumers at different stages in the consumer journey (i.e., informational vs. transactional) tend to also assume different levels of mental construal, or mindsets (i.e., abstract vs. concrete), which can be detected based on the language used in their search queries. When consumer mindsets are more abstract (more concrete), consumers generate textual search queries that are also more abstract (more concrete), and they respond more favorably to search engine results and ad content that match their mindset. Six empirical studies, including a pilot study, a survey, three lab experiments, and a field experiment involving over 128,000 ad impressions, provide strong support for this construal matching effect in online search.

Our work makes three contributions to research on the role of search in the consumer journey. First, we demonstrate the influence of construal level in online search and show that firms can increase click-through rates and search satisfaction by matching ad copy or their paid search result to a consumer's mindset. While previous research has documented the effects of construal matching in a variety of other contexts, this is the first work to show that marketers can influence perceptions of progress along the consumer decision journey by aligning paid content with consumers' mindsets. Second, we illustrate the important substantive consequences of matching construal for improving the consumer search experience and making search advertising more cost efficient and effective for firms. Specifically, this work outlines a scalable method for firms to identify consumers' search goals in text and facilitate their progression through the consumer journey. Finally, we highlight the contextual and changing nature of the decision journey itself. As noted by Hamilton and Price (2019), in some contexts, decision journeys may be "motivated by concrete goals, such as getting medications," while in others, consumers may be motivated by "more abstract goals, such as getting healthy and feeling good again" (p. 188). Although consumers' goals vary over time, in duration, and across contexts, their search queries can offer useful clues to marketers about their stage in the consumer decision journey. By outlining a straightforward and implementable process by which consumers' mindsets can be reliably categorized on the basis of their own search queries, we add to the burgeoning literature in marketing that shows how text analysis can provide nuanced insight into consumer needs and goals. Digital marketers can leverage this newfound knowledge to serve content that corresponds with consumers' construal level, resulting in greater consumer engagement at a lower cost to advertisers.

Conceptual Background

Psychological Distance

Our predictions for how online search queries will change as consumers move through the decision journey are derived from construal level theory, which proposes a fundamental link between the psychological distance separating an individual from a focal object and one's mental construal level. First introduced by Liberman and Trope (1998), this broad social psychology theory states that when objects, people, ideas, or goals are psychologically distant versus near, they are represented at different levels of mental construal. Trope and Liberman (2000) further posit that psychological distance may refer to spatial distance, temporal distance, social distance, or hypotheticality, and the link between psychological distance and construal level generalizes across all four dimensions.

In describing the relationship between psychological distance and construal, Trope, Liberman, and Wakslak (2007, p. 83) explain, "[Construal level theory] assumes that people mentally construe objects that are psychologically near in terms of low-level, detailed, and contextualized features, whereas at a distance they construe the same objects or events in terms of high-level, abstract, and stable characteristics." Low-level construals, or concrete mindsets, lead individuals to focus on the local, subordinate, and secondary features of an event, action, or goal. In contrast, high-level construals, or abstract mindsets, emphasize global, superordinate, and primary features (Fujita et al. 2006). For example, the goal of "studying" is more likely to be represented concretely as "reading a textbook" when it is proximal but abstractly as "doing well in school" when it is distal (Trope and Liberman 2010).

As consumers proceed in their journey, their purchase goal is likely to seem psychologically closer. Indeed, prior work has noted differences between the stages of browsing (a top-of-the-funnel activity) and buying (a bottom-of-the-funnel activity) that likely relate to psychological distance. For example, consumers in the browsing stage are likely to be "interested in gathering information for possible future use but not for the immediate purpose of picking one option to put into their shopping basket" (Hamilton and Chernev 2010, p. 52). Instead of having an informational goal, consumers in the buying stage are likely to have a transactional goal and "seek information for use in an immediate purchase decision" (Hamilton and Chernev 2010, p. 52). As psychological distance changes during their decision journey, consumers' construal levels are likely to shift as well.

Prior work has divided the decision journey into stages, be it 3 stages—prepurchase, purchase, and postpurchase—or as many as 12 stages (Batra and Keller 2016; Lemon and Verhoef 2016). While these stages may be analytically distinct, they may not always be empirically distinguishable, universal, or linear (Hamilton and Price 2019). Regardless of the precise number of stages, we expect that consumers will have different goals as they progress through their journey and move from one stage to another. These goals, in turn, are likely to vary in construal, and we posit that differences in consumer mindset can be detected in written text. Our theorizing is consistent with a conceptualization of the consumer decision journey as a spectrum that varies in terms of psychological distance from point of purchase, decision, or implementation. We acknowledge there may also be considerable individual and contextual difference in how consumers approach and advance through the journey (Mende et al. 2019; Nakata et al. 2019). Given this variation, it is even more vital for firms to be able to detect and adaptively respond to consumers who have different mindsets because they are at different stages in their journey.

Our proposition is therefore that consumers' construal level (or mindset) will be more abstract when they are at an earlier stage of the decision journey and more concrete when they are at a later stage (i.e., closer to purchase or implementation). For example, consumers who are about to purchase a washing machine may be more concerned with its specific product attributes (e.g., capacity, energy efficiency, location of purchase), whereas those who are further away from this point of purchase may focus more on understanding general product benefits or gathering information about the washing machine category as a whole (e.g., evaluating trade-offs between front-loading and top-loading models, exploring advantages/disadvantages of different brands). Although these predictions align with construal level theory, marketing researchers have not examined the consequences of construal matching on perceived goal progress, search satisfaction, and decision making in the context of the consumer journey.

Construal Level and Language Usage

If consumers' psychological distance from the point of purchase determines whether they assume a more abstract versus a more concrete mindset, this may influence the language that they use when conducting online searches. A number of empirical studies, albeit primarily in offline contexts, have shown that psychological distance exerts influence on language use. When communicating with larger (vs. smaller) audiences (i.e., when psychological distance is likely to be greater), Joshi and Wakslak (2014) find that speakers are more likely to use abstract (vs. concrete) language. Palmeira (2015) observes an association between power and abstract speech, as well as an association between action orientation and concrete speech. Stephan, Liberman, and Trope (2010) show that participants who adopt an abstract (vs. concrete) mindset use more polite and formal language when addressing another person, presumably because there is greater psychological distance between the two parties. In a marketing-relevant context, Hansen and Wänke (2011) demonstrate that consumers and advertisers tend to describe luxury products, which are more psychologically distant than ordinary products, with more abstract language.

Although fewer studies have examined the impact of psychological distance on language usage in online contexts, several recent articles have examined how construal level affects consumers' language when they use social media platforms or post online reviews. For instance, Snefjella and Kuperman (2015) show that Twitter users employ different language when discussing near versus far events. Huang et al. (2016) find evidence for a "distance boosting effect," in which online restaurant reviews contain more abstract language (and tend to be more positive) when the distance between the reviewer and the restaurant is greater, either spatially (i.e., the restaurant is geographically distant from the reviewer) or temporally (i.e., there is a long delay between the restaurant visit and the date the review was submitted online). In the context of online search behavior, we similarly predict that consumers' level of construal (as determined by their psychological distance from the point of purchase) will be reflected in the language of their search queries. Specifically, we hypothesize the following:

> $\mathbf{H_1}$ (language use in search queries): Consumers in the informational stage of the consumer journey use language in their search queries that is less concrete (more abstract) than consumers who are in the transactional stage.

Construal Matching Effects in Text Processing

We expect construal level to affect not only the search queries that consumers themselves generate but also the choices and decisions that they subsequently make online (e.g., the specific search result or digital ad that they click). This prediction is consistent with a large body of literature that has shown that mindsets shape consumer evaluations and behaviors in a variety of domains. For instance, abstract construal is known to alter consumer decisions by shifting attention away from low-level comparative details of products (Khan, Zhu, and Kalra 2011). Other research has found that consumers' savings decisions are jointly affected by their level of construal and the specificity of their savings goal (Ülkümen and Cheema 2011). Pham, Hung, and Gorn (2011) even demonstrate that when consumers are in a state of relaxation, they are more likely to adopt an abstract mindset, which in turn causes them to provide higher valuations for identical products.

In the present research, we hypothesize that when conducting an online search, consumers will be more likely to click on served marketing content (e.g., search results or digital ads) that matches their specific construal level. Although this proposition has never been tested in the context of online search, it is consistent with a large body of research showing that congruence between one's mindset and other psychological constructs affects decision making. Previous work has shown that an abstract (concrete) mindset is perceived to match with promotion-oriented (prevention-oriented) regulatory foci (Lee, Keller, and Sternthal 2010), desirability (feasibility) considerations (Liberman and Trope 1998), and independent (interdependent) self-views (Spassova and Lee 2013). Relative to a mismatch, a match between one's mindset and another psychological construct has positive effects on consumer attitudes and behaviors, such as by increasing receptivity to persuasive messages (Lee, Keller, and Sternthal 2010) or by boosting a product's appeal (Spassova and Lee 2013).

These matching effects have been demonstrated not only at the broad construct level but also at the level of the specific language used in persuasive communication attempts. For example, abstract appeals that emphasize the word "why" versus concrete appeals laden with the word "how" are more persuasive to voters whose decision is temporally distant, but the opposite is true if the voting decision is temporally near (Kim, Rao, and Lee 2009). In addition, individuals are more likely to judge concrete statements as being true when they have been primed with a concrete (vs. abstract) mindset prior to reading the statements (Hansen and Wänke 2010). Furthermore, consumers in an abstract (concrete) mindset are more willing to engage in recycling behavior after encountering a recycling appeal that uses a matching gain (loss) frame (White, MacDonnell, and Dahl 2011).

Matching has been described as an "it just feels right" experience with generally positive behavioral consequences (Camacho, Higgins, and Luger 2003; Lee and Aaker 2004, p. 212). For example, in proposing their "fit from construal" hypothesis, Lee, Keller, and Sternthal (2010, pp. 735–36) explain that construal matching "stimulates a subjective experience of engagement. This experience creates a motivational force that absorbs and engrosses people. In the context of a persuasive message, engagement is thought to intensify processing of the advocacy and thus positive reactions to it." Drawing on these principles, we predict that when searching online, congruence between a consumer's construal level (as determined by their search stage/goal) and the abstractness or concreteness of the language in an advertisement or a search result will increase consumers' propensity to click on the advertiser's content and enhance their search satisfaction.

> $\mathbf{H_2}$ (construal matching effect in online search): Consumers will be more satisfied with and likely to click on search results or advertising content that matches their construal level, and they will be less satisfied with and less likely to click on search results or advertising content that does not match their construal level.
>
> $\mathbf{H_{2a}}$: In the informational stage, consumers will be more satisfied with and likely to click on abstract results.
> $\mathbf{H_{2b}}$: In the transactional stage, consumers will be more satisfied with and likely to click on concrete results.

We further posit that the positive consequences of construal matching will carry over to perceptions of goal progress. Several studies have found, for instance, that people perceive feelings of movement toward a goal on account of fit between their goal orientation and external stimuli that matches this orientation (Higgins et al. 2003; Higgins, Kruglanski, and Pierro 2003; Kruglanski, Pierro, and Thompson 2000; Malaviya and Sternthal 2009).

> $\mathbf{H_3}$ (construal matching and goal progress): When consumers receive search results or advertising content that match their construal level, they will perceive greater goal progress than when they receive content that does not match their construal.
>
> $\mathbf{H_{3a}}$: In the informational stage, consumers will perceive greater goal progress when they receive abstract results.
>
> $\mathbf{H_{3b}}$: In the transactional stage, consumers will perceive greater goal progress when they receive concrete results.

Various psychological mechanisms have been proposed to explain matching effects in other domains, but most accounts implicate the construct of processing fluency (e.g., Higgins 2000; Reber, Schwarz, and Winkielman 2004; Reber, Wurtz, and Zimmermann 2004). Processing fluency (which can be either perceptual or conceptual in nature) reflects the ease of mental operations required to assign meaning to a stimulus and has been conceptualized as "a continuum from effortless to highly effortful" (Alter and Oppenheimer 2009, p. 220) along which every cognitive task can be described. Research on conceptual fluency (e.g., Alter and Oppenheimer 2009; Lee and Labroo 2004) indicates that it often arises from matching effects, such as "when a product is presented in a predictive context or when it is primed by a related construct" (Graf, Mayer, and Landwehr 2018, p. 395). Indeed, fluent (vs. disfluent) stimuli or experiences evoke feelings of metacognitive ease, which may be attributed to the target that is being evaluated (e.g., Reber, Schwarz, and Winkielman 2004) and consequently bolster evaluations. As such, we hypothesize that construal matching will yield processing fluency, which in turn will increase search satisfaction and perceived goal progress.

> $\mathbf{H_4}$ (processing fluency): Processing fluency mediates the effect of construal matching on both search satisfaction and perceived goal progress.

Overview of Studies

In a series of studies that combine archival, laboratory, and field data, we test these hypotheses and explore the relationship between search queries, construal level, and behavioral outcomes. Given that a fundamental premise of our theorizing is that consumers' search queries will vary in their level of abstractness/concreteness, we first conducted a pilot study in conjunction with the Intent Lab (a collaborative research unit established between academia and the digital marketing firm Performics). This study used archival data to verify that consumers do in fact express a range of construal in their search queries. High variability is not a foregone conclusion, particularly given that the concrete calls to action favored by most digital advertising campaigns seem to assume that online searchers are predominantly at the bottom of the purchase funnel.

For the pilot study, we acquired a data set of searches on 13 consumer packaged goods (CPG) brands across a range of categories. Using a dictionary developed by Paetzold and Specia (2016) and adapted from the Medical Research Council (MRC) machine readable dictionary (Coltheart 1981; Wilson

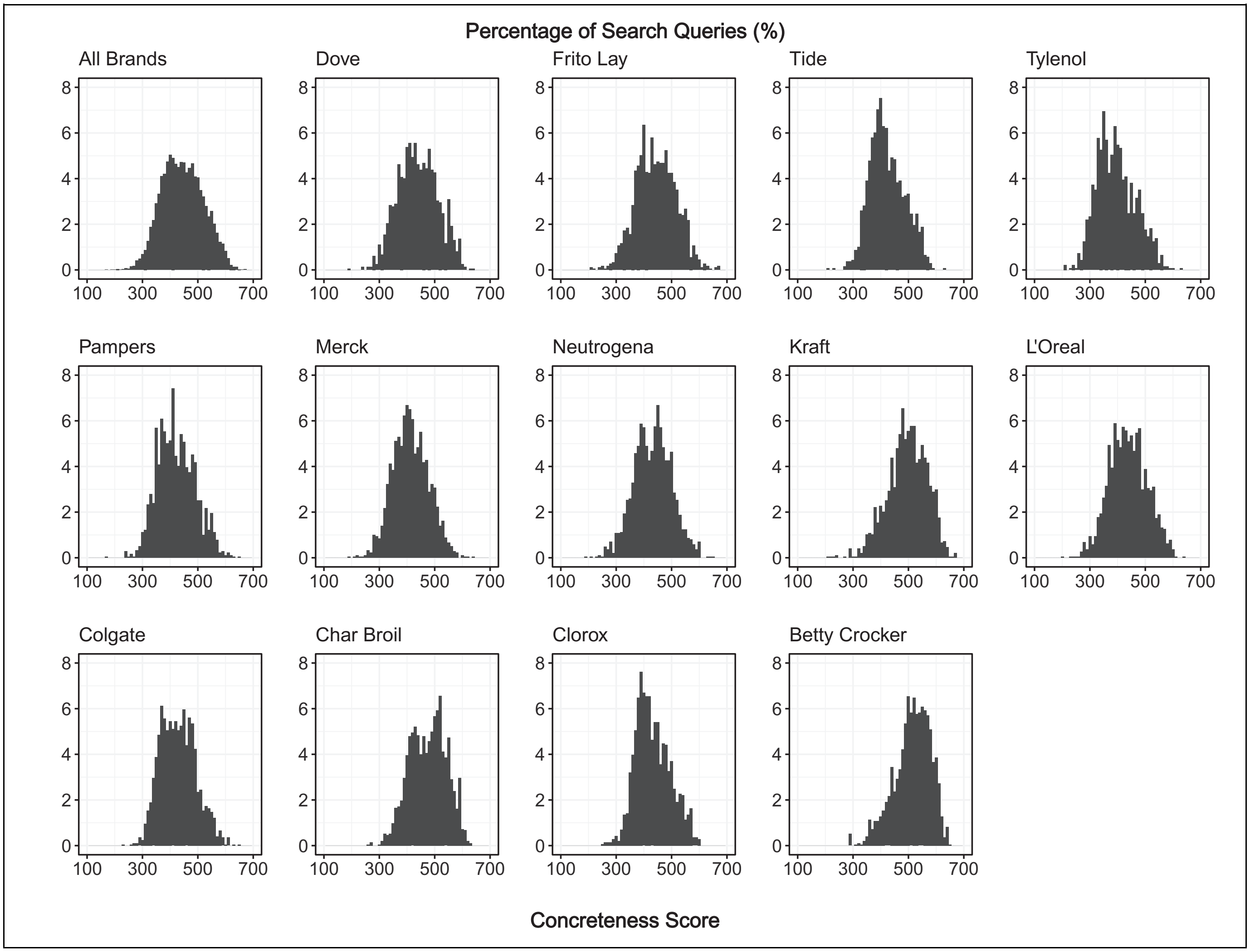

Figure 1. Pilot study: Distributions of search text concreteness scores by brand.
Notes: N = 22,675 distinct queries; 1,364 queries were dropped because searches contained only brand names or words that were not found in the MRC concreteness dictionary.

1988),[1] we calculated a concreteness score (range: 100 = abstract, 700 = concrete) for each of 24,039 unique search queries that drove traffic to 13 websites of leading CPG brands (for the list of brands, see Figure 1). We also developed an online tool at www.construalscore.com for academic researchers and marketing practitioners to quantify the concreteness of written communication using either the updated MRC dictionary or another dictionary developed by Brysbaert, Warriner, and Kuperman (2014). For additional information about these dictionaries and our efforts to establish the construct validity of our concreteness score metric, please see Web Appendix A.

Relying on the updated MRC dictionary, we plotted the concreteness score distribution of these queries, as shown in Figure 1. This figure illustrates that the mean concreteness score of search queries—at least those that ultimately send searchers to CPG-branded websites—tends to be near the scale midpoint (mean concreteness score = 444.72), but variability exists in the abstractness/concreteness of language used by consumers when performing an online search (SD = 74.347, range: [167.34, 668.76]). We observed this variation not only for queries within the same CPG brand query but also across CPG brands. For instance, queries that led consumers to health-related websites (e.g., Merck, Tylenol) were relatively abstract, whereas food-based categories (e.g., Kraft, Betty Crocker) were relatively concrete. These between-category differences suggest that consumer goals (and corresponding mindsets) may

[1] The original MRC dictionary (Coltheart 1981; Wilson 1988) contained concreteness scores for 8,228 words. Paetzold and Specia (2016) extended this original dictionary using a bootstrapping algorithm to calculate concreteness scores for 85,863 words. In this updated MRC concreteness dictionary, words typically have concreteness scores ranging from 100 to 700 (100 = abstract, 700 = concrete; e.g., "best" = 260.28, "sick" = 397.48, "throat" = 625.49). In analyzing the pilot study data set, we first removed all brand names from the queries and then calculated the concreteness of each search query as the mean of all words in the text that appeared in the updated MRC dictionary (Paetzold and Specia 2016).

be different in each context, depending on the journey (e.g., learning about a new drug vs. finding a recipe for dinner tonight). These findings have important managerial implications because they imply that certain easily tracked linguistic characteristics of online queries may provide useful signals about a consumer's construal level (and associated search goal) and potentially cover a range of concreteness and goals. For additional analyses and robustness checks related to the pilot study, please see Web Appendix B.

Having established that variation exists in the abstractness/concreteness of language used by consumers when performing an online search, our next step is to link the consumer decision journey with language concreteness to assess the downstream behavioral consequences of psychological distance (Study 1), and then to test the construal matching hypothesis in a series of controlled experiments (Studies 2–5).

Study 1 is a survey that investigates how consumers' search goals influence the concreteness of their online searches and the type of website that they ultimately visit. In Study 2, we examine whether consumers are more likely to click on content having abstract (concrete) language when their construal level is also abstract (concrete), which we refer to as the construal matching effect in online search. Studies 3 and 4 explore construal at different decision stages and demonstrate that matching can affect search satisfaction and perceived goal progress; Study 4 also implicates processing fluency as a causal mechanism. Finally, Study 5 establishes the external validity of this phenomenon through a large-scale field experiment that tests for construal matching in actual online search behavior (i.e., advertising click-through rates) and demonstrates the benefits to the firm. In summary, these studies offer novel insights on the causal links between consumer search goals, the concreteness of online searches, and preferences for different language and content.

Study 1

Study 1 is a survey that examines the link between participants' stage in the decision journey and their natural search queries. We ask participants to reflect on the specific goal that resulted in their last visit to a search engine. By collecting information about consumers' most recent search query, we can then assess whether consumers use language in their search queries that is more or less concrete (abstract) depending on their stage in the decision journey (H_1). Because we also ask participants to reveal the first website they visited following their query, Study 1 enables us to begin exploring whether consumers' stage in the decision journey predicts their search query concreteness and, in turn, their downstream click behavior.

Method

Study 1 was conducted online with 369 students at a large public university in the United States (60.4% female; M_{age} = 20.05 years, SD = 1.05) who participated in exchange for course credit. All participants were asked to reflect on the last time they had used a search engine. In case they did not remember their last online search, participants were given detailed information on how to access their recent search queries by using the "History" feature on their web browser. In an open-ended text box, participants were asked to specify in their own words the goal that they were trying to accomplish when they last used a search engine. Subsequently, participants classified this most recent search goal as informational (e.g., focused on browsing and learning) or transactional (e.g., focused on buying) along a seven-point bipolar scale (1 = "informational," and 7 = "transactional"). Next, in an open-ended text box, participants relayed the exact wording of the most recent search query that they had entered. We calculated the mean concreteness score of the words contained in participants' exact search query using the concreteness dictionary adapted from the MRC psycholinguistic database (Packard and Berger 2020; Paetzold and Specia 2016).[2] The usual range of MRC scores is from 100 to 700, with higher numbers indicating greater concreteness. Finally, participants provided the URL of the first website they visited as a result of their search and indicated on seven-point scales (1 = "not at all," and 7 = "very much") the extent to which this site was as an e-commerce website, a news website, or a corporate website (brief descriptions of each type of website were provided).[3] We reasoned that consumers at a later stage of the consumer journey would be more likely to visit e-commerce websites (where they could make a purchase) and even corporate websites (where they might be able to make a purchase) than those who were at an early stage in the journey.

Results

We first tested whether participants near the end of the decision journey executed more concrete searches than participants near the beginning of their journey. Specifically, we conducted a linear regression to predict search query concreteness based on participants' stage in the decision journey, as indicated by their rating on the seven-point bipolar scale (1 = "highly informational," and 7 = "highly transactional"). We obtained a significant result (B = 6.77, SE = 2.11, $p = .001$, 95% confidence interval [CI] = [2.62, 10.91]), which is consistent with H_1.

Next, we examined whether participants near the end of the decision journey visited different types of websites than participants at the beginning of their journey. Specifically, we conducted separate regressions to test whether participants' stage in the decision journey predicted their likelihood to classify their most recently visited website as an e-commerce website,

[2] Forty-five participants provided goals or search queries for which no words appeared in the updated MRC dictionary.

[3] E-commerce websites were described as sites where "buyers engage with sellers and actions such as ordering, payments, and shipping take place." News websites were described as sites offering "the latest news in entertainment, weather, and sports. They offer breaking news, news update concerning the country/the world and overall global news." Corporate websites were described as sites that "provide information about a particular company, brand, and/or products to the public. These websites may or may not have e-commerce functionality."

a corporate website, or a news website. Participants' stage in the decision journey positively predicted the extent to which they had visited an e-commerce website (B = .71, SE = .05, $p < .001$, 95% CI = [.61, .82]) and a corporate website (B = .32, SE = .06, $p < .001$, 95% CI = [.20, .44]). However, participants' stage in the decision journey did not significantly predict whether they had visited a news website (B = –.08, SE = .06, $p = .21$, 95% CI = [−.20, .04]).

Finally, we ran three separate mediations using the PROCESS macro developed by Hayes (2017) to examine whether concreteness of the language used in their search query mediated the relationship between participants' decision journey stage and their propensity to visit an e-commerce website, a corporate website, and a news website. Each mediation analysis utilized bootstrapping with repeated extraction of 10,000 samples (Hayes 2017, model 4). We found a significant indirect effect of decision stage on participants' relative preference for an e-commerce website through search query concreteness (B = .03, SE = .02, 95% CI = [.01, .06]). Similarly, we found a significant indirect effect of decision stage on participants' relative preference for a corporate website through search query concreteness (B = .03, SE = .02, 95% CI = [.01, .07]). However, we did not find an indirect effect of decision stage on participants' relative preference for a news website through search query concreteness (B = −.004, SE = .01, 95% CI = [−.03, .02]). Taken together, these results suggest that search query concreteness plays an important role in explaining the relationship between consumers' stage in the decision journey and their downstream click behavior.

Discussion

The results of Study 1 provide support for H_1. We found that consumers who had a transactional goal executed search queries that were more concrete (less abstract) than participants who had an informational goal. Furthermore, Study 1 provides correlational evidence that consumers at an earlier stage of the decision journey (i.e., who have informational goals) are relatively less likely to visit e-commerce websites and corporate websites than consumers at a later stage (i.e., who have transactional goals). Importantly, this behavioral consequence is mediated by the concreteness of their query.

Study 1 suggests that search query concreteness provides an important signal of consumers' stage in the decision journey. One reason that e-commerce sites and corporate websites may be particularly attractive to consumers who are in a transactional (vs. informational) stage of the consumer journey is because the language used in these sites' display ads and paid search results may also be relatively concrete. In our next study, we directly test for this construal matching effect in online search using a controlled laboratory experiment.

Study 2

In addition to reexamining H_1 in a controlled experiment that enables us to more readily make causal inferences, the purpose of Study 2 is to test our proposition that the congruence between a consumer's stage in the decision journey and the abstractness or concreteness of the language in an advertisement or a search result will increase consumers' propensity to click on the advertiser's content (H_2). In each of our lab experiments (Studies 2–4), participants were prompted to provide a search query related to a specific goal or task. We applied the same exclusion criteria in each experiment, consistently removing participants whose queries were missing values or contained language unrelated to the task at hand (as determined by a coder blind to experimental condition). The sample size reported in each experiment reflects the number of participants who met these criteria. For the stimuli used in each experiment, see Web Appendix C.

Method

Study 2 was conducted with 490 U.S. participants (57.7% female; $M_{age} = 36.76$ years, SD = 13.20) recruited using an online panel (Amazon Mechanical Turk Prime). We employed a 2 (consumer journey stage: informational vs. transactional) × 2 (search results: abstract vs. concrete) between-participants design.

All participants learned that they had been thinking about repainting their living room. Those assigned to the informational stage condition were advised that they wanted to know general strategies or ideas for repainting the room. In contrast, participants in the transactional stage condition were informed that they wanted to know specific techniques for repainting the room.

Subsequently, participants were asked to enter the exact search query that they would type into a search box if they consulted a search engine (e.g., Google, Bing) to help them with this task. We measured the mean concreteness score of each participant's search query using the updated MRC dictionary.

After reflecting on how they might use a search engine (e.g., Google, Bing) to help them with their task, participants were shown a single search result and informed that this was the top result returned by the search engine following their search. The result consisted of a title and a snippet (a brief description of the page appearing below the title). We manipulated the search result between participants that it contained language that was either predominantly abstract or predominantly concrete. Specifically, the abstract search result was titled "Making the Room Look Fresh—Repainting Your Living Room" and the accompanying snippet read, "Learn how a new coat of paint can freshen and cheer up a room" (concreteness score = 394.86 [title only], 387.84 [with snippet]). The concrete search result was titled "Applying Brush Strokes—Repainting Your Living Room" and the accompanying snippet read, "Learn ways to paint a room in your home with clear instructions" (concreteness score = 451.24 [title only], 403.11 [with snippet]).

The specific language used in the abstract and concrete search result conditions was taken directly from an established construal level operationalization, the Behavioral Identification Form (BIF; Vallacher and Wegner 1989). The BIF requires participants to classify a series of behaviors (e.g., eating) in one of two ways, either corresponding to a more abstract and high-level representation (e.g., "getting nutrition") or a more concrete and low-level representation (e.g., "chewing and swallowing"). In this study, we selected one behavior from the BIF (painting a room) that was described in abstract terms ("making the room look fresh"; concreteness score = 381.25) versus concrete terms ("applying brush strokes"; concreteness score = 487.60), and our search result titles retained the same wording used in the BIF.

Subsequently, participants were asked to indicate whether they would click on the specific search result they had been shown or if they would continue to search (presumably for a more suitable search result). Our key dependent variable was the proportion of participants in each condition who elected to click on the search result that they encountered.

Results

Our prediction is that consumers in the informational stage of the consumer decision journey will use more abstract language in their search queries whereas consumers in the transactional stage will use more concrete language. We tested this prediction by conducting a 2 (consumer journey stage) $\times$ 2 (search results) between-participants analysis of variance (ANOVA) on the mean concreteness score of their search queries. As expected, we observed a main effect of consumer journey stage ($F(1, 486) = 4.69$, $p = .031$, $\eta_p^2 = .010$), such that participants in the informational stage ($M = 424.44$, $SD = 53.99$, $N = 268$) executed searches that were less concrete (more abstract) than participants in the transactional stage ($M = 434.97$, $SD = 52.69$, $N = 222$). This finding is consistent with H_1.

To test for our matching effect, we conducted a binary logistic regression of consumer journey stage (-1 = informational, $+1$ = transactional), search results (-1 = abstract, $+1$ = concrete), and their interaction on participants' choice to click on the result provided versus keep searching (0 = keep searching, 1 = click). Neither the main effect of journey stage ($B = .10$, $SE = .10$, $Wald = 1.16$, $p > .28$) nor search results ($B = -.01$, $SE = .10$, $Wald = .004$, $p > .94$) was significant. However, as expected, we obtained a significant interaction ($B = .64$, $SE = .10$, $Wald = 44.58$, $p < .001$), indicating that when the consumer journey stage matched the concreteness of the search result, participants were more likely to choose the search result. Among participants in the informational stage condition, 68.1% of participants chose the abstract search result, whereas just 36.8% of participants chose the concrete search result ($\chi^2(1, N = 268) = 26.34$, $p < .001$). Furthermore, among participants in the transactional stage condition, 42.1% of participants chose the abstract search result, but 72.2% of participants chose the concrete search result ($\chi^2(1, N = 222) = 20.50$; $p < .001$). These results are indicative of our proposed construal matching effect in online search.

Discussion

The results of Study 2 provide strong support for H_1 and H_2. Unlike Study 1, where consumer journey stage was a measured (and self-reported) variable, Study 2 manipulated consumer journey stage. Yet we found a converging pattern of results whereby consumers who had a transactional goal executed search queries that were more concrete (less abstract) than participants who had an informational goal (H_1). Furthermore, we obtained evidence for our proposed construal matching effect (H_2) by showing that consumers are more likely to click on a search result that matches (vs. does not match) their construal level.

Thus far, we have mainly examined the endpoints of the consumer decision journey (i.e., informational stage vs. transactional stage). In our next study, we introduce a third stage (i.e., comparison stage) and examine whether the search queries and preferences of consumers in this intermediate stage differ from consumers in the other two stages.

Study 3

The purpose of Study 3 is to assess how construal level varies for multiple stages of the decision journey. As we have discussed, consumers may have different goals at each stage of the process. For example, during the informational stage, we have shown that consumers tend to display an abstract construal. During the transactional stage, they assume a more concrete construal. However, for an intermediate stage such as comparison, we do not know how construal will vary and/or if it will be distinguishable from stages at either end of the journey. The purpose of Study 3 is to empirically answer this question. Thus, Study 3 allows us to again validate H_1 and H_2 while also evaluating whether consumers at an intermediate comparison stage conduct search queries and exhibit click behaviors that differ from consumers at the informational or transactional stages.

Method

Study 3 was conducted with 357 U.S. participants (65.3% female; $M_{age} = 58.23$ years, $SD = 13.03$) recruited by the market research firm Dynata to complete a series of unrelated online studies. Participants were informed that they were considering purchasing a laptop computer. They learned that consumers tend to progress through three sequential stages when making a purchase decision and were provided with a diagram to review that described each stage (see Web Appendix C). The diagram depicted three chevrons, with Stage 1 corresponding to an initial informational stage, Stage 2 corresponding to an intermediate comparison stage, and Stage 3 corresponding to a final transactional stage. Participants were randomly assigned to one of these three stages and instructed to adopt the perspective of a decision maker in that particular stage of the consumer journey.

Subsequently, participants were asked to conduct an online search pertaining to their decision stage by entering their exact query into an open-ended text box. We measured the mean concreteness score of each participant's search query using the adapted MRC concreteness dictionary. Next, participants were instructed to open a new tab in their browser window and to visit the search engine of their choice (e.g., Google, Bing). They were asked to enter their search query into the search box of an actual search engine; to refresh their memory, participants' previously stated query (that they had provided earlier in the study) was displayed on the screen. Participants were directed to copy and paste the complete title and description of the actual result that they would be most likely to click on. We calculated the mean concreteness score for the title and description of this "preferred" result, again using the MRC concreteness dictionary.

Finally, participants were shown three search result descriptions for a popular laptop/tablet (i.e., Microsoft Surface) and asked to indicate the one that they would be most likely to click on. The descriptions varied in terms of construal level, with one being relatively abstract (i.e., "see what's new from Surface"; concreteness score = 307.15), one intermediate ("which Surface works for you?"; concreteness score = 362.93), and one being relatively concrete (i.e., "free next day delivery"; concreteness score = 390.93). For study stimuli, see Web Appendix C.[4]

Results

To confirm that our consumer journey manipulation had successfully migrated participants to three distinct stages in the journey, we conducted a pretest with 135 participants. Pretest participants reviewed one of the decision stage descriptions (i.e., informational, comparison, or transactional) and rated their stage in the consumer journey (0 = far from point of purchase, 100 = close to point of purchase). The results of a one-way ANOVA were significant ($F(2, 132) = 43.25$, $p < .001$, $\eta_p^2 = .40$). Participants in the intermediate stage ($M = 56.43$, $SD = 21.31$, $N = 49$) were more proximal to purchase than those the informational stage ($M = 28.84$, $SD = 34.33$, $N = 43$; $F(1, 132) = 22.95$, $p < .001$) and more distal from purchase than those in the transactional stage ($M = 84.12$, $SD = 26.26$, $N = 43$; $F(1, 132) = 23.11$, $p < .001$).

As in our previous studies, our prediction is that consumers in the informational stage of the consumer journey will use more abstract language in their search queries whereas consumers in the transactional stage will use more concrete language (H_1). However, we do not know whether consumers in the intermediate comparison stage will bear greater resemblance to consumers in the informational stage or the transactional stage, in terms of their search queries. We are also unsure whether our concreteness score measure will allow us to distinguish consumers in the comparison stage from consumers in the other stages of the journey.

We tested this prediction by conducting a one-way (consumer journey stage: informational vs. comparison vs. transactional) ANOVA on the mean concreteness score of participants' search queries. We observed a main effect of consumer journey stage ($F(2, 285) = 3.40$, $p = .035$, $\eta_p^2 = .023$).[5] Participants in the transactional stage ($M = 393.04$, $SD = 78.45$, $N = 96$) executed searches that were more concrete (less abstract) than either participants in the informational stage ($M = 373.20$, $SD = 63.64$, $N = 101$; $F(1, 285) = 3.77$, $p = .053$) or participants in the comparison stage ($M = 367.08$, $SD = 72.60$, $N = 91$; $F(1, 285) = 6.13$, $p = .014$). However, search queries of participants in the informational stage and participants in the comparison stage did not differ in terms of concreteness ($F(1, 285) = .35$, $p = .56$). We will revisit this finding in a follow-up analysis.

Next, we conducted a one-way ANOVA on the mean concreteness score of participants' preferred search result (after combining the words that appeared in the result's title and description). We obtained a pattern similar to the search queries. Specifically, there was a main effect of consumer journey stage ($F(2, 342) = 7.12$, $p = .001$, $\eta_p^2 = .04$). The preferred result of participants in the transactional stage ($M = 385.17$, $SD = 56.02$, $N = 116$) was more concrete (less abstract) than either participants in the informational stage ($M = 360.41$, $SD = 53.33$, $N = 130$; $F(1, 342) = 12.97$, $p < .001$) or participants in the comparison stage ($M = 364.88$, $SD = 51.82$, $N = 99$; $F(1, 342) = 7.59$, $p = .006$). However, the preferred result of participants in the informational stage and participants in the comparison stage did not differ in terms of concreteness ($F(1, 342) = .39$, $p = .53$).

Finally, we examined whether participants in different stages of the decision journey exhibited different preferences for the three search result descriptions, all of which related to the Microsoft Surface laptop/tablet but varied in terms of concreteness. For choice shares, see Table 1. A chi-square test of independence showed a significant association between choice and decision stage; ($\chi^2(2, N = 357) = 39.20$, $p < .001$). Notably, 38.7% of participants in the transactional stage opted to click on the most concrete result, which is a greater proportion than participants in either the informational stage (12.9%; $\chi^2(1) = 22.52$, $p < .001$) or the comparison stage (9.9%; $\chi^2(1) = 24.15$, $p < .001$). Participants in the informational and comparison stages were equally likely to click on the concrete result ($\chi^2(1) = .50$, $p = .48$). Only 37.9% of participants in the transactional condition opted to click on the intermediate result, which is a lower proportion than participants in either the informational condition (61.4%; $\chi^2(1) = 14.08$, $p < .001$) or the comparison condition (67.3%; $\chi^2(1) = 19.29$, $p < .001$). Participants in the informational and comparison stages were equally likely to click on the intermediate result ($\chi^2(1) = .88$, $p = .25$). Although a higher proportion of participants in the

[4] The word "surface" was excluded from the concreteness score calculation because it refers to a brand name.

[5] For all dependent variables, degrees of freedom reflect text responses for which a score could be computed using the updated MRC dictionary (i.e., text with at least one nonbranded word appearing in the dictionary).

Table 1. Study 3: Result Choice Share by Decision Stage.

Concreteness of Search Result	MRC Concreteness Score	Informational Stage	Comparison Stage	Transactional Stage
Abstract ("See What's New From Surface")	307.15	25.8%	22.8%	23.4%
Intermediate ("Which Surface Works For You")	362.93	61.4%	67.3%	37.9%
Concrete ("Free Next Day Delivery")	390.93	12.9%	9.9%	38.7%

informational condition opted to click on the abstract result than those in either the comparison or transactional stages, these differences were nonsignificant ($\chi^2(1) < 1$).

Discussion

Study 3 provides additional support for H_1 and H_2, in a different decision context and with a different consumer journey manipulation. Furthermore, our results generally indicate that consumers in an intermediate comparison stage provide search queries and prefer search results that are more abstract than transactional-stage consumers. However, we were not able to empirically distinguish comparison-stage consumers from informational-stage consumers on the basis of our concreteness score measure.

We conducted an exploratory follow-up analysis on the data from Study 3 to evaluate whether refinements to our concreteness score measure would allow us to better capture differences in construal preferences between comparison-stage consumers and informational-stage consumers. Recall that all prior linguistic analyses excluded brand terms because they either do not appear in the concreteness dictionary (e.g., nonword brands) or may not be scored appropriately given the branding context (e.g., Surface). Our follow-up analysis indicates that 23.8% of all comparison queries contained brand names, which is similar to the 29.8% of transactional queries with brand names ($\chi^2(1) = 1.04$, $p = .31$) but significantly higher than the 9.8% of informational queries with brand names ($\chi^2(1) = 8.29$, $p < .01$). To the extent that the use of brand names reflects concrete mindsets, this difference might explain why our concreteness scoring—which excluded brand names—was not able to detect differences in queries between participants in the informational and comparison stages. Indeed, a separate pretest ($N = 101$) of concreteness ratings for the brands mentioned by participants in this study (e.g., Lenovo, Macintosh, HP) indicates that brand names are perceived as more concrete than other words ($M_{brands} = 4.41$ vs. $M_{other} = 2.28$; $p < .001$) on a seven-point scale (1 = "abstract," and 7 = "concrete"). We return to this point in the "General Discussion" section, where we advocate for the development of a modified concreteness score measure that accounts for product and brand names, rather than excluding them.

Study 4

Having established the construal matching effect in terms of click behavior, the purpose of Study 4 is to explore how this effect influences consumers' search satisfaction (H_2) and perceived goal progress (H_3). We also test our hypothesis that processing fluency is the underlying mechanism for the construal matching effect. We expect processing fluency to mediate the effect of construal matching on both search satisfaction and perceived goal progress (H_4).

Method

Study 4 was conducted with 427 U.S. participants (59.3% female; $M_{age} = 38.77$ years, $SD = 12.79$) recruited using an online panel (Amazon Mechanical Turk Prime). This experiment utilized a 2 (decision stage: informational vs. transactional) × 2 (search results: abstract vs. concrete) between-participants design.

Participants randomly assigned to the informational stage condition were instructed that they wanted to learn more about grilling to determine whether they might need a charcoal grill sometime in the future. In contrast, those in the transactional stage condition learned that they needed to choose a charcoal grill for a summer barbecue that they would be hosting next week.

Subsequently, all participants were told that they had decided to consult a search engine to help them with their task. Participants were shown a search engine logo and a search box underneath, and they were asked to enter their exact search query in the box. We measured the mean concreteness score of each participant's search query using the adapted MRC concreteness dictionary.

Participants then advanced to a new screen in which they were shown a set of search results that had purportedly been obtained after their search. Each set included six search result titles along with brief descriptions. Participants either encountered a set of search results that contained more abstract language (e.g., "Must-Know Grilling Tricks") or more concrete language (e.g., "Bang for Your Buck Charcoal Grills"). Each set of results was constructed so that the abstract set of search results (concreteness score = 365.25 [title only], 359.81 [with snippet]) was less concrete than the concrete set (concreteness score = 440.11 [title only], 392.81 [with snippet]). Thus, depending on the decision journey stage to which they had been assigned, participants either encountered a matching or mismatching set of results. For the exact wording and concreteness scores of each search result, see Web Appendix D.

Next, participants were asked to use a slider to indicate the amount of progress they would make on their decision journey if they visited each of the six search result websites (−50 =

"further away," 0 = "about the same," and +50 = "closer"). Participants also rated how satisfied they were by the search results (1 = "not very satisfied," and 7 = "very satisfied") and how likely they were to be helpful (1 = "not at all likely," and 7 = "very likely"). These last two items were aggregated to form a composite satisfaction index (r = .72).

On a new screen, participants were instructed to evaluate the entire search process, including their query and results (1 = "incomprehensible/unclear/difficult/disfluent/effortful," and 5 = "comprehensible/clear/easy/fluent/effortless"). These items comprised a five-item processing fluency measure (α = .93; adapted from Graf, Mayer, and Landwehr [2018]), whose inclusion in the study allowed us to test for the mediating role of fluency. Finally, as a manipulation check of matching/mismatching results, we asked participants to rate the extent to which the search results they encountered matched (1 = "matched poorly," and 7 = "matched well") and were consistent (1 = "not at all consistent," and 7 = "very consistent") with their goal. These two items were averaged to create a composite matching manipulation check measure (r = .94).

Results

Our decision stage manipulation in this study relied in part on a temporal distance cue (i.e., "in the future" vs. "next week), unlike our previous studies. We chose this manipulation because consumers naturally view the decision journey as a temporal progression toward a purchase or decision. To validate our manipulation, we conducted a pretest with 100 nonoverlapping participants from the same pool. Pretest participants read the two decision stage descriptions (i.e., informational vs. transactional) and rated the stage in the consumer journey conveyed by each description (0 = "far from point of purchase," and 100 = "close to point of purchase"). As we expected, participants rated the transactional stage (M = 58.57, SD = 30.37, N = 60) as more proximal to purchase than the informational stage (M = 32.78, SD = 32.12, N = 60; t(99) = 6.04, $p < .001$).

A 2 (decision stage) × 2 (search results) between-participants ANOVA on the matching manipulation check measure returned a significant interaction (F(1, 423) = 120.30, $p < .001$, $\eta_p^2 = .22$), which indicates that our manipulation was successful. Specifically, participants in the informational stage condition felt that the abstract search results (M = 5.62, SD = 1.36, N = 113) matched their goal more than the concrete search results (M = 4.65, SD = 1.85, N = 106; F(1, 423) = 18.63, $p < .001$, $\eta_p^2 = .04$). In contrast, participants in the transactional stage condition felt that the concrete search results (M = 5.64, SD = 1.42, N = 98) matched their goal more than the abstract search results (M = 3.07, SD = 1.92, N = 110; F(1, 423) = 123.32, $p < .001$, $\eta_p^2 = .23$).

Our prediction is that consumers in the informational stage of the consumer decision journey will use more abstract language in their search queries, whereas consumers in the transactional stage will use more concrete language. We tested this prediction by conducting a 2 (decision stage) × 2 (search results) between-participants ANOVA on the mean concreteness score of their search queries. As expected, we observed only a main effect of decision stage (F(1, 422) = 93.86, $p < .001$, $\eta_p^2 = .18$), such that participants in the informational stage (M = 445.71, SD = 73.02, N = 219) executed searches that were less concrete (more abstract) than participants in the transactional stage (M = 509.05, SD = 60.64, N = 207). This result again supports H_1.

Next, we conducted the 2 (decision stage) × 2 (search results) between-participants ANOVA on the goal progress measure, which served as our key dependent variable. We observed a main effect of decision stage (F(1, 423) = 18.91, $p < .001$, $\eta_p^2 = .04$) and a main effect of search results (F(1, 423) = 33.29, $p < .001$, $\eta_p^2 = .07$). More germane to our theorizing, we also observed a significant interaction (F(1, 423) = 73.05, $p < .001$, $\eta_p^2 = .15$). Participants in the informational stage (M = +25.23, SD = 17.40, N = 113) felt that they had made greater goal progress when they encountered abstract search results as opposed to concrete search results (M = +19.82, SD = 20.10, N = 106; F(1, 423) = 3.96, $p = .047$, $\eta_p^2 = .01$). However, participants in the transactional stage condition felt that they had made greater goal progress when they encountered concrete search results (M = +28.00 SD = 15.59, N = 98) as opposed to abstract search results (M = .11, SD = 25.52, N = 110; F(1, 423) = 99.80, $p < .001$, $\eta_p^2 = .19$).

A similar pattern of results was obtained when the satisfaction index was the dependent variable. We observed a main effect of decision stage (F(1, 423) = 17.25, $p < .001$, η_p^2 = .04) and a main effect of search results (F(1, 423) = 9.07 $p = .003$, $\eta_p^2 = .02$). More germane to our theorizing, we also observed a significant interaction (F(1, 423) = 85.85, $p < .001$, $\eta_p^2 = .17$). Participants in the informational stage (M = 5.62, SD = 1.25, N = 113) were more satisfied with the abstract search results versus the concrete search results (M = 4.69, SD = 1.71, N = 106; F(1, 423) = 20.10, $p < .001$, η_p^2 = .05). However, participants in the transactional stage condition were more satisfied when they encountered concrete search results (M = 5.45, SD = 1.38, N = 98) as opposed to abstract search results (M = 3.63, SD = 1.74, N = 110; F(1, 423) = 73.40, $p < .001$, $\eta_p^2 = .15$). Our findings with respect to perceived goal progress and search satisfaction lend support to H_3.

Finally, we conducted a mediation analysis using the PROCESS macro developed by Hayes (2017) to examine whether fluency of the search process mediated the relationship between this matching effect and participants' perceived goal progress. This mediation analysis utilized bootstrapping with repeated extraction of 10,000 samples (Hayes 2017, model 4). We created a match variable, which was coded as 1 if participants were in an informational (transactional) stage and had received abstract (concrete) search results, and coded as 0 otherwise. Consistent with our theorizing, we found a significant indirect effect of matching/mismatching on the goal progress measure through fluency (B = 6.42, SE = 1.25, 95% CI = [4.11, 8.99]). A second mediation analysis, with the satisfaction index as the dependent variable, returned a similar result such that a significant indirect effect of matching/mismatching through fluency (B = .57, SE = .10, 95% CI = [.38, .78]) was observed. These analyses support H_4.

Discussion

Study 4 provides support for all four of our hypotheses. In addition to again showing that decision stage affects the abstractness/concreteness of consumers' search queries (H_1), we find that the construal matching effect influences search satisfaction (H_2) and perceived goal progress (H_3). Furthermore, we show that construal matching leads to greater processing fluency which in turn yields positive effects on perceived goal progress and search satisfaction (H_4). In Study 5, we revert our attention to the effect of construal matching on click-throughs (i.e., H_2 only) and design a field experiment that aims to document real-world consequences of construal matching in online search.

Study 5

The purpose of Study 5 is to validate the construal matching effect in online search by examining actual click behavior in a real-world context. Specifically, we conducted a field experiment via a search engine to test whether consumers who have an informational goal (i.e., when point of purchase is distal) are more likely to click on an abstract advertisement and if consumers who have a transactional goal (i.e., when point of purchase is proximal) are more likely to click on a concrete advertisement. Because this study focuses on actual click behavior, it allows for a strong test of the external validity of H_2.

Method

In this study, we chose to identify consumers at different stages of the purchase funnel based on their online search queries. Using the website of the skin care and cosmetics company Philosophy (philosophy.com) as our landing page, we bid on three modified broad match keywords that appeared in search queries: +buy, +best, and +how in combination with one of ten cosmetics-related terms.[6] We selected the keywords of "buy" (concreteness score = 389.96) versus "best" (concreteness score = 260.28) because their relative MRC dictionary concreteness scores (Paetzold and Specia 2016; Packard and Berger 2020) suggest that they are strong indicators of transactional and informational goals, respectively.

We selected the "how" keyword for a more theoretical reason. Prior construal research has shown that considering questions of "how" tends to prime concrete construal, whereas considering questions of "why" leads to abstract construal (e.g., Freitas, Gollwitzer, and Trope 2004; Fujita et al. 2006). Given these findings, the keyword of "how" might seem to imply a more transactional search goal. Yet the MRC dictionary revealed that the word "how" is itself relatively abstract (concreteness score = 272.64), which suggests that its use may be indicative of an informational search goal. Given these opposing perspectives, it is unclear whether the use of the word "how" in a search query denotes that point of purchase is proximal or distal. Therefore, we did not have a hypothesis about +how a priori and included it primarily for exploratory purposes.

We considered also including the keyword of "why" because it is considered abstract and has been contrasted with "how" in prior construal research (e.g., Freitas, Gollwitzer, and Trope 2004; Fujita et al. 2006). Furthermore, it is also classified as abstract in the MRC dictionary (concreteness score = 286.84). However, a brief test campaign revealed that +why searches in combination with one of the chosen cosmetics terms are relatively uncommon, making "why" a poor keyword to use in this field experiment. The same test campaign confirmed that the frequencies of +buy, +best, and +how cosmetics-related searches were relatively similar, making these keywords suitable candidates for our experiment.

Each of three keywords (i.e., "buy," "best," and "how") was set up to be its own search campaign. We also created two versions of an online advertisement for Philosophy that were written using either relatively abstract language (concreteness score = 301.67) or relatively concrete language (concreteness score = 415.61) (i.e., "Be Your Beautiful Best" [abstract] vs. "Makeup, Moisturizer, Skin Care" [concrete]). For study stimuli, see Web Appendix D. Although the ads differed in terms of concreteness, a pretest confirmed that they were similar on most other dimensions, including likeability.[7] Conditional on a successful search bid, a consumer who performed a search with one of the three keywords plus one of the ten cosmetics-related terms was randomly shown one of the two versions of the Philosophy ad. The search campaigns on Google AdWords ran

[6] A modified broad match is specified by an advertiser during the bidding process and triggers an advertisement if certain keywords are included in a search, irrespective of keyword order. In this experiment, we ran three campaigns for +buy, +best, and +how, which were mandatory keywords for each campaign that needed to be paired with any of the following terms: lip, lotion, hair color, skin care, lipstick, eyeliner, eyeshadow, anti-aging, mascara, and moisturizer.

[7] We conducted a pretest with 96 female participants (M_{age} = 36.90 years, SD = 12.59) who saw one of two versions of the online advertisement and rated it on several dimensions. Controlling for participants' age and familiarity with the Philosophy brand and the beauty product category, the pretest confirmed that the two ads were not perceived to differ in terms of likeability (B = .20, SE = .13, p = .13), goal orientation (B = .08, SE = .14, p = .55), and solution orientation (B = .19, SE = .15, p = .22). Although the ads differed in terms of relatability (B = .34, SE = .15, p = .03), this difference would not predict the matching effect we observed.

Table 2. Study 5: Summary Counts.

Ad Copy	Search	Impressions	Clicks
Abstract	Buy	14,910	313
	Best	25,861	649
	How	16,706	631
Concrete	Buy	36,725	733
	Best	20,351	363
	How	13,590	413
Total		128,143	3,102

from March to May 2018. For each of the campaigns, we collected data on the number of click-throughs and the number of impressions generated for each of the two versions of the advertisement.

Results

Our full search engine data set included 128,143 impressions and 3,102 click-throughs. Table 2 shows the summary counts of impressions and click-throughs for each combination of keywords and advertisements. In line with our theorizing, consumers should respond to advertising whose language matches their level of construal.

To test for the construal matching effect, we first restricted our data set to only +buy and +best keywords, for which we had clear predictions. This restricted data set included 97,847 impressions and 2,058 click-throughs. We predicted an interaction effect (H_2), such that consumers with a transactional goal (i.e., those who entered a +buy cosmetics-related search query) would be more likely to click on an advertisement containing concrete language. Conversely, we expected that consumers with an informational goal (i.e., those who entered a +best cosmetics-related search query) would be more likely to click on an advertisement containing abstract language. We conducted a binary logistic regression analysis to test for this interaction while simultaneously controlling for the advertisement's rank and its cost per click.[8] The model estimates and fit statistics for this analysis appear as Model 1 in Table 3.

The dependent variable of interest is whether or not an advertisement was clicked (1 = clicked, 0 = not clicked). There was a main effect of advertisement, such that the abstract ad received more clicks. There was also a main effect of keyword, such that "best" searches led to a higher rate of click-throughs. More germane to our theorizing, we also observed that when there was a match between search goal (as indicated by the use of "best" or "buy" in the query) and advertisement type (i.e., abstract vs. concrete language), the likelihood of a click-through was higher ($\hat{\beta} = .65$, $p < .01$). This finding supports H_2.

[8] Ad rank refers to the advertisement's vertical position on the search engine result page. Cost per click is dynamically determined by the search engine and varies based on time of day and other exogenous factors.

Table 3. Study 5: Logistic Regression Estimates, Impact of Matching Advertising Content with Search Goal on Impression Click-Throughs.

	Dependent Variable: Search Result Click-Through	
	Model 1 (1)	Model 2 (2)
Avg. position	1.77***	1.81***
	(0.39)	(0.37)
Avg. cost per click	1.22***	1.24***
	(0.24)	(0.23)
Is abstract ad	0.60***	–.06
	(0.10)	(0.07)
Is best search	0.41***	–.25
	(0.13)	(0.16)
Is how search		2.06***
		(0.31)
Ad-search matched	0.65**	
	(0.12)	
Abstract ad × Best search		1.33***
		(0.23)
Abstract ad × How search		0.33***
		(0.10)
Observations	97,847	128,143
Nagelkerke R^2	0.003	0.009
Likelihood ratio test	$\chi^2 = 58.63$, d.f. = 5 $p < .0001$	$\chi^2 = 238.57$, d.f. = 7 $p < .0001$
Log-likelihood	−9,954.20	−14,487.71
Akaike information criterion	19,920.41	28,991.42

*$p < .05$.
**$p < .01$.
***$p < .001$.
Notes: Standard errors in parentheses. Model 1 comprises data pertaining to best and buy queries only, comparing abstract and concrete construal levels in consumers' search behavior. Model 2 comprises all data records.

In addition to +best and +buy keywords, we included +how keywords in Model 2 of Table 3. In Model 2, we included two interaction effects, representing the pairing of an abstract ad and a "best" search, and an abstract ad and a "how" search ("buy" was used as the baseline in the logistic regression for both models). We observed a positive interaction effect between a "best" search and showing an abstract ad ($\hat{\beta} = 1.33$, $p < .01$), which suggests that consumers prefer an abstract ad when they have an informational goal in mind. Although we had no a priori expectations regarding +how, finding a positive interaction effect between a "how" search and showing an abstract ad would indicate that +how coincides with a more abstract construal and an informational goal. However, finding a negative interaction would indicate that +how coincides with a more concrete construal and a transactional goal. We found evidence of the former; the interaction between a "how" search and showing an abstract ad was positive ($\hat{\beta} = .33$, $p < .001$). Thus, in this particular context, +how appears to coincide with a more abstract construal and an informational goal.

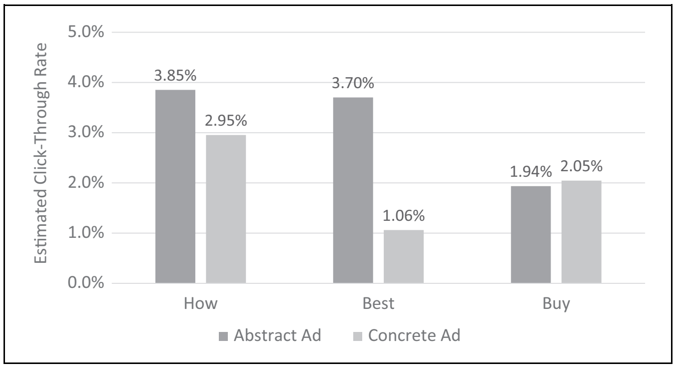

Figure 2. Study 5: Estimated average click-through probabilities by ad type and search type.

Notes: Based on predicted values from the logistic model. Numerical control variables (i.e., average position and CPC) are assumed to be their sample means by search type.

The results of Model 2 (in Table 3) align with the data shown in Table 2, which indicates that consumers clicked on the abstract (vs. concrete) advertisement more frequently when their beauty-related search query included the word "how." This may be because in the search context of beauty and cosmetics products, learning how to perform a task, (e.g., apply eyeshadow) is a more of a top-of-the-funnel activity as compared with deciding which product to purchase.

Figure 2 illustrates our logistic regressions, showing the estimated means of click-through probabilities by advertisement type and keyword. When the search contained the word "best" or "how," displaying the abstract ad yielded higher click-throughs than the concrete ad. When the search contained the word "buy," the two ads have similar click-through rates. This figure visually demonstrates that in spite of the main effect of abstract advertisement, there is a positive effect of matching abstract (concrete) search queries with abstract (concrete) advertisements, which corroborates the matching effect. The findings from all of these analyses strongly support H_2.

Drawing on the findings from Study 5, we evaluate the financial benefits of matching advertising content with search goals for the three keywords (i.e., "best," "how," and "buy"). We perform this exercise from both a search engine's perspective and a brand's perspective. For a search engine, the financial benefits come from consumers' click-throughs, as revenue is equal to click-through rate multiplied by cost per click. For a brand, we predict how quickly it can fulfill its advertising campaign goals, namely, how fast it can achieve a target number of click-throughs.

We begin with a search engine's perspective. We refer to the estimates of Model 2 and first calculate the average click-through rate of each combination of ad copy and search type, as shown in Table 4. On average, cost per click for "buy" queries is \$2.82, "best" is \$1.83, and "how" is \$1.79. Now we have both click-through rate and cost per click, the expected revenues from an impression for a search engine company is click-through rate × cost per click, as summarized

Table 4. Study 5: Predicted Click-Through Rates by Advertising Content and Search Goal.

Ad Copy	Search	Predicted Click-Through Rate	Expected Revenues from an Impression
Abstract	Buy	1.94%	5.46¢
	Best	3.70%	6.78¢
	How	3.85%	6.90¢
Concrete	Buy	2.05%	5.77¢
	Best	1.06%	1.94¢
	How	2.95%	5.29¢
Ad type	Buy	Best	How
Abstract	1.94%/5.46¢	**3.70%/6.78¢**	**3.85%/6.90¢**
Concrete	**2.05%/5.77¢**	1.06%/1.94¢	2.95%/5.29¢

Notes: The two reported statistics are click-through rate and expected revenue per impression. Specifically, predicted click-through rate is calculated based on parameter estimates of Model 2. Expected revenues from an impression is equal to predicted click-through rate multiplied by cost per click, which are \$2.82 for +buy, \$1.83 for +best, and \$1.79 for +how. Boldfaced cells are results from serving a matched ad.

in the last column in Table 4. In Table 5, we calculate the increase in expected revenues from serving a matched ad versus three baseline scenarios (i.e., always serving a concrete ad, always serving an abstract ad, and serving a concrete or abstract ad half of the time at random). Based on the ad campaigns in our field experiment, expected increase in search engine revenue due to increased click-through rates would have been \$2,714.35 if queries were served with matching advertisement content as opposed to always serving a concrete ad. Compared with the other two baseline scenarios, the increase in search engine revenue is \$164.52 (compared with always serving an abstract ad) and \$1,439.43 (compared with matching abstract or concrete ad half of the time), respectively.

From a brand's perspective, one potential benefit from matching advertisements is time saving (i.e., reduction in the needed duration of the digital advertising campaign). We evaluate how long it takes to reach the goal of a digital advertising campaign, which is often specified by the number of click-throughs achieved. Drawing on our field experiment, which took 56 days to achieve 3,102 clicks, we calculated the average impressions and clicks per day, as summarized in Table 6. Assuming that a campaign goal is to achieve 500 clicks, we then calculated the expected number of days it takes to reach the goal based on our recommendation (i.e., matched advertisement) and the three baseline scenarios as before (i.e., always serving an abstract ad, always serving a concrete ad, and serving 50–50 of concrete and abstract ad). Summarized in Table 7, it takes 20.13 days if the campaign always serves a concrete ad, while serving a matched ad takes 11.80 days, which is a time saving of 8.33 days. Compared with the other two baseline scenarios, time saving is 2.97 days (compared with always serving an abstract ad) and 5.24 days (compared with matching abstract or concrete ad half of the time), respectively.

Table 5. Study 5: Comparing Search Engine Revenues by Advertisement Strategy.

Scenario	Expected Revenue Increase per 100 Impressions			Total Increase
	Buy Search	How Search	Best Search	
# impressions	51,635	30,296	46,212	
Serving a matched ad versus always serving a concrete ad	$0	$1.60	$4.82	$2,714.35
Serving a matched ad versus always serving an abstract ad	$.32	$0	$0	$164.52
Serving a matched ad versus serving a concrete or abstract ad 50–50	$.16	$.80	$2.41	$1,439.43

Notes: Expected revenue increase per 100 impressions is the revenue difference between matching and the baseline scenario per impression, multiplied by 100.

Table 6. Study 5: Comparing Campaign Time Saving by Advertisement Strategy.

Ad Copy	Search	Impressions per Day	Expected Clicks per Day
Abstract	Buy	266	5.32
	Best	462	17.07
	How	298	11.46
Concrete	Buy	656	13.83
	Best	363	3.85
	How	243	7.15

Notes: Expected clicks per day is equal to impressions per day multiplied by click-through rate.

Table 7. Study 5: Comparing Campaign Time Saving by Advertisement Strategy.

Scenario	Expected Clicks per Day	Days to Reach 500 Clicks	Time Saving in Days (Compared with Serving a Matched Ad)
Serving a matched ad	42.36	11.80	—
Always serving a concrete ad	24.83	20.13	8.33
Always serving an abstract ad	33.84	14.78	2.97
Serving a concrete or abstract ad 50–50	29.34	17.04	5.24

Notes: Expected clicks per day is equal to the sum of clicks by query type and the corresponding ad type.

Discussion

In Study 5, we provide evidence of the construal matching effect on clicks (H_2) in a real, field test setting. Conceptually replicating Study 2, we find that participants who enter a search term suggesting they are in an abstract mindset (e.g., "best") are more likely to click search advertising that matches this mindset. Conversely, consumers who search for a term suggesting they are in a concrete mindset ("buy"), are more likely to engage with an ad that matches this mindset. In Study 5, we show real potential gains in click engagement based on matching construal level of participants. Furthermore, it is worth noting that words like "best" are less expensive than "buy" from a cost-per-click perspective,[9] which suggests that marketers may be able to more efficiently allocate their search marketing budget (Tables 5–7).

General Discussion

We view online shopping and search as an experience (Lemon and Verhoef 2016)—one in which consumers do not simply search and buy, but rather learn, play, and explore. This experience is not necessarily transaction-oriented; it can be involved and extended as consumers learn about products, read reviews, and explore options. Yet in digital marketing, current marketing practice places a great deal of firm attention and resources on the last part of the experience, the purchase. On the one hand, this makes sense; purchase is where consumers actually give money to the company. However, in a competitive environment, vying for this space means that firms pay more for something that may, we argue, be worth less. Our work suggests that this may not be the most efficient use of marketing resources, and that an approach personalized to consumers may be more optimal. Consumers are different; they are at different phases of the shopping experience and may have different mindsets.

Analyzing a combination of archival, laboratory, and field data, we demonstrate a method for quantifying construal level of search strings and ad content. We find that consumers express a range of construal levels in search queries that predict where consumers are in the decision journey. Results from both our laboratory experiments and a field study illustrate that matching consumer mindset can increase engagement with search advertising. We also show that this difference in construal has downstream consequences for the type of information that consumers prefer. We find that consumers may prefer product category or brand/retailer information, depending on their goal and construal level. Specifically, consumers prefer to visit brand websites and commerce/retailer websites when they are farther in the journey, and they are more likely to visit informational sites when they are earlier in the decision journey. Matching construal has favorable outcomes for both consumers and firms. Consumers are more satisfied with the search experience and perceive more progress toward their goal. Firms

[9] In our field experiment, the average cost per click for a +buy keyword was $2.82, compared with $1.83 for +best and $1.79 for +how.

can benefit from higher click-through rates on less expensive advertising, leading to efficiencies.

Theoretically, this article makes several contributions to how we think about the mindsets of consumers as they navigate their journey. Previous work on construal has examined the ways in which various inputs and contextual factors may affect psychological distance and therefore construal level. However, this research has not linked construal directly to the consumer decision journey or explored some of its important consequences—linguistic output, desire for particular types of content, and progression in the consumer journey. We show that consumers with different goals approach search with distinct mindsets that are reflected in search queries. Using text analysis, we can measure and classify consumer mindset. In addition to detecting consumer mindset, we show that matching it can not only produce a more satisfying search experience, but that it can also change consumers' perceptions about their progress toward their goals. This meaningfully builds on the idea that the consumer journey is varied and various, and firms should better cater to these differences (Hamilton and Price 2019). We find, for instance, contextual variation for different product categories. The journeys for travel, services, healthcare, or experiences may be relatively abstract, as consumers take longer to explore and understand while journeys for CPGs may be relatively transactional and concrete, given that the products themselves are physical, often utilitarian objects.

Journeys can also be varied due to contextual factors and individual differences. Contextual factors such as risk, search strategy (i.e., maximizing vs. satisficing), involvement, and expertise might influence how construal varies through the decision journey. Individual differences such as resource scarcity (Hamilton et al. 2019) or attachment style (Mende et al. 2019) may have a profound impact on how consumers progress through their journey. Being able to detect and adapt to these differences is therefore a valuable insight that can serve both consumers and firms.

Despite these different starting baselines for concreteness across categories, within the journey, we find that consumers move from relatively more abstract to relatively more concrete as they move closer to purchase. Future research should explore the nature of other contextual factors that exist not only for different product categories but perhaps also for different brands or different consumer segments. Our core insight—that consumer mindset can be detected through linguistic analysis and then matched to produce a more fluent shopping experience—can be used to inform many of these contextual differences.

And yet, as illustrated by one of our studies, the journey may not always be clearly defined. For instance, in Study 3, we find that participants' construal between an informational stage and a comparison stage in online search may not always be distinguishable through text analysis. Future research might probe deeper into other textual indicators of stage in the decision journey such as brand mentions or other linguistic cues. The concreteness score metric used in this research may serve as a useful starting point for more sophisticated attempts at linking construal level to the consumer journey. As previously mentioned, we developed an online tool for this purpose (www.construalscore.com) that quantifies the concreteness of written communication and may facilitate the work of other researchers on this important topic.

Our results also speak to the extant literature on search advertising. Prior research regarding consumer search has examined the text of search to better optimize ad buying and increase consumer response based on keyword matching. For instance, some work has shown that the use of less (vs. more) popular keywords is an indicator that consumers are closer to purchase and therefore search advertising from these keywords is more valuable (e.g., Jerath, Ma, and Park 2014). In this research, we build on this work by showing that although less popular keywords are indeed more valuable search terms that focus on product attributes (e.g., more concrete), keywords that focus on product benefits (e.g., more abstract) may be valuable as well, particularly when they match consumer mindset. A match between construal level of the keyword and the sponsored ad is expected to produce more clicks per search, thereby making marketing dollars more efficient. An objective of this work was to go beyond research on search order effects and firm competition for keywords to better understand how firms can actually connect to consumer goals in online search.

With respect to managerial recommendations, we propose that firms view online search as part of the consumer journey, not merely the last stage in it. Consumers use online search as an embedded tool through which they learn about the world. As we have shown, search advertising platforms can monitor consumers' search behavior to predict where they are in the journey and respond adaptively rather than uniformly. They can then produce the search results, ad copy, and extended online content and storytelling to provide the appropriate information that meets consumer needs. Future work might explore the benefits of taking this customization further, by offering consumer experiences in landing pages and other communication for consumers in different mindsets. Customers expect different experiences throughout the consumer journey (Lemon and Verhoef 2016). In addition to having disparate preferences for how search results should be presented and communicated to them, consumers may also demand larger or smaller assortments of choice options depending on their construal level (Xu, Jiang, and Dhar 2016) and stage of the decision journey. Next generations of search advertising platforms could incorporate both textual and voice search to provide personalized, relevant results, combining not just semantic information but also sentence structure and even voice tonality, with machine learning models, to provide a unique consumer experience.

Findings from this research can help both consumers and marketers improve the online shopping experience. Our aim was to enrich our understanding of how consumers' mindsets change throughout the journey and to provide specific and actionable guidance to managers on how to most effectively communicate with customers, particularly when they are engaged in online search. More broadly, using traces of consumer activity online, particularly cues offered in textual communication, presents many fruitful avenues whereby consumer needs can be identified and met in unobtrusive, personalized, and useful ways. For instance, future research might examine whether device type (i.e., laptop vs. mobile) interacts with textual communication to provide an even more predictive cue regarding construal level or another marker of consumers' stage in their decision journey.

In summary, consumers have different and varied goals when moving through their journey. As Hamilton and Price (2019, p. 190) argue, recent research on the consumer journey has "uncovered challenges in neatly defining a pre-core-post customer journey process, suggesting the value of more adaptive and customized time frames depending on the consumer journey context." Consumers use online search, often counter to current marketing practice, as a tool for learning, exploring, gaining social information, and a varied number of other goals. As we have demonstrated, one important dimension of these goals is construal level. Detecting and matching consumer mindset enables firms to meet consumers where they are in the journey, producing a better overall experience.

Acknowledgments

The authors thank Angela Y. Lee, Grant Packard, and participants of the 2020 Wharton Behavioral Insights from Text Conference for their helpful comments and suggestions. They also gratefully acknowledge the support of Microsoft and the Intent Lab, a collaborative research unit created by Performics and Northwestern University, as well as the following individuals: Burcu Agma, Alice Air, Minna Du, Kelli Kemery, Shawn Mohamed, Ines Paez, Joe Reinstein, and Esteban Ribero.

Associate Editor

Stephen Nowlis

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Alter, Adam L. and Daniel M. Oppenheimer (2009), "Uniting the Tribes of Fluency to Form a Metacognitive Nation," *Personality and Social Psychology Review*, 13 (3), 219–35.

Batra, Rajeev and Kevin Lane Keller (2016), "Integrating Marketing Communications: New Findings, New Lessons, and New Ideas," *Journal of Marketing*, 80 (6), 122–45.

Berger, Jonah, Ashlee Humphreys, Stephan Ludwig, Wendy W. Moe, Oded Netzer, and David A. Schweidel (2020), "Uniting the Tribes: Using Text for Marketing Insight," *Journal of Marketing*, 84 (1), 1–25.

Brysbaert, Marc, Amy Beth Warriner, and Victor Kuperman (2014), "Concreteness Ratings for 40 Thousand Generally Known English Word Lemmas," *Behavior Research Methods*, 46 (3), 904–11.

Camacho, Christopher J., E.Tory Higgins, and Lindsay Luger (2003), "Moral Value Transfer from Regulatory Fit: What Feels Right Is Right and What Feels Wrong Is Wrong," *Journal of Personality and Social Psychology*, 84 (3), 498–510.

Coltheart, Max (1981), "The MRC Psycholinguistic Database," *Quarterly Journal of Experimental Psychology Section A*, 33 (4), 497–505.

Enberg, Jasmine (2019), "Global Digital Ad Spending 2019," *eMarketer* (March 28), https://www.emarketer.com/content/global-digital-ad-spending-2019.

Freitas, Antonio L., Peter Gollwitzer, and Yaacov Trope (2004), "The Influence of Abstract and Concrete Mindsets on Anticipating and Guiding Others' Self-Regulatory Efforts," *Journal of Experimental Social Psychology*, 40 (6), 739–52.

Fujita, Kentaro, Yaacov Trope, Nira Liberman, and Maya Levin-Sagi (2006), "Construal Levels and Self-Control," *Journal of Personality and Social Psychology*, 90 (3), 351–67.

Graf, Laura K.M., Stefan Mayer, and Jan R. Landwehr (2018), "Measuring Processing Fluency: One Versus Five Items," *Journal of Consumer Psychology*, 28 (3), 393–411.

Hamilton, Rebecca W. and Linda L. Price (2019), "Consumer Journeys: Developing Consumer- Based Strategy," *Journal of the Academy of Marketing Science*, 47 (2), 187–91.

Hamilton, Rebecca, Debora Thompson, Sterling Bone, Lan Nguyen Chaplin, Vladas Griskevicius, Kelly Goldsmith, et al. (2019), "The Effects of Scarcity on Consumer Decision Journeys," *Journal of the Academy of Marketing Science*, 47 (3), 532–50.

Hamilton, Ryan and Alexander Chernev (2010), "The Impact of Product Line Extensions and Consumer Goals on the Formation of Price Image," *Journal of Marketing Research*, 47 (1), 51–62.

Hansen, Jochim and Michaela Wänke (2010), "Truth from Language and Truth from Fit: The Impact of Linguistic Concreteness and Level of Construal on Subjective Truth," *Personality and Social Psychology Bulletin*, 36 (11), 1576–88.

Hansen, Jochim and Michaela Wänke (2011), "The Abstractness of Luxury," *Journal of Economic Psychology*, 32 (5), 789–96.

Hayes, Andrew F. (2017), *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York: Guilford Press.

Higgins, E. Tory (2000), "Making a Good Decision: Value from Fit," *American Psychologist*, 55 (11), 1217–30.

Higgins, E. Tory, Lorraine Chen Idson, Antonio L. Freitas, Scott Spiegel, and Daniel C. Molden (2003), "Transfer of Value from Fit," *Journal of Personality and Social Psychology*, 84 (6), 1140–53.

Higgins, E. Tory, Arie W. Kruglanski, and Antonio Pierro (2003), "Regulatory Mode: Locomotion and Assessment as Distinct Orientations," in *Advances in Experimental Social Psychology*, Vol. 35, M.P. Zanna, ed. Cambridge: Elsevier Academic Press, 293–344.

Hoban, Paul R. and Randolph E. Bucklin (2015), "Effects of Internet Display Advertising in the Purchase Funnel: Model-Based Insights from a Randomized Field Experiment," *Journal of Marketing Research*, 52 (3), 375–93.

Huang, Ni, Gordon Burtch, Yili Hong, and Evan Polman (2016), "Effects of Multiple Psychological Distances on Construal and Consumer Evaluation: A Field Study of Online Reviews," *Journal of Consumer Psychology*, 26 (4), 474–82.

Humphreys, Ashlee and Rebecca Jen-Hui Wang (2017), "Automated Text Analysis for Consumer Research," *Journal of Consumer Research*, 44 (6), 1274–306.

Jerath, Kinshuk, Liye Ma, and Young-Hoon Park (2014), "Consumer Click Behavior at a Search Engine: The Role of Keyword Popularity," *Journal of Marketing Research*, 51 (4), 480–86.

Joshi, Priyanka D. and Cheryl J. Wakslak (2014), "Communicating with the Crowd: Speakers Use Abstract Messages When Addressing Larger Audiences." *Journal of Experimental Psychology: General*, 143 (1), 351–62.

Khan, Uzma, Meng Zhu, and Ajay Kalra (2011), "When Trade-Offs Matter: The Effect of Choice Construal on Context Effects," *Journal of Marketing Research*, 48 (1), 62–71.

Kim, Hakkyun, Akshay R. Rao, and Angela Y. Lee (2009), "It's Time to Vote: The Effect of Matching Message Orientation and Temporal Frame on Political Persuasion," *Journal of Consumer Research*, 35 (6), 877–89.

Kruglanski, Arie W., Antonio Pierro, and Erik P. Thompson (2000), "To 'Do the Right Thing' or To 'Just Do It': Locomotion and Assessment as Distinct Self-Regulatory Imperatives," *Journal of Personality-and Social Psychology*, 79 (5), 793–815.

Lee, Angela Y. and Jennifer L. Aaker (2004), "Bringing the Frame into Focus: The Influence of Regulatory Fit on Processing Fluency and Persuasion," *Journal of Personality and Social Psychology*, 86 (2), 205–18.

Lee, Angela Y., Punam Anand Keller, and Brian Sternthal (2010), "Value from Regulatory Construal Fit: The Persuasive Impact of Fit Between Consumer Goals and Message Concreteness," *Journal of Consumer Research*, 36 (5), 735–47.

Lee, Angela Y. and Aparna A. Labroo (2004), "The Effect of Conceptual and Perceptual Fluency on Brand Evaluation," *Journal of Marketing Research*, 41 (2), 151–65.

Lemon, Katherine N. and Peter C. Verhoef (2016), "Understanding Customer Experience Throughout the Customer Journey," *Journal of Marketing*, 80 (6), 69–96.

Liberman, Nira and Yaacov Trope (1998), "The Role of Feasibility and Desirability Considerations in Near and Distant Future Decisions: A Test of Temporal Construal Theory," *Journal of Personality and Social Psychology*, 75 (1), 5–18.

Liberman, Nira, Yaacov Trope, and Elena Stephan (2007), "Psychological Distance," in *Social Psychology: Handbook of Basic Principles*, 2nd ed., A.W. Kruglanski and E.T. Higgins, eds. Vol. 2. New York: Guilford Press, 353–433.

Malaviya, Prashan and Brian Sternthal (2009), "Parity Product Features Can Enhance or Dilute Brand Evaluation: The Influence of Goal Orientation and Presentation Format," *Journal of Consumer Research*, 36 (1), 112–21.

Mende, Martin, Maura L. Scott, Aaron M. Garvey, and Lisa E. Bolton (2019), "The Marketing of Love: How Attachment Styles Affect Romantic Consumption Journeys," *Journal of the Academy of Marketing Science*, 47 (2), 255–73.

Nakata, Cheryl, Elif Izberk-Bilgin, Lisa Sharp, Jelena Spanjol, Anna Shaojie Cui, Stephanie Y. Crawford, et al. (2019), "Chronic Illness Medication Compliance: A Liminal and Contextual Consumer Journey," *Journal of the Academy of Marketing Science*, 47 (2), 192–215.

Packard, Grant and Jonah Berger (2020), "How Concrete Language Shapes Customer Satisfaction," *Journal of Consumer Research* (published online July 18, 2020), DOI: 10.1093/jcr/ucaa038.

Paetzold, Gustavo and Lucia Specia (2016), "Inferring Psycholinguistic Properties of Words," in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Stroudsburg, PA: Association for Computational Linguistics, 435–40.

Palmeira, Mauricio (2015), "Abstract Language Signals Power, but Also Lack of Action Orientation," *Journal of Experimental Social Psychology*, 61, 59–63.

Pham, Michel Tuan, Iris W. Hung, and Gerald J. Gorn (2011), "Relaxation Increases Monetary Valuations," *Journal of Marketing Research*, 48 (5), 814–26.

Reber, Rolf, Norbert Schwarz, and Piotr Winkielman (2004), "Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience?" *Personality and Social Psychology Review*, 8 (4), 364–82.

Reber, Rolf, Pascal Wurtz, and Thomas D. Zimmermann (2004), "Exploring "Fringe" Consciousness: The Subjective Experience of Perceptual Fluency and its Objective Bases," *Consciousness and Cognition*, 13 (1), 47–60.

Snefjella, Bryor and Victor Kuperman (2015), "Concreteness and Psychological Distance in Natural Language Use," *Psychological Science*, 26 (9), 1449–60.

Spassova, Gerri and Angela Y. Lee (2013), "Looking into the Future: A Match Between Self-View and Temporal Distance," *Journal of Consumer Research*, 40 (1), 159–71.

Stephan, Elena, Nira Liberman, and Yaacov Trope (2010), "Politeness and Psychological Distance: A Construal Level Perspective," *Journal of Personality and Social Psychology*, 98 (2), 268–80.

Trope, Yaacov and Nira Liberman (2000), "Temporal Construal and Time-Dependent Changes in Preference," *Journal of Personality and Social Psychology*, 79 (6), 876–89.

Trope, Yaacov and Nira Liberman (2003), "Temporal Construal," *Psychological Review*, 110 (3), 403–21.

Trope, Yaacov and Nira Liberman (2010), "Construal-Level Theory of Psychological Distance," *Psychological Review*, 117 (2), 440–63.

Trope, Yaacov, Nira Liberman, and Cheryl Wakslak (2007), "Construal Levels and Psychological Distance: Effects on Representation, Prediction, Evaluation, and Behavior," *Journal of Consumer Psychology*, 17 (2), 83–95.

Ülkümen, Gülden and Amar Cheema (2011), "Framing Goals to Influence Personal Savings: The Role of Specificity and Construal Level," *Journal of Marketing Research*, 48 (6), 958–69.

Vallacher, Robin R. and Daniel M. Wegner (1989), "Levels of Personal Agency: Individual Variation in Action Identification," *Journal of Personality and Social Psychology*, 57 (4), 660–71.

White, Katherine, Rhiannon MacDonnell, and Darren W. Dahl (2011), "It's the Mind-Set That Matters: The Role of Construal Level and Message Framing in Influencing Consumer Efficacy and Conservation Behaviors," *Journal of Marketing Research*, 48 (3), 473–85.

Wilson, M. (1988), "MRC Psycholinguistic Database: Machine-Usable Dictionary, Version 2.00," *Behavior Research Methods, Instruments, & Computers*, 20 (1), 6–10.

Xu, Jing, Zixi Jiang, and Ravi Dhar (2013), "Mental Representation and Perceived Similarity: How Abstract Mindset Aids Choice from Large Assortments," *Journal of Marketing Research*, 50 (4), 548–59.

Industrial Marketing Management 94 (2021) A2–A10

Contents lists available at ScienceDirect

Industrial Marketing Management

journal homepage: www.elsevier.com/locate/indmarman

Editorial

How to develop great conceptual frameworks for business-to-business marketing

ARTICLE INFO

Keywords
Conceptual articles
Theory development
Conceptual framework
Theorizing

ABSTRACT

Robust conceptual frameworks are essential to building academic knowledge. Theory development involves high-quality conceptualization that integrates and builds on existing knowledge, possibly using a multi-disciplinary approach. Further, especially in an applied research area such as business-to-business marketing, the emerging theory will have meaningful implications for managerial decision-makers. Insightful conceptual framework development advances theory substantially, not incrementally. Theoretical development can be either purely conceptual or based on empirical data. Nevertheless, there are comparatively few guidelines for the process of conceptual framework development. This editorial discusses pathways to developing conceptual frameworks to support academic research, with emphasis on business-to-business marketing research. As guidelines and conventions are available for data-driven approaches such as grounded theory, we focus on theorizing processes in which existing theory plays a pivotal role.

1. Introduction

Robust conceptual frameworks play a critical role in advancing academic and practical knowledge. Development of theory requires high-quality, novel conceptualizations and advancements that integrate existing theories, link research across disciplines, provide multi-level insights to move the field forward with substantial leaps rather than incremental steps. For academic researchers in business-to-business marketing, it is essential for emerging theory to be a source of practical insight to support decision-makers. Despite the importance and contribution of insightful conceptual frameworks, existing methods, books, and articles have seldom elaborated on how to develop these frameworks.. The issue of developing conceptual frameworks is relevant for different research approaches, both purely conceptual and those based on empirical data. As far as business-to-business marketing research is concerned, *Industrial Marketing Management*, as the leading business-to-business marketing journal, needs to take a leadership position by prioritizing the development of conceptual work, which meaningfully advances theory.

Conceptual frameworks can offer a substantive contribution to warrant publication on their own, without empirical data; or such frameworks can be motivated, illustrated, and fleshed out with empirical data. Articles of this type are typically labelled as conceptual or theoretical articles, as they use existing literature as their primary source for developing novel frameworks (Jaakkola, 2020) (see Fig. 1). Conceptual frameworks also can be developed through an abductive process where researchers move between theory and empirical data to develop a framework. The theorizing process is guided, but not determined, by existing theory, as is typical for many qualitative studies (e.g., Dubois & Gadde, 2002; Nenonen, Brodie, Storbacka, & Peters, 2017). In contrast, the role of empirical data is substantial when the focus is on testing conceptual frameworks. This is also the case when the framework is developed solely on the basis of empirical data (as in the development of grounded theory).

The purpose of this editorial is to consider explicitly various pathways to developing conceptual frameworks and how these pathways can be used to create new conceptual frameworks to support research, with a particular focus on business-to-business marketing. This editorial focuses on theorizing processes where existing theory plays a pivotal role (as shown in Fig. 1). Following Brodie and Peters (2020), we see conceptual frameworks emerging from interfaces between i) general theoretic perspectives and midrange theories; ii) multiple midrange theories, and iii) applied theories and midrange theories. Less attention is given to the more data-driven approaches such as grounded theory (e.g., Glaser & Strauss, 1967; Strauss & Corbin, 1990) because there are robust guidelines and established conventions to guide the theorizing.

To shed light on some of the challenges involved in developing, and writing about, conceptual frameworks, we frame our discussion around three specific styles of conceptual writing introduced by Cornelissen (2017) in his review of conceptual articles published in the *Academy of Management Review*. His examination identified three conceptual styles: i) articles that are centred on a set of propositions (propositional style), ii) articles that develop a process model (narrative style), and iii) articles that build or elaborate on a theoretical typology (typological style). We outline practical research design considerations (see Jaakkola, 2020) for each style and present an illustrative example of each style.

Insightful conceptual frameworks are essential to integrating existing knowledge and setting the agendas for future business-to-business marketing research. A significant weakness of many conceptual frameworks is that the theory used is too narrow in scope; and that the

https://doi.org/10.1016/j.indmarman.2020.04.005

Available online 15 April 2020

resulting conceptual development lacks strong theoretical foundations and fails to bridge theory and practice. We contribute value by providing researchers with templates and guidance in developing conceptual frameworks is a more explicit and systematic manner. Influential.

2. How theory informs the development of conceptual frameworks

Weick (1995), in his essay "What Theory is Not, Theorizing Is," laments that in organizational studies, there are few attempts to develop and use what he refers to as strong theory. In particular, he criticizes researchers of "lazy theorizing in which researchers try to graft theory onto stark sets of data" (p. 385). This occurs because there is confusion between theory as the outcome and theorizing as a process. Theory often is presented in the form of references, data, lists of diagrams, and hypotheses, which, while they are essential parts of the theorizing process, are not theory per se. Thus, when considering how theory informs the development of conceptual manuscripts, it is important place emphasis on the theorizing processes.

While the question "what is theory?" has been debated extensively, for the purpose of this editorial we use a simple general definition that "theory is a statement of concepts and their interrelationships that shows how and/or why a phenomenon occurs" (Corley & Gioia, 2011: p. 12). In a previous editorial, we drew on Brodie and Peters (2020) to distinguish between three levels of theoretical abstraction (Lindgreen, Di Benedetto, Brodie, & van der Borgh, 2020b).

General theories: These theories are conceptions and perspectives utilizing theory that is framed at the highest conceptual level and provides a perspective or logic of explanation for a domain. The theories are broad in scope, integrative, and context-free, and thus the theories do not directly lead to empirical investigation. The theories provide the foundations for understanding and explanation and are informed by a paradigmatic perspective.

Midrange theories: Midrange theories are context-specific, which relates to specific phenomena. Hence, midrange theories provide conceptual frameworks to undertake empirical observation and models to guide managerial practices. Most of the theories currently used in business-to-business marketing research have these characteristics, so midrange theories characterize most conceptual frameworks.

Applied theories: Applied theories are embedded in the domain of empirical research and the research context. While the focus of applied theories traditionally has been with empirical research, "theories-in-use" can play an important role (Argyris & Schon, 1974). "Theories-in-use" (TIU) recognize that practicing managers, customers, and other stakeholders in a service system use theory.

Zeithaml et al. (2020) argue that the TIU approach should be used to create theory that is developed from the mental models used by marketing stakeholders and thereby specific to the marketing issue being studied. The TIU approach builds on constructs, which are guided by marketing practitioners and grounded in marketing-specific contexts. Accordingly, this approach produces conceptual frameworks that is not only meaningful to marketing stakeholders, but can be communicated to them in language they use. Zeithaml et al. (2020) point out that this approach contrasts with the more commonly used approach of basing marketing theory on established academic theories developed in related disciplines. The authors note that this traditional approach limits the researcher's ability to find new, interesting marketing phenomena, ultimately widening the disconnect between academics and practitioners of marketing (Reibstein, Day, & Wind, 2009).

It has further been argued that neither traditional academic theory building, nor the TIU approach, optimally guides academics and practitioners in co-producing knowledge that is mutually beneficial (Crespin-Mazet & Ingemansson-Havenvit, 2020). Academics and practitioners inherently have different interests and contexts and require different kinds of knowledge (Di Benedetto, Lindgreen, Storgaard, & Clarke, 2019). To bridge this divide, the solution is to co-produce knowledge that is useful to both groups; this issue is known as the knowledge production problem (van de Ven & Johnson, 2006). To solve this problem, the interaction between academics and practitioners should produce context-specific knowledge; this knowledge will be useful to the extent that it can be combined with other contextual resources (Håkansson & Waluszewski, 2007). That is, both academics and practitioners should bring their own knowledge, competencies, and partner networks to their collaboration so that knowledge is not just co-produced, but also can be used and further produced in other contexts over time (Crespin-Mazet & Ingemansson-Havenvit, 2020).

In Fig. 2, we outline the domains of knowledge at three levels of theoretical abstraction. Midrange theory can be seen as the intermediary (bridging) body of theory that interfaces between the empirical and theoretical domains and hence are the foundation for conceptual frameworks. Within the theoretical and empirical domains, it is recognized that the boundaries between marketing and other management disciplines overlap. Thus, the theorizing processes in business-to-business marketing research can draw on these different disciplines.

The theorizing process for developing conceptual frameworks typically results in advancement of midrange theories and consists of the interplay between general theories and applied theories. However, prior to this theorizing process, it is essential to recognize the paradigmatic perspective that informs the theorizing process. As discussed by Lindgreen et al. (2020b), contemporary business-to-business marketing research mostly adopts a network or systems paradigmatic perspective leading to general theories that are inherently cross-disciplinary and drawing on sociological and institutional foundations. While a focal general theoretic perspective can interface directly with midrange theory to develop a conceptual framework, the theorizing process also can be informed from other general theoretic perspectives and other midrange theories.

Recently, Jaakkola (2020) has elaborated on the process of theorizing for developing conceptual frameworks. She distinguishes between two starting points. One way to start is with a focal phenomenon that is observable, but not adequately addressed in the existing research. The researcher inductively develops a conceptual framework (midrange theory) in terms of particular concepts that reflect the phenomena. Nenonen et al. (2017) recognize that insight about the focal phenomena can come not only from researchers' observations but can also be initiated with researchers interfacing with managers and other actors involved with practice. An alternative starting point for developing conceptual manuscripts begins with a focal (midrange) theory and extends and refines this theory to reflect the phenomena of interest better. The process can be enhanced by taking into account the meta-level conceptual system provided by the general theory. As discussed by Lindgreen et al. (2020b) and Brodie and Peters (2020), the process of theorizing for developing conceptual frameworks should not be considered linear, but should be seen as iterative, drawing on pathways.

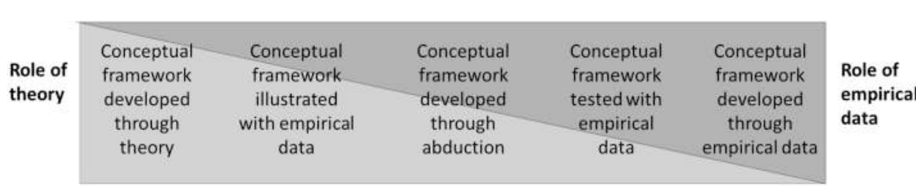

Fig. 1. Role of Theory vs. Empirical Data in Developing Conceptual Frameworks.

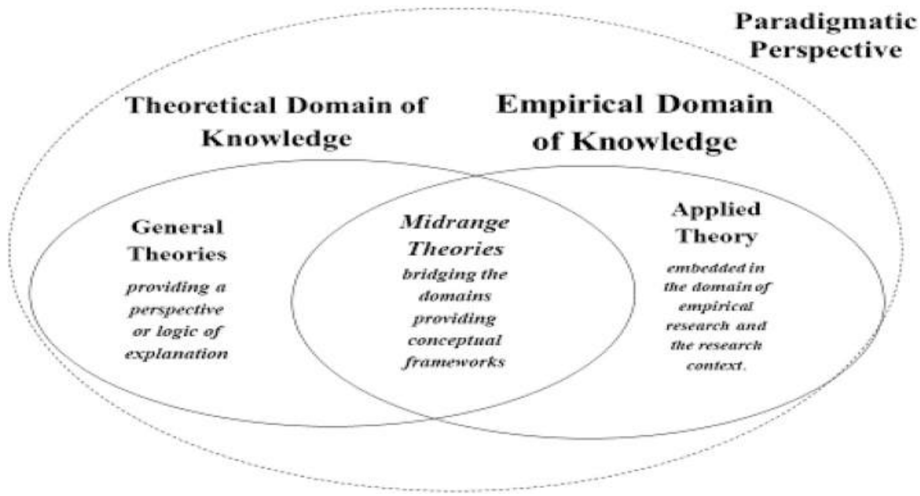

Fig. 2. Domains of Knowledge and Levels of Theory.
Source: Lindgreen et al. (Lindgreen, Di Benedetto, Brodie, & van der Borgh, 2020a; Lindgreen et al., 2020b, p. 2).

3. Styles for developing conceptual frameworks

To identify the different ways for developing conceptual frameworks, Cornelissen (2017) examined articles published in the *Academy of Management Review*. While he did not find one straightforward formula, he does identify three styles:

1. *Proposition-based style*: The statement of theoretical propositions that introduces new constructs and cause-effect relationships.
2. *Narrative-based style*: The specification of a process model that lays out a set of mechanisms explaining events and outcomes.
3. *Typology-based style:* The specification of a typology that interrelates different dimensions to flesh out new constructs and causal interactions.

To develop guidelines for these styles, Cornelissen (2017) went through the reviewer reports and editorial letters for all of the *Academy of Management Review* articles he had handled as the editor for this journal. These styles and associated guidelines apply to advance the craft of developing conceptual frameworks. We also draw on four alternative templates for conceptual research design identified by Jaakkola (2020): *Theory Synthesis, Theory Adaptation, Typology*, and *Model* that can guide the theorizing process for developing conceptual frameworks. Next, we will discuss the three styles identified by Cornelissen (2017).

The *propositional style* refers to a theoretical framework that outlines a set of formally stated theoretical propositions (Cornelissen, 2017). These propositions can introduce new constructs and cause-effect relationships. This type of article develops a conceptual framework that takes the form a research model detailing the antecedents, outcomes, and contingencies related to the focal construct (MacInnis, 2011). The propositional style suggests making claims about causal relationships and specifying testable relationships. However, Cornelissen (2017) suggests that insightful conceptual frameworks should make propositions that cover novel theoretical ground rather than merely summarize prior literature. The creative scope of such arguments is wider in articles that present the conceptual framework as their primary outcome, drawing on theoretical or empirical domain of knowledge to model emerging phenomena instead of testing well-charted constructs (Yadav, 2010). The researcher should carefully justify the choice and role of different sources of knowledge in building the propositions: typically, the literature that addresses key elements of the phenomenon/concept to be explained is informed by another theory that enables the explanation of relationships between the studied variables (Jaakkola, 2020).

The narrative-based style of developing conceptual frameworks focuses on specifying a process model that lays out a set of mechanisms explaining events and outcomes (Cornelissen, 2017). This style represents a form of theorizing that emphasizes narrative reasoning that seeks to unveil “big picture” patterns, connections, and mechanisms rather than specific causal relationships (Cornelissen, 2017; Delbridge & Fiss, 2013). This type of conceptual framework is often a process model involving the dynamics of constructs and critical events or turning points for the phenomenon. The framework contributes to extant knowledge by not only describing what is known, but making novel arguments about how a concept changes, the processes by which it operates, or why and how elements of a process lead to a particular outcome (Cornelissen, 2017; MacInnis, 2011). This type of framework can develop by synthesizing existing literature across multiple theoretical perspectives to form novel, higher-order understanding. It can also come from problematizing an existing theory and resolving the identified shortcomings by introducing a new theoretical lens that enables organizing the elements of the studied process in a better way (Jaakkola, 2020).

The typology-based style aims to logically and causally combine different constructs into a coherent and explanatory set of types (Cornelissen, 2017). A typology provides a more precise and nuanced understanding of a phenomenon or concept, as the typology dimensionalizes or categorizes existing knowledge of a phenomenon or construct (Jaakkola, 2020; MacInnis, 2011). As a theoretical framework, a typology delineates how variants of an entity differ and may help to recognize the entity’s differing antecedents, manifestations, or effects (MacInnis, 2011) and causal relationships (Fiss, 2011). When building a typology, the researcher should carefully justify the logic of identifying dimensions of types. The dimensions of a typology can be identified by applying some general theory, or other midrange theories that are equipped to explain logically the differences between variants of the concept (Jaakkola, 2020). Another option is to tease out relevant dimensions through iterations between theories and knowledge in the empirical domain.

The three styles presented by Cornelissen (2017) ultimately aim at explaining relationships between concepts to answer questions *why, how*, and *when* something happens. The propositional style results in a research model depicting cause-effect relationships; the typology outlines variants of a concept that have different drivers, outcomes, or contingencies; and the narrative style lays out sequences of events (Fig. 3). According to Jaakkola (2020), conceptual research also can be designed to increase understanding, thus answering the ‘*what*’ question. A *theory synthesis* framework seeks to achieve conceptual integration

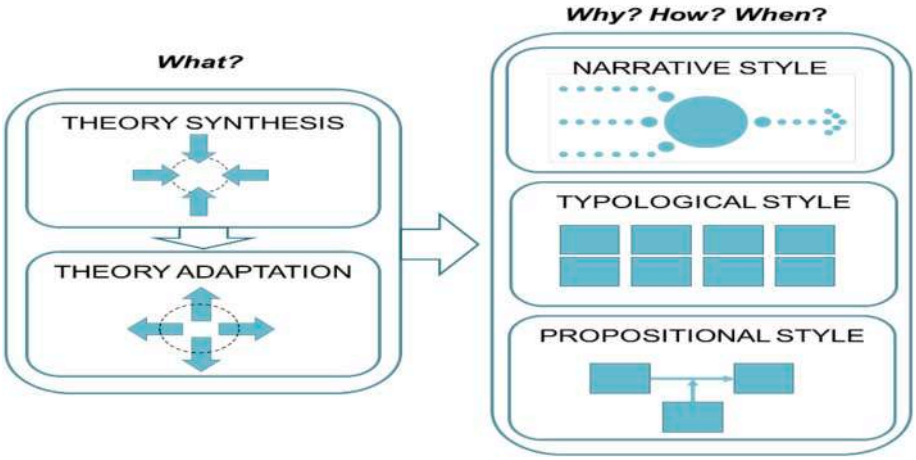

Fig. 3. Styles of Theorizing for Building Conceptual Frameworks.

across multiple theories or fragmented literature streams, to offer and enhance the view of a concept or phenomenon by linking previously unconnected elements in a novel way (Jaakkola, 2020). Theory synthesis frameworks summarize and integrate existing knowledge into a manageable whole and produce a new higher-order understanding of the concept under study (MacInnis, 2011). Another approach is to broaden, extend, or amend an existing theory by using other theories. *Theory adaptation* aims at revising extant conceptualizations by introducing alternative frames of reference to propose a novel, enhanced perspective (Jaakkola, 2020; MacInnis, 2011). For example, the researcher might draw from practical insights, that is, TIU, or other midrange theories to argue that an existing conceptualization is insufficient or conflicted, and suggest that broadening of perspective or scope is needed to align better the concept to its purpose (Jaakkola, 2020; Nenonen et al., 2017). Research aiming at theory synthesis or adaptation often serves as a stepping stone towards building frameworks that can explain (Fig. 3).

4. Examples

In this section, we discuss three business-to-business marketing studies, each of which is illustrative of one of the styles of developing conceptual frameworks (Cornelissen, 2017).

4.1. Proposition-based style

The first study, by Du, Swaen, Lindgreen, and Sen (2013), examines the interrelationship between leadership styles and institutional corporate social responsibility (CSR) practices. The authors noted the critical role of organizational leadership style in developing organizational strategy, yet the shortage of research of how leadership style affects the practice of CSR. In particular, the authors investigated the relationship between transformational and transactional leadership styles and organizational CSR outcomes.

Du et al. (2013) developed a theoretical framework (Fig. 4) based on their literature review of the literature on transformational leadership. They proposed three testable hypotheses, which can be summarized as follow:

First, transformational leaders are most likely to recognize the interrelationships and interdependencies between the organization's stakeholders, including the local community and the natural environment. Therefore:

H1. : Transformational (not transactional) leadership is positively associated with an organization's institutional CSR practices.

Second, stakeholder-oriented marketing provides a wider

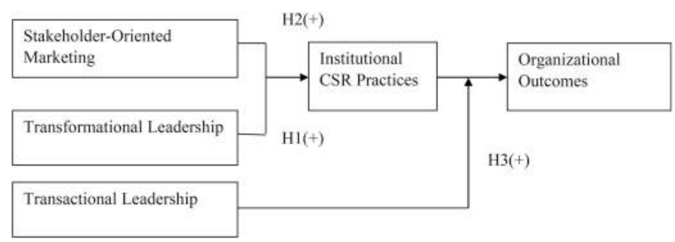

Fig. 4. Conceptual Framework of Du et al. (2013).
Source: Adapted from Du et al. (2013, p. 160).

environmental view, which provides deeper knowledge of the organization's stakeholders and their concerns. Therefore, transformational leaders can form even stronger relationships with stakeholders, and can work with stakeholders to provide CSR practices that better suit the needs of the stakeholders. Du et al. (2013) hypothesized a moderation effect:

H2. : Stakeholder-oriented marketing positively moderates the relationship between transformational leadership and institutional CSR practices (that is, the relationship is more positive for organizations effectively practicing stakeholder-oriented marketing).

Finally, the value provided to secondary stakeholders is a societal impact allowing institutional CSR to generate positive organizational outcomes. If the organization uses its core competencies appropriately, the organization will be able to effectively implement its corporate CSR initiatives. This is also modeled as a moderation effect:

H3. : Transactional (not transformational) leadership positively moderates the relationship between institutional CSR practices and organizational outcomes (that is, the relationship is more favorable for organizations with higher transactional leadership).

Based on a survey of managers in 440 U.S. organizations, Du et al. (2013) found support for their new conceptual model. Transformational leadership was positively related to institutional CSR practices, and stakeholder-oriented marketing did moderate this positive relationship. By contrast, transactional leadership positively moderated the relationship between institutional CSR practices and organizational outcomes. Thus, all hypotheses were supported, and the authors concluded that transformational and transactional leadership styles affect institutional CSR practices differently.

Du et al. (2013) also identified several important managerial implications derived from their conceptual model. Transformational leadership is most appropriate for initiating CSR practices, while transactional leadership may have an advantage in deriving the organizational benefits from these practices. The authors noted that both types of leadership are needed in order for the organization to achieve the "circle of virtue" (the organization investing in CSR achieves its business objectives and therefore, can ensure sustained CSR investment). The authors recommended that managers should consider how both transformational and transactional leadership can be implemented. Also, stakeholder-oriented marketing has significant implications in that organizational members must keep the welfare of all stakeholders in mind and create an organizational climate that fosters CSR practices. Stakeholder-oriented marketing supports transformational leadership and increases its effectiveness in promoting CSR practices. This finding suggests that transformational leaders should consider developing complementary capabilities that help them attain their institutional CSR objectives.

The Du et al. (2013) study is a clear illustration of Cornelissen's (2017) proposition-based style. The authors identified a significant gap in the literature: the organizational leadership literature had not investigated the effects of leadership on organizational CSR policy despite the growing importance of the latter. They proposed a new conceptual framework, built on organizational leadership theory, which included new cause-and-effect relationships among constructs from leadership theory to organizational CSR practices and outcomes. It is also notable that the authors employed the remedies recommended by Cornelissen (2017) for the proposition-based style. They develop and empirically test their conceptual model. They also use a broad theoretical perspective that required cross-disciplinary inputs from different business disciplines (organizational behaviour, marketing, and corporate strategy).

4.2. Narrative-based Style

To illustrate the narrative-based style, consider Vallaster, Maon, Lindgreen, and Vanhamme (2020), a multiple-case qualitative study of for-profit hybrid organizations. Due to their hybrid nature, sustainability-driven hybrids design organizational activities in line with social and environmental objectives and economic objectives simultaneously, which lead to tension. The most successful hybrids will need to manage these tensions, yet the process by which they accomplish this has been under-researched in the literature. A much greater understanding of this process is warranted, from the viewpoints of both individual actions and collective organizational practices. To gain the required depth of understanding, Vallaster et al. (2020) conducted a qualitative study of for-profit hybrids.

Based on a review of the literature, the authors developed an initial conceptual framework. The existing literature discussed the tensions found within for-profit organizations attempting to transform their industrial and social environments, identifying four categories: learning, belonging, organizing, and performing tensions. To handle these tensions, organizations need to develop the ability to deal with specific issues that arise from each of these four categories; the literature provided some discussion of dynamic capabilities required for the organization to integrate, build, and transform internal and external resources in response to changing environmental conditions. Specifically, the literature mentioned sensing, seizing, and transforming dynamic capabilities. Nevertheless, little research on how these dynamic capabilities are developed and applied by for-profit hybrids was found.

To assess empirically how actors at the individual and collective levels develop these capabilities in a for-profit hybrid setting, they used practice-based theory to conceptualize the practices (individual and collective behaviors, activities, and processes) undertaken by actors to address each of the specific tensions and, ultimately, to create economic, social, and environmental value.

Next, Vallaster et al. (2020) applied a theory-generative approach, carrying out a qualitative study comprising several stages over a 15-month period. Each stage comprised several workshops. Stage A was designed to understand industry context, and experience in handling multiple goals. During Stage B, middle managers were interviewed about the activities, processes, and capabilities involved in developing a for-profit hybrid orientation. Stage C was devoted to identifying challenges, which had arisen from the initiatives undertaken. Post-workshop interviews, personal reflections, and diaries were used to take notes and to drive discussion at upcoming workshops. Finally, in Stage D, supplementary interviews with managers from other for-profit hybrids were conducted.

The theory-generative approach involved several phases. First, initial first-order codes were identified by the authors from the interview findings. Second, related concepts across case organizations were identified and linked by the authors, creating second-order concepts. These concepts primarily involved how the respondents addressed the tensions arising from their organization's hybrid orientation. Third, the authors conducted a cross-case analysis to identify consistent patterns. Finally, the authors completed the theoretical framework by refocusing on the hybrid-related tensions identified in the existing literature, and determining how dynamic capabilities and micro-foundations addressed these tensions.

The result of the analysis was a new conceptual framework of the micro-foundations of the dynamic capabilities of for-profit hybrids. Fig. 5 shows the four dynamic capabilities of for-profit hybrids (sensing, seizing, transforming, and liaising; extant literature had not previously discussed the dynamic capability of liaising), and how these dynamic capabilities are supported by micro-foundations. Table 1 indicates the micro-foundations that were identified and how they aligned with specific for-profit hybrid tensions. The authors also draw several managerial implications. For-profit hybrid managers should recognize the need to constantly nurture both the individual and collective practices that support the micro-foundations. It is also important to monitor hybridity-related tensions since not all practices will be suitable for all business decisions. The organization should be able to capitalize on economic opportunities and also identify sustainability-oriented solutions.

In sum, Vallaster et al. (2020) provides a good illustration of the narrative-based style of conceptual model. The authors started with a 'theory adaptation' approach (Jaakkola, 2020), as they combined the dynamic capabilities theory with practice theory to conceptualize how actors develop capabilities, and then informed this theory-based understanding with knowledge from the empirical domain (cf. Nenonen et al., 2017). As a result, the authors developed a useful managerial framework that presents the underlying micro-foundations leading to dynamic capabilities. Managers of for-profit hybrids can make use of the framework to improve current practices, identify which practices are lacking, and ultimately support sustainable value creation. While some components of the model were available in the existing literature, the process by which for-profit hybrids could best manage ongoing tensions was not yet well understood. This conceptual framework helps managers understand the individual actions and collective practices that best support environmental and social objectives.

4.3. Typology-based Style

In this study of purchasing practices, by Lindgreen et al. (2013), the authors develop a new framework and a measurement instrument. The purchasing literature usually assumes two categories of purchasing practices (transactional and relational), though, in practice, purchasing often involves both types. The issue of how and why organizations choose one type or the other, or to combine types, and how these choices affect organizational performance is not well understood. To remedy this situation, Lindgreen et al. (2013) needed to develop a new framework, based on the existing literature, which includes a broader range of purchasing practices, as well as a new measurement instrument to measure organizational purchase practice with greater precision.

The existing literature described two purchasing management practices. Transaction purchasing emphasizes the aggressive and continuous search for new suppliers to achieve the best terms. In

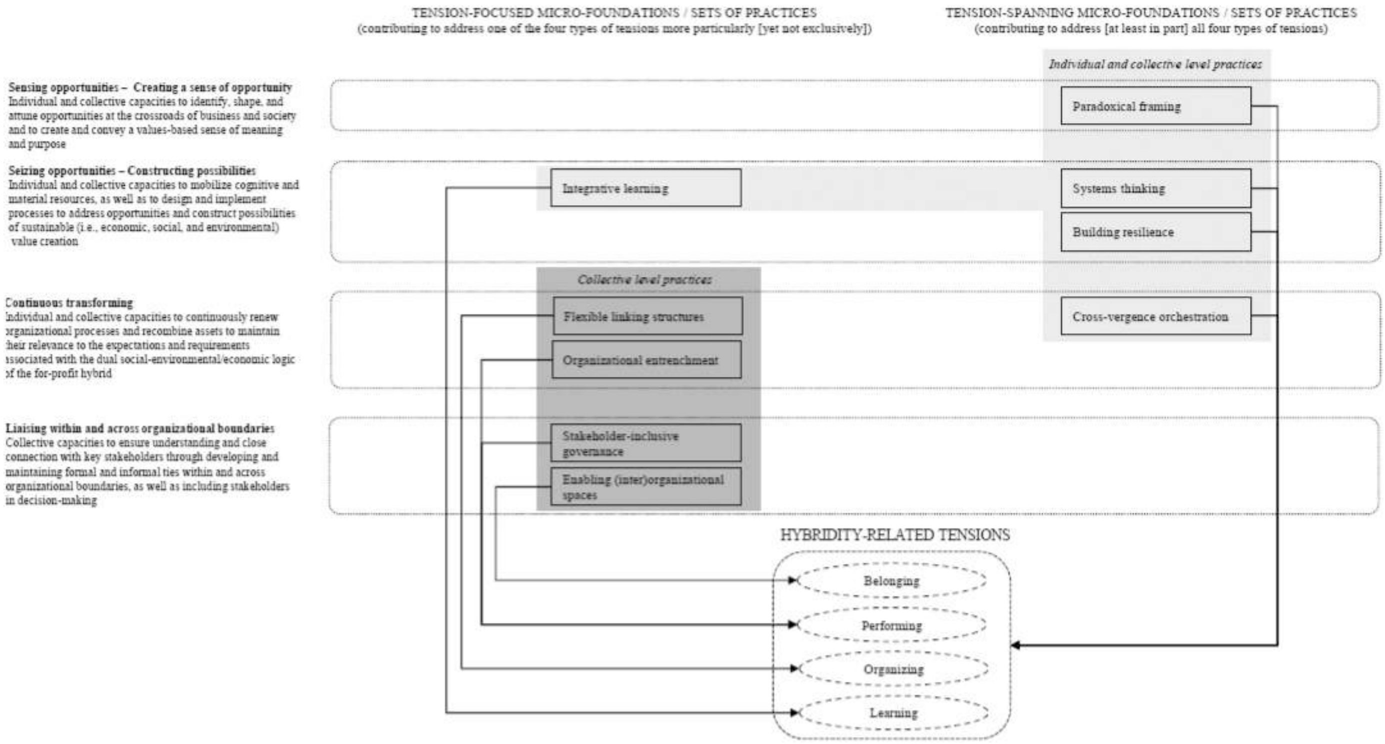

Fig. 5. Addressing Hybridity-Related Tensions through Micro-Foundations of Dynamic Capabilities.
Source: Vallaster et al. (2020, in print).

Table 1
Micro-foundations of dynamic capabilities of for-profit hybrids.

Micro-foundations	Individual or collective
Sensing opportunities: Creating a sense of opportunity	
Experiential /grounded scouting	*Individual*
Attention to functional core	*Individual*
Paradoxical framing	*Individual & collective*
Seizing opportunities: Constructing possibilities	
Systems thinking	*Individual & collective*
Bending institutional norms	*Individual & collective*
Integrative learning	*Individual & collective*
Building resilience	*Individual & collective*
Continuous transforming	
Cross-vergence orchestration	*Individual & collective*
Flexible linking structures	*Collective*
Organizational entrenchment	*Individual & collective*
Liaising within and across organizational boundaries	
Enabling (inter) organizational spaces	*Collective*
Stakeholder-inclusive governance	*Collective*

Notes: Shaded cells relate to practices and micro-foundations that do not contribute to addressing hybridity-related tensions.
Source: After Vallaster et al. (2020, in print).

contrast, network purchasing refers to the organization relying on a more extensive organizational system for purchasing from suppliers. Lindgreen et al. (2013) suggested adding two practices. Electronic purchasing refers to the use of the Internet or other one-to-one or one-to-many technologies to support the supplier relationship, and interactive purchasing refers to interpersonal interaction between organizational employees and suppliers. Further, Lindgreen et al. (2013) identified eight formative indicators, which together describe the four different kinds of purchasing practices. These formative indicators are: purpose of exchange, managerial intent, nature of communication, type of contact, duration of exchange relationship, formality of exchange, managerial focus, and managerial investment. The authors also identified a general indicator that provides an overall view of each purchasing practice. Table 2 shows the four purchasing practices and how they are characterized by the formative indicators.

The authors surveyed 202 purchasing managers in the U.S. Each organization was scored on each of the eight indicators, for all four purchasing practices. By summing the eight indicators for each purchasing practice, an index was created, which indicates the extent to which each organization practices transaction, electronic, interactive, and network purchasing. The purchasing practice types are not mutually exclusive, so, as a result, each organization will have its characteristic mix of indices.

Lindgreen et al. (2013) used cluster analysis to group organizations into configurations of purchasing practice, based on these indices. This procedure resulted in a new conceptual framework of purchasing practices, consisting of four identifiable clusters or patterns:

Transactional configuration: High on transaction purchasing index and low on the other three indices.

Integrative relational configuration: Low on transaction purchasing index and high on the other three indices.

Interpersonal dyadic configuration: Medium on transaction purchasing index, high on interactive purchasing index, and low on the other indices.

Interpersonal network configuration: Medium on transaction purchasing index, high on interactive purchasing and network purchasing indices, and low on electronic purchasing index.

The authors also gathered marketing performance outcomes (customer attraction, retention, and satisfaction, sales growth, and market share), as well as a financial performance outcome (profitability). Therefore, once the framework was in place, the authors were able to link purchasing practices to performance outcomes in a very detailed manner. For example:

Electronic purchasing practices needed to be combined with the interactive purchasing and network purchasing practices in order to achieve high levels of performance.

Many organizations are pluralistic; that is, they utilize two or more purchasing practices.

The organizations that use the integrative relational configuration

Table 2
Indicators Pertaining to Purchasing Practices.

Aspects	Transactional Perspective	Relational Perspective		
	Transaction Purchasing	Electronic Purchasing	Interactive Purchasing	Network Purchasing
Purpose of exchange: When dealing with our direct suppliers, our purpose is to:	Achieve cost savings or other 'financial' measure(s) of performance (monetary transactions)	Create information-generating dialogue with many identified suppliers	Build a long-term relationship with specific supplier(s)	Form relationships with a number of organizations in our supply market(s) or wider purchasing system
Nature of communication: Our communication with direct suppliers can be characterized as:	Our organization using undifferentiated communications with all suppliers	Our organization using technology to communicate with and possibly among many individual suppliers	Individuals at various levels in our organization personally interacting with individual suppliers	Senior managers networking with other managers from a variety of organizations in our supply market(s) or wider purchasing system
Type of contact: Our organization's contact with our direct suppliers is:	Arm's-length, impersonal with no individualized or personal contact	Interactive via technology such as the Internet	Interpersonal (e.g. involving one-to-one interaction between people)	Across firms in the broader network (from impersonal to interpersonal contact)
Duration of exchange: The type of contact with our direct suppliers is characterized as:	Transactions that are discrete or one-off (i.e. not ongoing)	Technology-based interactivity that is ongoing and real-time	Interpersonal interaction that is ongoing	Contact with people in our organization and wider purchasing system that is ongoing
Formality of exchange: hen people from our organization meet with our direct suppliers, it is:	Mainly at a formal business level	Mainly at a formal level, yet customized and/or personalized via interactive technologies	At both a formal business level and informal social level on a one-to-one basis	At both a formal business level and informal social level in a wider organizational system / network
Managerial intent: Our purchasing exchanges are intended to:	Continuously search for new suppliers to find the best deal (i. e., low prices)	Create two-way, technology-enabled data exchanges with our suppliers	devElop cooperative relationships with our suppliers	Coordinate activities between ourselves, suppliers, and other parties in our wider purchasing and supply system (e.g., second-tier suppliers, key customers, service providers, and other organizations with which we interact through our purchasing activities)
Managerial focus: Our purchasing strategy is focused on issues related to:	The purchase item and its price	Managing IT-enabled relationships with many individual suppliers	One-to-one relationships with suppliers, or individuals in supplier organizations we deal with	The network of relationships between individuals and organizations in our wider supply system
Managerial investment: Our purchasing resources (i.e. people, time, and money) are invested in:	Specifying products, negotiations, ordering, and expediting activities	Operational assets (IT, website, logistics) and functional systems integration (e.g., purchasing with IT)	Establishing and building personal relationships with individual suppliers	Developing our organization's network relationships within our supply market (s) or wider purchasing system
General indicator: Overall, our organization's general approach to our direct suppliers (of product-related items) involves:	Using aggressive sourcing (continuously search for new suppliers) to obtain purchase items at the most favorable conditions	Using the Internet and other interactive technologies to create and mediate data exchanges between our organization and our suppliers	Developing personal interactions between employees and individual suppliers	Positioning organization within a wider organizational system or network

Source: Lindgreen et al. (2013, p. 76).

outperform other organizations on all the marketing and financial performance outcome measures.

The organizations that use the integrative relational configuration outperform all or most other organizations on some of the specific formative indicators as well, such as supplier quality and delivery reliability. However, in the case of supplier lead time, there were no noticeable differences among the configurations.

Organizations tend to use interactive and network purchasing more with direct suppliers, and transactional purchasing more with indirect suppliers.

Lindgreen et al. (2013) were able to draw several critical managerial implications from their conceptual framework. A manager can, for example, set targets for each type of purchasing practice to use per cluster, and evaluate performance outcomes. Gaps between actual and target achievement levels can be identified, and ways to bridge the gaps can be discussed. Organizations can also use this information to specify (and, when necessary, adjust and re-specify) their strategies on how to achieve purchasing practice and performance outcomes.

In sum, this study effectively illustrates the typology-based style of Cornelissen (2017). The previous literature tended to view only two types of purchasing practices (transactional and relational), without delving into how the organization chooses one or the other, or whether to mix types. Lindgreen et al. (2013) built on this literature stream and extended it in meaningful directions. First, the relational purchasing practices type was defined too broadly, so they essentially split this type into three new purchasing categories (electronic, interactive, and network purchasing) each with different characteristics. The authors identified a set of specific formative indicators, developed a new measurement instrument, and empirically developed a typology of purchasing patterns based on cluster analysis. Finally, they identified relationships between the purchasing patterns and performance outcomes. Overall, the new cluster-based typology resulted in a more precise view of purchasing practices and key performance metrics, making an important conceptual contribution and offering actionable managerial implications as well.

5. Guidelines for developing strong conceptual frameworks

In a previous editorial, Lindgreen et al. (2020a) discussed which types of articles typically get cited. One type of article that tends to get highly cited is the article that introduces a new conceptual framework. Di Benedetto et al. (2019) consider the necessary conditions that determine the success of authors' conceptual and theoretical development. In a summary of these conditions, the authors highlight the following points:

- Develop a clear and convincing logic to their theory so that researchers can see how the theory fits in the field,
- Define concepts clearly and concisely so that other researchers can use them in their own research,

- Ensure that there is a clear rationale for the conceptual development so that other researchers can understand why they should use the concepts, methods, or theories,
- Ensure that propositions and hypotheses are specific, well-argued, grounded in theory, and not tautological.

For any type of conceptual framework, it is elementary that the authors explicate and justify the choice of theories and concepts, as well as the role those different domains of knowledge play in the analysis (Jaakkola, 2020). There is no single best template for building a conceptual framework. Still, authors can use general theories, midrange theories, and theories-in-use in many different ways, as long as they make their approach clear for the reader. For example, the article should communicate if empirical data was used to illustrate a theoretical framework developed through conceptual analysis, or if the elements of the framework are derived from empirical data.

In an examination of how to undertake cross-disciplinary research, Lindgreen et al. (2020b) recognize that there are multiple pathways to develop midrange theory and hence undertake empirical research. First, the focal general theoretic perspective can interface directly with midrange theory. The second option is that other general theoretic perspectives provide pathways that can lead to other midrange theories, which then leads to a focal midrange theory that can be used in business-to-business research (Lindgreen et al., 2020b, p. 1).

One pathway to develop midrange theory is by having managerial practices inform research processes. For a more in-depth discussion of this pathway, we refer to the editorial by Lindgreen et al. (2020b). However, little is known about university-business collaborations. Thus, Di Benedetto et al. (2019) discuss this type of collaboration including, for example, offering advice to business managers about how to collaborate with university academics.

In addition to these more general guidelines, Table 3 outlines more specific guidelines related to the three styles discussed in this editorial.

As outlined in Table 3, the common problem with these theorizing styles for developing conceptual manuscripts was that the theory used is too narrow in scope. Thus, the resulting conceptual development lacks strong theoretical foundations. Cornelissen's (2017) remedies for all three theorizing styles are to introduce a stronger theorizing process that facilitates the interface between general theory and applied theory and to explicitly recognize the social causal mechanisms that underpin midrange theory (Mason, Easton, & Lenny, 2013).

Table 3
Three Styles of Developing Theoretical Frameworks and Associated Guidelines.

Attributes of Each Style	Proposition-based Style	Narrative-based Style	Typology-based Style
Definition	The statement of theoretical propositions that introduces new constructs and cause-effect relationships	The specification of a process model that lays out a set of mechanisms explaining events and outcomes	The specification of a typology that interrelates different dimensions to flesh out new constructs and causal interactions
Basic form	Identify cause-effect relationships that act as broad signposts and implications for further research	Provide generalized causal mechanism, as the underlying storyline of a process model	Explains the fuzzy nature of many subjects by combining different constructs into a coherent and explanatory set of types
Common problems	• Propositions are too narrow in scope and merely summarize the prior literature • Propositions include multiple clauses • Propositions lack detail on the causal agent	• Narrative and process model are too descriptive • Narrative and process model lack explanatory detail • Narrative features stylized arguments and claims (lacking nuance and contingent variation) • Narrative features complex compounds and phrases as constructs	• Typology is descriptive and does not offer multidimensional ideal types • Typology only systematizes and summarizes existing research but lacks explanation • Typology features various degrees of causal entanglement (including circularity and tautology)
Remedies	• Broaden the scope of the propositions and develop an original line of argument, with a novel set of assumptions as theorized grounds • Develop the arguments first, before formalizing them into propositions	• Elaborate the underlying conceptual linkages of a process model, foregrounding a clear mechanism or set of mechanisms • Add details and more • contingent variation to the overall narrative, strengthening its explanatory potential	• Identify whether the proposed typology has a review or theory contribution, or both • Develop the typology from a theoretical angle, incorporating multiple theoretical dimensions • Draw out patterns of causality (using fuzzy set reasoning) and explicate the basic line of argument

Adapted from Cornelissen (2017).

6. Conclusions

The purpose of this editorial is to draw attention to the important but demanding craft of developing insightful theoretical frameworks. We urge authors to consider the role of different domains of knowledge in building their frameworks, and explain their approach clearly in the article. Too often, reviewers face manuscripts have conceptual frameworks based on descriptive literature reviews, which devoid of more in-depth conceptual analysis or integration. Hopefully, business-to-business marketing scholars will be inspired by the different styles for building theoretical frameworks discussed in this editorial and make use of the guidelines and research design considerations we have outlined.

Acknowledgments

This editorial is part in a series of the co-editors-in-chiefs' reflections on important aspects of planning, undertaking, and publishing research in business-to-business marketing management, reflections that should help prospective and early-career researchers eventually to see their findings published in *Industrial Marketing Management* and other top journals.

References

Argyris, C., & Schon, D. (1974). *Theory in practice: Increasing professional effectiveness*. Oxford, U.K.: Jossey-Bass.
Brodie, R. J., & Peters, L. (2020). Increasing contribution in service research: Fresh thinking about the process of theorizing. *Journal of Services Marketing* (in press).
Corley, K. G., & Gioia, D. A. (2011). Building theory about theory building: What constitutes a theoretical contribution? *Academy of Management Review, 36*(1), 12–32.
Cornelissen, J. (2017). Editor's comments: Developing propositions, a process model, or a typology? Addressing the challenges of writing theory without a boilerplate. *Academy of Management Review, 42*(1), 1–9.
Crespin-Mazet, F., & Ingemansson-Havenvit, M. (2020). Rethinking the theory-practice divide: How academia-industry collaboration contributes to theorizing. *Industrial Marketing Management*. https://doi.org/10.1016/j.indmarman.2020.01.003.
Delbridge, R., & Fiss, P. C. (2013). Editors' comments: Styles of theorizing and the social organization of knowledge. *Academy of Management Review, 38*(3), 325–331.
Di Benedetto, C. A., Lindgreen, A., Storgaard, M., & Clarke, A. H. (2019). Editorial: How to collaborate really well with practitioners. *Industrial Marketing Management, 82*, 1–8.
Du, S., Swaen, V., Lindgreen, A., & Sen, S. (2013). The roles of leadership styles in corporate social responsibility. *Journal of Business Ethics, 114*(1), 155–169.
Dubois, A., & Gadde, L.-E. (2002). Systematic combining: An abductive approach to case research. *Journal of Business Research, 55*(7), 553–560.

Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. *Academy of Management Journal, 54*(2), 393–420.
Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine Publishers.
Håkansson, H., & Waluszewski, A. (Eds.). (2007). *Knowledge and innovation in business and industry: The importance of using others*. London: Routledge.
Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*. https://doi.org/10.1007/s13162-020-00161-0.
Lindgreen, A., Di Benedetto, C. A., Brodie, R. J., & van der Borgh, M. (2020a). Editorial: How to get great research cited. *Industrial Marketing Management, 89*. https://doi.org/10.1016/j.indmarman.2020.03.023.
Lindgreen, A., Di Benedetto, C. A., Brodie, R. J., & van der Borgh, M. (2020b). Editorial: How to undertake great cross-disciplinary research. *Industrial Marketing Management, 90*. https://doi.org/10.1016/j.indmarman.2020.03.025.
MacInnis, D. J. (2011). A framework for conceptual contributions in marketing. *Journal of Marketing, 75*(4), 136–154.
Mason, K., Easton, G., & Lenny, P. (2013). Causal social mechanisms: From the what to the why. *Industrial Marketing Management, 42*(3), 347–355.
Nenonen, S., Brodie, R. J., Storbacka, K., & Peters, L. D. (2017). Theorizing with managers: How to achieve both academic rigor and practical relevance? *European Journal of Marketing, 51*(7/8), 1130–1152.
Reibstein, D. J., Day, G., & Wind, J. (2009). Guest editorial: Is marketing academic losing its way? *Journal of Marketing, 73*(4), 1–3.
Strauss, A. L., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage Publications.
Vallaster, C., Maon, F., Lindgreen, A., & Vanhamme, J. (2020). Serving multiple masters: The role of micro-foundations of dynamic capabilities in addressing tensions in for-profit hybrid organizations. *Organization Studies*. https://doi.org/10.1177/0170840619856034.
van de Ven, A. H., & Johnson, P. E. (2006). Knowledge for theory and practice. *Academy of Management Review, 31*(4), 802–821.
Weick, K. E. (1995). What theory is not theorizing is. *Administrative Science Quarterly, 40*(3), 385–390.
Yadav, M. S. (2010). The decline of conceptual articles and implications for knowledge development. *Journal of Marketing, 74*(1), 1–19.
Zeithaml, V. A., Jaworski, B. J., Kohli, A. K., Tuli, K. R., Ulaga, W., & Zaltman, G. (2020). A theories-in-use approach to building marketing theory. *Journal of Marketing, 84*(1), 32–51.

Adam Lindgreen[a,e,*,1], C. Anthony Di Benedetto[b], Roderick J. Brodie[c], Elina Jaakkola[d]

[a] *Department of Marketing, Copenhagen Business School, Solbjerg Plads 3, 2000 Frederiksberg C, Denmark*
[b] *Fox School of Business, Temple University, Alter Hall 523, 1801 Liacouras Walk, Philadelphia, PA 19122-6083, USA*
[c] *Department of Marketing, University of Auckland Business School, Owen G. Glenn Building, 12 Grafton Road, Auckland 1010, New Zealand*
[d] *Turku School of Economics, University of Turku, Rehtorinpellonkatu 3, 20500 Turku, Finland*
[e] *University of Pretoria's Gordon Institute of Business Science, 26 Melville Road, Illovo, Johannesburg, South Africa*

[*] Corresponding author at: Department of Marketing, Copenhagen Business School, Solbjerg Plads 3, 2000 Frederiksberg C, Denmark.
E-mail addresses: adli.marktg@cbs.dk (A. Lindgreen), tonyd@temple.edu (C.A. Di Benedetto), r.brodie@auckland.ac.nz (R.J. Brodie), Elina.jaakkola@utu.fi (E. Jaakkola).

[1] All authors contributed equally.

Article

AMERICAN MARKETING ASSOCIATION

Journal of Marketing
2020, Vol. 84(1) 1-25

DOI: 10.1177/0022242919873106
journals.sagepub.com/home/jmx
SAGE

Uniting the Tribes: Using Text for Marketing Insight

Jonah Berger, Ashlee Humphreys, Stephan Ludwig, Wendy W. Moe, Oded Netzer, and David A. Schweidel

Abstract

Words are part of almost every marketplace interaction. Online reviews, customer service calls, press releases, marketing communications, and other interactions create a wealth of textual data. But how can marketers best use such data? This article provides an overview of automated textual analysis and details how it can be used to generate marketing insights. The authors discuss how text reflects qualities of the text producer (and the context in which the text was produced) and impacts the audience or text recipient. Next, they discuss how text can be a powerful tool both for prediction and for understanding (i.e., insights). Then, the authors overview methodologies and metrics used in text analysis, providing a set of guidelines and procedures. Finally, they further highlight some common metrics and challenges and discuss how researchers can address issues of internal and external validity. They conclude with a discussion of potential areas for future work. Along the way, the authors note how textual analysis can unite the tribes of marketing. While most marketing problems are interdisciplinary, the field is often fragmented. By involving skills and ideas from each of the subareas of marketing, text analysis has the potential to help unite the field with a common set of tools and approaches.

Keywords

computational linguistics, machine learning, marketing insight, interdisciplinary, natural language processing, text analysis, text mining

Online supplement: https://doi.org/10.1177/0022242919873106

The digitization of information has made a wealth of textual data readily available. Consumers write online reviews, answer open-ended survey questions, and call customer service representatives (the content of which can be transcribed). Firms write ads, email frequently, publish annual reports, and issue press releases. Newspapers contain articles, movies have scripts, and songs have lyrics. By some estimates, 80%–95% of all business data is unstructured, and most of that unstructured data is text (Gandomi and Haider 2015).

Such data has the potential to shed light on consumer, firm, and market behavior, as well as society more generally. But, by itself, all this data is just that—data. For data to be useful, researchers must be able to extract underlying insight—to measure, track, understand, and interpret the causes and consequences of marketplace behavior.

This is where the value of automated textual analysis comes in. Automated textual analysis[1] is a computer-assisted methodology that allows researchers to rid themselves of measurement straitjackets, such as scales and scripted questions, and to quantify the information contained in textual data as it naturally occurs. Given these benefits, the question is no longer whether to use automated text analysis but *how* these tools can best be used to answer a range of interesting questions.

This article provides an overview of the use of automated text analysis for marketing insight. Methodologically, text analysis approaches can describe "what" is being said and "how" it is said, using both qualitative and quantitative inquiries with various degrees of human involvement. These

[1] Computer-aided approaches to text analysis in marketing research are generally interchangeably referred to as computer-aided text analysis (Pollach 2012), text mining (Netzer et al. 2012), automated text analysis (Humphreys and Wang 2017), or computer-aided content analysis (Dowling and Kabanoff 1996).

Jonah Berger is Associate Professor of Marketing, Wharton School, University of Pennsylvania, USA (email: jberger@wharton.upenn.edu). Ashlee Humphreys is Associate Professor, Medill School of Journalism, Media, and Integrated Marketing Communications, Northwestern University, USA (email: a-humphreys@northwestern.edu). Stephan Ludwig is Associate Professor of Marketing, University of Melbourne, Australia (email: stephan.ludwig@unimelb.edu.au). Wendy W. Moe is Associate Dean of Master's Programs, Dean's Professor of Marketing, and Co-Director of the Smith Analytics Consortium, University of Maryland, USA (email: wmoe@rhsmith.umd.edu). Oded Netzer is Professor of Business, Columbia Business School, Columbia University, USA (email: onetzer@gsb.columbia.edu). David A. Schweidel is Professor of Marketing, Goizueta Business School, Emory University, USA (email: dschweidel@emory.edu).

approaches consider individual words and expressions, their linguistic relationships within a document (within-text interdependencies) and across documents (across-text interdependencies), and the more general topics discussed in the text. Techniques range from computerized word counting and applying dictionaries to supervised or automated machine learning that helps deduce psychometric and substantive properties of text.

Within this emerging domain, we aim to make four main contributions. First, we illustrate how text data can be used for both prediction and understanding, to gain insight into who produced that text, as well as how that text may impact the people and organizations that consume it. Second, we provide a how-to guide for those new to text analysis, detailing the main tools, pitfalls, and challenges that researchers may encounter. Third, we offer a set of expansive research propositions pertaining to using text as a means to understand meaning making in markets with a focus on how customers, firms, and societies construe or comprehend marketplace interactions, relationships, and themselves. Whereas previous treatments of text analysis have looked specifically at consumer text (Humphreys and Wang 2017), social media communication (Kern et al. 2016), or psychological processes (Tausczik and Pennebaker 2010), we aim to provide a framework for incorporating text into marketing research at the individual, firm, market, and societal levels. By necessity, our approach includes a wide-ranging set of textual data sources (e.g., user-generated content, annual reports, cultural artifacts, government text).

Fourth, and most importantly, we discuss how text analysis can help "unite the tribes." As a field, part of marketing's value is its interdisciplinary nature. Unlike core disciplines such as psychology, sociology, or economics, the marketing discipline is a big tent that allows researchers from different traditions and research philosophies (e.g., quantitative modeling, consumer behavior, strategy, consumer culture theory) to come together to study related questions (Moorman et al. 2019a, b). In reality, however, the field often seems fragmented. Rather than different rowers all simultaneously pulling together, it often feels more like separate tribes, each independently going off in separate directions. Although everyone is theoretically working toward similar goals, there tends to be more communication within groups than between them. Different groups often speak different "languages" (e.g., psychology, sociology, anthropology, statistics, economics, organizational behavior) and use different tools, making it increasingly difficult to have a common conversation. However, text analysis can unite the tribes. Not only does it involve skills and ideas from each of these areas, doing it well *requires* such integration because it borrows ideas, concepts, approaches, and methods from each tribe and incorporates them to achieve insight. In so doing, the approach also adds value to each of the tribes in ways that might not otherwise be possible.

We start by discussing two distinctions that are useful when thinking about how text can be used: (1) whether text reflects or impacts (i.e., says something about the producer or has a downstream impact on something else) and (2) whether text is used for prediction or understanding (i.e., predicting something or understanding what caused something). Next, we explain how text may be used to unite the tribes of marketing. Then we provide an overview of text analysis tools and methodology and discuss key questions and measures of validity. Finally, we close with a future research agenda.

The Universe of Text

Communication is an integral part of marketing. Not only do firms communicate with customers, but customers communicate with firms and one another. Moreover, firms communicate with investors and society communicates ideas and values to the public (through newspapers and movies). These communications generate text or can be transcribed into text.

A simple way to organize the world of textual data is to think about producers and receivers—the person or organization that creates the text and the person or organization who consumes the text (Table 1). While there are certainly other parties that could be listed, some of the main producers and receivers are consumers, firms, investors, and society at large. Consumers write online reviews that are read by other consumers, firms create annual reports that are read by investors, and cultural producers represent societal meanings through the creation of books, movies, and other digital or physical artifacts that are consumed by individuals or organizations.

Consistent with this distinction between text producer and text receiver, researchers may choose to study how text reflects or impacts. Specifically, text reflects information about, and thus can be used to gain insight into, the text *producer* or one can study how text impacts the text *receiver*.

Text as a Reflection of the Producer

Text reflects and indicates something about the text producer (i.e., the person, organization, or context that created it). Customers, firms, and organizations use language to express themselves or achieve desired goals, and as a result, text signals information about the actors, organization, or society that created it and the contexts in which it was created. Like an anthropologist piecing together pottery shards to learn about a distant civilization, text provides a window into its producers.

Take, for example, a social media post in which someone talks about what they did that weekend. The text that person produces provides insight into several facets. First, it provides insight into the individual themselves. Are they introverted or extraverted? Neurotic or conscientious? It sheds light on who they are in general (i.e., stable traits or customer segments; Moon and Kamakura 2017) as well as how they may be feeling or what they may be thinking at the moment (i.e., states). In a sense, language can be viewed as a fingerprint or signature (Pennebaker 2011). Just like brush strokes or painting style can be used to determine who painted a particular painting, researchers use words and linguistic style to infer whether a play was written by Shakespeare, or if a person is depressed (Rude, Gortner, and Pennebaker 2004) or being deceitful

Table 1. Text Producers and Receivers.

Text Producers	Text Receivers: Consumers	Firms	Investors	Institutions/Society
Consumers	• Online reviews (Anderson and Simester 2014; Chen and Lurie 2013; Fazio and Rockledge 2015[a]; Kronrod and Danziger 2013[a]; Lee and Bradlow 2011; Liu, Lee, and Srinivasan 2019[a]; Melumad, Inman, and Pham 2019; Moon and Kamakura 2017; Puranam, Narayan, and Kadiyali 2017) • Social media (Hamilton, Schlosser, and Chen 2017[a]; Netzer et al. 2012; Villarroel Ordenes et al. 2017) • Offline word of mouth (Berger and Schwartz 2011[a]; Mehl and Pennebaker 2003[a])	• Forms and applications (Netzer, Lemaire, and Herzenstein 2019) • Idea-generation contexts (Bayus 2013[a]; Toubia and Netzer 2017) • Social media/brand communities (Herhausen et al. 2019) • Consumer complaints (Ma, Baohung, and Kekre 2015) • Customer language on service calls • Tweeting at companies (Liu, Singh, and Srinivasan 2016[a])	• Stock market reactions to consumer text (Bollen, Mao, and Zeng 2011; Tirunillai and Tellis 2012) • Protests • Petitions	• Crowdsourcing knowledge • Letters to the editor • Online comments section • Activism (e.g., organizing political movements and marches)
Firms	• Owned media (e.g., company website and social media; Villarroel Ordenes et al. 2018) • Advertisements (Fossen and Schweidel 2017[a], 2019; Liaukonyte, Teixeira, and Wilbur 2015[a]; Rosa et al. 1999; Stewart and Furse 1986) • Customer service agents (Packard and Berger 2019a; Packard, Moore, and McFerran 2018) • Packaging, including labels • Text used in instructions	• Trade publications (Weber, Heinze, and DeSoucey 2008[a]) • Interfirm communication emails (Ludwig et al. 2016) • White papers	• Financial reports (Loughran and McDonald 2016) • Corporate communications (Hobson, Mayhew, and Venkatachalam 2012) • Chief executive officer letters to shareholders (Yadav, Prabhu, and Chandy 2007	• Editorials by firm stakeholders • Interviews with business leaders
Investors		• Letters to shareholders (Yadav, Prabhu, and Chandy 2007) • Shareholder feedback (Wies et al. 2019)	• Sector reports	
Institutions/society	• News content (Berger, Kim, and Meyer 2019; Berger and Milkman 2012; Humphreys 2010) • Movies (Berger, Moe, and Schweidel 2019; Eliashberg, Hui, and Zhang 2007, 2014; Toubia et al. 2019) • Songs (Berger and Packard 2018; Packard and Berger 2019a) • Books (Akpinar and Berger 2015; Sorescu et al. 2018[a])	• Business section • Specialty magazines (e.g., *Wired*, *Harvard Business Review*)	• *Wall Street Journal* • *Fortune* • Various forms of investment advice that come from media	• Government documents, hearings, and memoranda (Chappell et al. 1997[a]) • Forms of public dialogue or debate

[a]Reference appears in the Web Appendix.

(Ludwig et al. 2016). The same is true for groups, organizations, or institutions. Language reflects something about who they are and thus provides insight into what they might do in the future.

Second, text can provide insight into a person's attitudes toward or relationships with other attitude objects—whether that person liked a movie or hated a hotel stay, for example, or whether they are friends or enemies with someone. Language used in loan applications provides insight into whether people will default (Netzer, Lemaire, and Herzenstein 2019), language used in reviews can provide insight into whether they are fake (Anderson and Simester 2014; Hancock et al. 2007;

Ott, Cardie, and Hancock 2012), and language used by political candidates could be used to study how they might govern in the future.

These same approaches can also be used to understand leaders, organizations, or cultural elites through the text they produce. For example, the words a leader uses reflect who they are as an individual, their leadership style, and their attitudes toward various stakeholders. The language used in ads, on websites, or by customer service agents reflects information about the company those pieces of text represent. Aspects such as brand personality (Opoku, Abratt, and Pitt 2006), how much a firm is thinking about its customers (Packard and Berger 2019a), or managers' orientation toward end users (Molner, Prabhu, and Yadav 2019) can be understood through text. Annual reports provide insight into how well a firm is likely to perform in the future (Loughran and McDonald 2016).

Yet beyond single individuals or organizations, text can also be aggregated across creators to study larger social groups or institutions. Given that texts reflect information about the people or organizations that created them, grouping people or organizations together on the basis of shared characteristics can provide insight into the nature of such groups and differences between them. Analyzing blog posts, for example, can shed light on how older and younger people view happiness differently (e.g., as excitement vs. peacefulness; Mogilner, Kamvar, and Aaker 2011). In a comparison of newspaper articles and press releases about different business sectors, text can be used to understand the creation and spread of globalization discourse (Fiss and Hirsch 2005). Customers' language use further gives insight into the consumer sentiment in online brand communities (Homburg, Ehm, and Artz 2015).

More broadly, because texts are shaped by the contexts (e.g., devices, cultures, time periods) in which they were produced, they also reflect information about these contexts. In the case of culture, U.S. culture values high-arousal positive affective states more than East Asian cultures (Tsai 2007), and these differences may show up in the language these different groups use. Similarly, whereas members of individualist cultures tend to use first-person pronouns (e.g., "I"), members of collectivist cultures tend to use a greater proportion of third-person pronouns (e.g., "we").

Across time, researchers were able to examine whether the national mood changed after the September 11 attacks by studying linguistic markers of psychological change in online diaries (Cohn, Mehl, and Pennebaker 2004). The language used in news articles, songs, and public discourse reflects societal attitudes and norms, and thus analyzing changes over time can provide insight into aspects such as attitudes toward women and minorities (Boghrati and Berger 2019; Garg et al. 2018) or certain industries (Humphreys 2010). Journal articles provide a window into the evolution of topics within academia (Hill and Carley 1999). Books and movies serve as similar cultural barometers and could be used to shed light on everything from cultural differences in customs to changes in values over time.

Consequently, text analysis can provide insights that may not be easily (or cost-effectively) obtainable through other methods. Companies and organizations can use social listening (e.g., online reviews and blog posts) to understand whether consumers like a new product, how customers feel about their brand, what attributes are relevant for decision making, or what other brands fall in the same consideration set (Lee and Bradlow 2011; Netzer et al. 2012). Regulatory agencies can determine adverse reactions to pharmaceutical drugs (Feldman et al. 2015; Netzer et al. 2012), public health officials can gauge how bad the flu will be this year and where it will hit the hardest (Alessa and Faezipour 2018), and investors can try to predict the performance of the stock market (Bollen, Mao, and Zeng 2011; Tirunillai and Tellis 2012).

Text's Impact on Receivers

In addition to reflecting information about the people, organizations, or society that created it, text also impacts or shapes the attitudes, behavior, and choices of the audience that consumes it. For example, take the language used by a customer service agent. While that language certainly reflects something about that agent (e.g., their personality, how they are feeling that day), how they feel toward the customer, and what type of brand they represent, that language also impacts the customer who receives it (Packard and Berger 2019a; Packard, Moore, and McFerran 2018). It can change customer attitudes toward the brand, influence future purchase, or affect whether customers talk about the interaction with their friends. In that sense, language has a meaningful and measurable impact on the world. It has consequences.

This can be seen in a myriad of different contexts. Ad copy shapes customers' purchase behavior (Stewart and Furse 1986), newspaper language changes customers' attitudes (Humphreys and LaTour 2013), trade publications and consumer magazines shift product category perceptions (e.g., Rosa et al. 1999), movie scripts shape audience reactions (Berger, Kim, and Meyer 2019; Eliashberg, Hui, and Zhang 2014; Reagan et al. 2016), and song lyrics shape song market success (Berger and Packard 2018; Packard and Berger 2019b). The language used in political debates shapes which topics get attention (Berman et al. 2019), the language used in conversation shapes interpersonal attitudes (Huang et al. 2017), and the language used in news articles shapes whether people read (Berger, Moe, and Schweidel, 2019b) or share (Berger and Milkman 2012) them.

Firms' language choice has impact as well. For example, nuances in language choices by firms when responding to customer criticism online directly impacts consumers and, thus, the firms' success in containing social media firestorms (Herhausen et al. 2019). Language used in YouTube ads is correlated with their virality (Tellis et al. 2019). Shareholder complaints about nonfinancial concerns and topics that receive high media attention substantially increase firms' advertising investments (Wies et al. 2019).

Note that while the distinction between text reflecting and impacting is a useful one, it is not an either/or. Text almost always simultaneously reflects *and* impacts. Text always reflects information about the actor or actors that created it, and as long as some audience consumes that text, it also impacts that audience.

Despite this relationship, researchers studying reflection versus impact tend to use text differently. Research that examines what the text reflects often treats it as a dependent variable and investigates how it relates to the text creator's personality, the social groups they belong to, or the time period or culture in which it was created.

Research that examines how text impacts others often treats it as an independent variable, examining if and how text shapes outcomes such as purchase, sharing, or engagement. In this framework, textual elements are linked with outcomes that are believed to be theoretical consequences of the textual components or some latent variable that they are thought to represent.

Contextual Influences on Text

Importantly, text is also shaped by contextual factors; thus, to better understand its meaning and impact, it is important to understand the broader situation in which it was produced. Context can affect content in three ways: through technical constraints and social norms of the genre, through shared knowledge specific to the speaker and receiver, and through prior history.

First, different types of texts are influenced by formal and informal rules and norms that shape the content and expectations about the message. For example, newspaper genres such as opinion pieces or feature stories will contain a less "objective" point of view than traditional reporting (Ljung 2000). Hotel comment cards and other feedback are usually dominated by more extreme opinions. On Snapchat and other social media platforms, messages are relatively recent, short, and often ephemeral. In contrast, online reviews can be longer and are often archived dating back several years. Synchronic text exchanges, in which two individuals interactively communicate in real time may be more informal and contain dialogue of short statements and phatic responses (i.e., communication such as "Hi," which serves a social function) that indicate affiliation rather than semantic content (Kulkarni 2014). Some genres (e.g., social media) are explicitly public, whereas on others, such as blogs, information that is more private may be conveyed.

Text is also shaped by technological constraints (e.g., the ability to like or share) and physical constraints (e.g., character length limitations). Tweets, for example, necessarily have 280 characters or fewer, which may shape the ways in which they are used to communicate. Mobile phones have constraints on typing and may shape the text that people produce on them (Melumad, Inman, and Pham 2019; Ransbotham, Lurie, and Liu 2019).

Second, the relationship between the text producer and consumer may affect what is said (or, more often, unsaid). If the producer and consumer know each other well, text may be relatively informal (Goffman 1959) and lack explicit information that a third party would need to make sense of the conversation (e.g., past events, known likes/dislikes). If both have an understanding of the goal of the communication (e.g., that the speaker wants to persuade the receiver), this may shape the content but be less explicit.

These factors are important to understand when interpreting the content of the text itself. Content has been shown to be shaped by the creator's intended audience (Vosoughi, Roy, and Aral 2018) and anticipated effects on the receiver (Barasch and Berger 2014). Similarly, what consumers share with their best friend may be different (e.g., less impacted by self-presentational motivations) than what they post online for everyone to see.[2] Firms' annual reports may be shaped by the goals of appearing favorably to the market. What people say on a customer service call may be driven by the goal of getting monetary compensation. Consumer protests online are meant to inspire change, not merely inform others.

Finally, history may affect the content of the text. In message boards, prior posts may shape future posts; if someone raised a point in a previous post, the respondent will most likely refer to the point in future posts. If retweets are included in an analysis, this will bias content toward most circulated posts. More broadly, media frames such as #metoo or #blacklivesmatter might make some concepts or facts more accessible to speakers and therefore more likely to emerge in text, even if seemingly unrelated (McCombs and Shaw 1972; Xiong, Cho, and Boatwright 2019).

Using Text for Prediction Versus Understanding

Beyond reflecting information about the text creator and shaping outcomes for the text recipient, another useful distinction is whether text is used for prediction or understanding.

Prediction

Some text research is predominantly interested in prediction. Which customer is most likely to default on their loan (Netzer, Lemaire, and Herzenstein 2019)? Which movie will sell the most tickets (Eliashberg et al. 2014)? How will the stock market perform (Bollen, Mao, and Zeng 2011; Tirunillai and Tellis 2012)? Whether focusing on individual-, firm-, or market-level outcomes, the goal is to predict with the highest degree of accuracy. Such work often takes many textual features and uses machine learning or other methods to combine these features in a way that achieves the best prediction. The authors care less

[2] Note that intermediaries can amplify (e.g., retweet) an original message and may have different motivations than the text producer.

about any individual feature and more about how the set of observable features can be combined to predict an outcome.

The main difficulty involved with using text for predictions is that text can generate hundreds and often thousands of features (words) that are all potential predictors for the outcome of interest. In some cases, the number of predictors is larger than the number of observations, making traditional statistical predictive models largely impractical. To address this issue, researchers often resort to machine learning–type methods, but overfitting needs to be carefully considered. In addition, inference with respect to the role of each word in the prediction can be difficult. Methods such as feature importance weighing can help extract some inference from these predictive models.

Understanding

Other research is predominantly interested in using text for understanding. How does the language consumers use shape word of mouth's impact (Packard and Berger 2017)? Why do some online posts get shared, songs become popular, or brands engender greater loyalty? How do cultural attitudes or business practices change? Whether focusing on individual-, firm-, or market-level outcomes, the goal is to understand why or how something occurred. Such work often involves examining only one or a small number of textual features or aspects that link to underlying psychological or sociological processes and aims to understand which features are driving outcomes and why.

One challenge with using textual data for understanding is drawing causal inferences from observational data. Consequently, work in this area may augment field data with experiments to allow key independent variables to be manipulated. Another challenge is interpreting relationships with textual features (we discuss this further in the closing section). Songs that use more second-person pronouns are more popular (Packard and Berger 2019b), for example, but this relationship alone does not necessarily explain *why* this is the case; second-person pronouns may indicate several things. Consequently, deeper theorizing, examination of links observed in prior research, or further empirical work is often needed.

Note that research that can use either a prediction or understanding lens to study either what text reflects or what it impacts. On the prediction side, researchers interested in what text reflects could use it to predict states or traits of the text creator such as customer satisfaction, likelihood of churn, or brand personality. Researchers interested in the impact of text could predict how text will shape outcomes such as reading behavior, sharing, or purchase among consumers of that text.

On the understanding side, someone interested in what text reflects could use it to shed light on why people might use certain personal pronouns when they are depressed or why customers might use certain types of emotional language when they are talking to customer service. Someone interested in the impact of text could use it to understand why text that evokes different emotions might be more likely to be read or shared.

Furthermore, while most research tends to focus on either prediction or understanding, some work integrates both aspects. Netzer, Lemaire, and Herzenstein (2019), for example, both use a range of available textual features to predict whether a given person will default on a loan and analyze the specific language used by people who tend to default (e.g., language used by liars).

Uniting the Tribes of Marketing

Regardless of whether the focus is on text reflection versus impact, or prediction versus understanding, doing text analysis well requires integrating skills, techniques, and substantive knowledge from different areas of marketing. Furthermore, textual analysis opens up a wealth of opportunity for each of these areas as well.

Take consumer behavior. While hypothetical scenarios can be useful, behavioral economics has recently gotten credit for many applications of social or cognitive psychology because these researchers have demonstrated phenomena in the field. Given concerns about replication, researchers have started to look for new tools that enable them to ensure validity and increase relevance to external audiences. Previously, use of secondary data was often limited because it addressed the "what" but not the "why" (i.e., what people bought or did, but not why they did so). But text can provide a window into the underlying process. Online reviews, for example, can be used to understand why someone bought one thing rather than another. Blog posts can help marketers understand consideration sets (Lee and Bradlow 2011; Netzer et al. 2012) and the customer journey (Li and Du 2011). Text even helps address the age-old issue of telling more than we can know (Nisbett and Wilson 1977). While people may not always know why they did something, their language often provides traces of explanation (Pennebaker 2011), even beyond what they can consciously articulate.

This richness is attractive to more than just behavioral researchers. Text opens a large-scale window into the world of "why" in the field and does so in a scalable manner. Quantitative modelers are always looking for new data sources and tools to explain and predict behavior. Unstructured data provides a rich set of predictors that are often readily available, at large scale, and able to be combined with structured measures as either dependent variables or independent variables. Text, through product reviews, user-driven social media activity, and firm-driven marketing efforts, provides data in real time that can shed light on consumer needs/preferences. This offers an alternative or supplement to traditional marketing research tools. In many cases, text can be retraced to an individual, allowing distinction between individual differences and dynamics. It also offers a playground where new methodologies from other disciplines can be applied (e.g., deep learning; LeCun, Bengio, and Hinton 2015; Liu et al. 2019).

Marketing strategy researchers want logic by which business can achieve its marketing objectives and to better understand what affects organizational success. A primary challenge to these researchers is to obtain reliable and generalizable survey or field data about factors that lie deep in the firm's culture and structure or that are housed in the mental models and beliefs of marketing leaders and employees. Text analysis offers an objective and systematic solution to assess constructs in naturally occurring data (e.g., letters to shareholders, press releases, patent text, marketing messages, conference calls with analysts) that may be more valid. Likewise, marketing strategy scholars often struggle with valid measures of a firm's marketing assets, and text may be a useful tool to understand the nature of customer, partner, and employee relationships and the strength of brand sentiments. For example, Kübler, Colicev, and Pauwels (2017) use dictionaries and support vector machine methods to extract sentiment and relate it to consumer mindset metrics.

Scholars who draw from anthropology and sociology have long examined text through qualitative interpretation and content analysis. Consumer culture theory–oriented marketing researchers are primarily interested in understanding underlying meanings, norms, and values of consumers, firms, and markets in the marketplace. Text analysis provides a tool for quantifying qualitative information to measure changes over time or make comparisons between groups. Sociological and anthropological researchers can use automated text analysis to identify important words, locate themes, link them to text segments, and examine common expressions in their context. For example, to understand consumer taste practices, Arsel and Bean (2013) use text analysis to first identify how consumers talk about different taste objects, doings, and meanings in their textual data set (comments on a website/blog) before analyzing the relationship between these elements using interview data.

For marketing practitioners, textual analysis unlocks the value of unstructured data and offers a hybrid between qualitative and quantitative marketing research. Like qualitative research, it is rich, exploratory, and can answer the "why," but like quantitative research, it benefits from scalability, which often permits modeling and statistical testing. Textual analysis enables researchers to explore open-ended questions for which they do not know the range of possible answers a priori. With text, scholars can answer questions that they did not ask or for which they did not know the right outcome measure. Rather than forcing on participants a certain scale or set of outcomes from which to select, for example, marketing researchers can instead ask participants broad questions, such as why they like or dislike something, and then use topic modeling tools such as latent Dirichlet allocation (LDA; explained in detail subsequently) to discover the key underlying themes.

Importantly, while text analysis offers opportunities for a variety of research traditions, such opportunities are more likely to be realized when researchers work across traditional subgroups. That is, the benefits of computer-aided text analysis are best realized if we include both quantitative, positivist analyses of content and qualitative, interpretive analyses of discourse. Quantitative researchers, for example, have the skills to build the right statistical models, but they can benefit from behavioral and qualitative researchers' ability to link words to underlying psychological or social processes as well as marketing strategy researchers' understanding of organizational and marketing activities driving firm performance. This is true across all of the groups.

Thus, to really extract insights from textual data, research teams must have the interpretative skills to understand the meaning of words, the behavioral skills to link them to underlying psychological processes, the quantitative skills to build the right statistical models, and the strategy skills to understand what these findings mean for firm actions and outcomes. We outline some potential areas for fruitful collaboration in "Future Research Agenda" section.

Text Analysis Tools, Methods, and Metrics

Given the recent work using text analysis to derive marketing insight, some researchers may wonder where to start. This section reviews methodologies often used in text-based research. These include techniques needed to convert text into constructs in the research process as well as procedures needed to incorporate extracted textual information into subsequent modeling and analyses. The objective of this section is not to provide a comprehensive tutorial but, rather, to expose the reader to available techniques, discuss when different methods are appropriate, and highlight some of the key considerations in applying each method.

The process of text analysis involves several steps: (1) data preprocessing, (2) performing a text analysis of the resulting data, (3) converting the text into quantifiable measures, and (4) assessing the validity of the extracted text and measures. Each of these steps may vary depending on the research objective. Table 2 provides a summary of the different steps involved in the text analysis process from preprocessing to commonly used tools and measures and validation approaches. Table 2 can serve as a starter kit for those taking their first steps with text analysis.

Data Preprocessing

Text is often unstructured and "messy," so before any formal analyses can take place, researchers must first preprocess the text itself. This step provides structure and consistency so that the text can be used systematically in the scientific process. Common software tools for text analysis include Python (https://www.nltk.org/) and R (https://cran.r-project.org/web/packages/quanteda/quanteda.pdf, https://quanteda.io/). For both software platforms, a set of relatively easy-to-use tools has been developed to perform most of the data preprocessing steps. Some programs, such as Linguistic Inquiry and Word Count (LIWC; Tausczik and Pennebaker 2010) and WordStat (Peladeau 2016), require minimal preprocessing. We detail the data preprocessing steps next (for a summary of the steps, see Table 3).

Table 2. The Text Analysis Workflow.

Data Preprocessing	Common Tools	Measurement	Validity
• ***Data acquisition***: Obtain or download (often in an HTML format) text. • ***Tokenization***: Break text into units (often words and sentences) using delimiters (e.g., periods). • ***Cleaning***: Remove nonmeaningful text (e.g., HTML tags) and nontextual information. • ***Removing stop words***: Eliminate common words such as "a" or "the" that appear in most documents. • ***Spelling***: Correct spelling mistakes using common spellers. • ***Stemming and lemmatization***: Reduce words into their common stem or lemma.	• ***Entity extraction***: Tools used to extract the meaning of one word at a time or simple cooccurrence of words. These tools include dictionaries; part-of-speech classifiers; many sentiment analysis tools; and, for complex entities, machine learning tools. • ***Topic modeling***: Topic modeling can identify the general topics (described as a combination of words) that are discussed in a body of text. Common tools include LDA and PF. • ***Relation extraction***: Going beyond entity extraction, the researcher may be interested in identifying textual relationships among extracted entities. Relation extraction often requires the use of supervised machine learning approaches.	• ***Count measures***: The set of measures used to represent the text as count measures. The tf-idf measure allows the researcher to control for the popularity of the word and the length of the document. • ***Similarity measures***: Cosine similarity and the Jaccard index are often used to measure the similarity of the text between documents. • ***Accuracy measures***: Often used relative to human-coded or externally validated documents. The measures of recall, precision, F1, and the area under the curve of the receiver operating characteristic curve are often used. • ***Readability measures***: Measures such as the simple measure of gobbledygook (SMOG) are used to assess the readability level of the text.	• ***Internal Validity*** – Construct: Dictionary validation and sampling-and-saturation procedures ensure that constructs are correctly operationalized in text. – Concurrent: Compare operationalizations with prior literature. – Convergent: Multiple operationalizations of key constructs. – Causal: Control for factors related to alternative hypotheses. • ***External Validity*** – Predictive: Use conclusions to predict key outcome variable (e.g., sales, stock price). – Generalizability: Replicate effects in other domains. – Robustness: Test conclusions on holdout samples (k-fold); compare different categories within the data set.

Note: PF = Poisson factoring.

Data acquisition. Data acquisition can be well defined if the researcher is provided with a set of documents (e.g., emails, quarterly reports, a data set of product reviews) or more open-ended if the researcher is using a web scraper (e.g., Beautiful Soup) that searches the web for instances of a particular topic or a specific product. When scraping text from public sources, researchers should abide by the legal guidelines for using the data for academic or commercial purposes.

Tokenization. Tokenization is the process of breaking the text into units (often words and sentences). When tokenizing, the researcher needs to determine the delimiters that define a token (space, period, semicolon, etc.). If, for example, a space or a period is used to determine a word, it may produce some nonsensical tokens. For example, "the U.S." may be broken to the tokens "the," "U," and "S." Most text-mining software has smart tokenization procedures to alleviate such common problems, but the researcher should pay close attention to instances that are specific to the textual corpora. For cases that include paragraphs or threads, depending on the research objective, the researcher may wish to tokenize these larger units of text as well.

Cleaning. HTML tags and nontextual information, such as images, are cleaned or removed from the data set. The cleaning needs may depend on the format in which the data was provided/extracted. Data extracted from the web often requires heavier cleaning due to the presence of HTML tags. Depending on the purpose of the analysis, images and other nontextual information may be retained. Contractions such as "isn't" and "can't" need to be expanded at this step. In this step, researchers should also be mindful of and remove phrases automatically generated by computers that may occur within the text (e.g., "html").

Removing stop words. Stop words are common words such as "a" and "the" that appear in most documents but often provide no significant meaning. Common text-mining tools (e.g., the tm, quanteda, tidytext, and tokenizers package in R; the Natural Language Toolkit package in Python; exclusion words in WordStat) have a predefined list of such stop words that can be amended by the researcher. It is advisable to add common words that are specific to the domain (e.g., "Amazon" in a corpora of Amazon reviews) to this list. Depending on the research objective, stop words can sometimes be very meaningful, and researchers may wish to retain them for their analysis. For example, if the researcher is interested in extracting not only the content of the text but also writing style (e.g., Packard, Moore, and McFerran 2018), stop words can be very informative (Pennebaker 2011).

Spelling. Most text-mining packages have prepackaged spellers that can help correct spelling mistakes (e.g., the Enchant speller). In using these spellers, the researcher should be aware of language that is specific to the domain and may not appear in the speller—or even worse, that the speller may incorrectly "fix." Moreover, for some analyses the researcher may want to record the number of spelling mistakes as an additional

Table 3. Data Preprocessing Steps.

Data Processing Step	Issues to Consider	Illustration
Data acquisition	• Is the data readily available in textual format or does the research needs to use a web scraper to find the data? • What are the legal guidelines for using the data (particularly relevant for web-scraped data)?	Tweets mentioning different brands from the same category during a particular time frame are downloaded from Twitter.
Tokenization	• What is the unit of analysis (word, sentence, thread, paragraph)? • Use smart tokenization for delimiters and adjust to specific unique delimiters found in the corpora.	The unit of analysis is the individual tweet. The words in the tweet are the tokens of the document.
Cleaning	• Web-scraped data often requires cleaning of HTML tags and other symbols. • Depending on the research objective, certain textual features (e.g., advertising on the page) may or may not be cleaned. • Expansion of contractions such as “isn’t” to “is not.”	URLs are removed and emojis/emoticons are converted to words.
Removing stop word	• Use a stop word list available by the text-mining software, but adapt it to a specific application by adding/removing relevant stop words. • If the goal of the analysis is to extract writing style, it is advisable to keep all/some of the stop words.	Common words are removed. The remaining text contains brand names, nouns, verbs, adjectives, and adverbs.
Spelling	• Can use commonly used spellers in text-mining packages (e.g., the Enchant speller). • Language that is specific to the domain may be erroneously coded as a spelling mistake. • May wish to record the number of spelling mistakes as an additional textual measure.	Spelling mistakes are removed, enabling analysis into consumer perceptions (manifest through word choice) of different brands.
Stemming and lemmatization	• Can use commonly used stemmers in text-mining packages (e.g., Porter stemmer). • If the goal of the analysis is to extract writing style, stemming can mask the tense used.	Verbs and nouns are “standardized” by reducing to their stem or lemma.

textual measure reflecting important states or traits of the communicator (e.g., Netzer, Lemaire, and Herzenstein 2019).

Stemming and lemmatization. Stemming is the process of reducing the words into their word stem. Lemmatization is similar to stemming, but it returns the proper lemma as opposed to the word’s root, which may not be a meaningful word. For example, with stemming, the entities “car” and “cars” are stemmed to “car,” but “automobile” is not. In lemmatization, the words “car,” “cars,” and “automobile” are all reduced to the lemma “automobile.” Several prepackaged stemmers exist in most text-mining tools (e.g., the Porter stemmer). Similar to stop words, if the goal of the analysis is to extract the writing style, one may wish to skip the stemming step, because stemming often masks the tense used.

Text Analysis Extraction

Once the data has been preprocessed, the researcher can start analyzing the data. One can distinguish between the extraction of individual words or phrases (entity extraction), the extraction of themes or topics from the collective set of words or phrases in the text (topic extraction), and the extraction of relationships between words or phrases (relation extraction). Table 4 highlights these three types of analysis, the typical research questions investigated with each approach, and some commonly used tools.

Entity (word) extraction. At the most basic level, text mining has been used in marketing to extract individual entities (i.e., count words) such as person, location, brands, product attributes, emotions, and adjectives. Entity extraction is probably the most commonly used text analysis approach in marketing academia and practice, partly due to its relative simplicity. It allows the researcher to explore both what was written (the content of the words) as well as how it was written (the writing style). Entity extraction can be used (1) to monitor discussions on social media (e.g., numerous commercial companies offer buzz monitoring services and use entity extraction to track how frequently a brand is being mentioned across alternative social media), (2) to generate a rich set of entities (words) to be used in a predictive model (e.g., which words or entities are associated with fake or fraudulent statements), and (3) as input to be used with dictionaries to extract more complex forms of textual expressions, such as a particular concept, sentiment, emotion, or writing style.

In addition to programming languages such as Python and R’s tm tool kits, software packages such as WordStat make it possible to extract entities without coding. Entity extraction can also serve as input in commonly used dictionaries or lexicons. Dictionaries (i.e., a predefined list of words, such as a list of brand names) are often used to classify entities into the categories (e.g., concepts, brands, people, categories, locations). In more formal text, capitalization can be used to help

Table 4. Taxonomy of Text Analysis Tools.

Approach	Common Tools	Research Questions	Benefits	Limitations and Complexities	Marketing Examples
Entity (word) extraction: Extracting and identifying a single word/n-gram	• Named entity extraction (NER) tools (e.g., Stanford NER) • Dictionaries and lexicons (e.g., LIWC, EL 2.0, SentiStrength, VADER) • Rule-based classification • Linguistic-based NLP tools • Machine learning classification tools (conditional random fields, hidden Markov models, deep learning)	• Brand buzz monitoring • Predictive models where text is an input • Extracting psychological states and traits • Sentiment analysis • Consumer and market trends • Product recommendations	• Can extract a large number of entities • Can uncover known entities (people, brands, locations) • Can be combined with dictionaries to extract sentiment or linguistic styles • Relatively simple to use	• Can be unwieldy due to the large number of entities extracted • Some entities have multiple meanings that are difficult to extract (e.g., the laundry detergent brand "All") • Slang and abbreviations make entity extraction more difficult in social media • Machine learning tools may require large human-coded training data • Can be limited for sentiment analysis	• Lee and Bradlow (2011) • Berger and Milkman (2012) • Ghose et al. (2012)[a] • Tirunillai and Tellis (2012) • Humphreys and Thompson (2014)[a] • Berger, Moe, and Schweidel (2019) • Packard, Moore, and McFerran (2018)
Topic extraction: Extracting the topic discussed in the text	• LSA • LDA • PF • LDA2vec word embedding	• Summarizing the discussion • Identifying consumer and market trends • Identifying customer needs	• Topics often provide useful summarization of the data • Data reduction permits the use of traditional statistical methods in subsequent analysis • Easy-to-assess dynamics	• The interpretation of the topics can be challenging • No clear guidance on the selection of the number of topics • Can be difficult with short text (e.g., tweets)	• Tirunillai and Tellis (2014) • Büschken and Allenby (2016) • Puranam, Narayan, and Kadiyali (2017) • Berger and Packard (2018) • Liu and Toubia (2018) • Toubia et al. (2019) • Zhong and Schweidel (2019) • Ansari, Li, and Yang (2018)[a] • Timoshenko and Hauser (2019) • Liu, Singh, and Srinivasan (2016)[a] • Liu, Lee, and Srinivasan (2019)[a]
Relation extraction: Extracting and identifying relationships among words	• Co-occurrence of entities • Handwritten rule • Supervised machine learning • Deep learning • Word2vec word embedding • Stanford Sentence and Grammatical Dependency Parser	• Market mapping • Identifying problems mentioned with specific product features • Identifying sentiment for a focal entity • Identifying which product attributes are mentioned positively/negatively • Identifying events and consequences (e.g., crisis) from consumer- or firm-generated text • Managing service relationships	• Relaxes the bag-of-words assumption of most text-mining methods • Relates the text to a particular focal entity • Advances in text-mining methods will offer new opportunities in marketing	• Accuracy of current approaches is limited • Complex relationships may be difficult to extract • It is advised to develop domain-specific sentiment tools as sentiment signals can vary from one domain to another	• Netzer et al. (2012) • Toubia and Netzer (2017) • Boghrati and Berger (2019)

[a]Reference appears in the Web Appendix.

extract known entities such as brands. However, in more casual text, such as social media, such signals are less useful. Common dictionaries include LIWC (Pennebaker et al. 2015), EL 2.0 (Rocklage, Rucker, and Nordgren 2018), Diction 5.0, or General Inquirer for psychological states and traits (for example applications, see Berger and Milkman [2012]; Ludwig et al. [2013]; Netzer, Lemaire, and Herzenstein [2019]).

Sentiment dictionaries such as Hedonometer (Dodds et al. 2011), VADER (Hutto and Gilbert 2014), and LIWC can be used to extract the sentiment of the text. One of the major limitations of the lexical approaches for sentiment analysis commonly used in marketing is that they apply a "bag of words" approach—meaning that word order does not matter—and rely solely on the cooccurrence of a word of interest (e.g., "brand") with positive or negative words (e.g., "great," "bad") in the same textual unit (e.g., a review). While dictionary approaches may be an easy way to measure constructs and comparability across data sets, machine learning approaches trained by human-coded data (e.g., Borah and Tellis 2016; Hartmann et al. 2018; Hennig-Thurau, Wiertz, and Feldhaus 2015) tend to be the most accurate way of measuring such constructs (Hartmann et al. 2019), particularly if the construct is complex or the domain is uncommon. For this reason, researchers should carefully weigh the trade-off between empirical fit and theoretical commensurability, taking care to validate any dictionaries used in the analysis (discussed in the next section).

A specific type of entity extraction includes linguistic-type entities such as part-of-speech tagging, which assigns a linguistic tag (e.g., verb, noun, adjective) to each entity. Most text analysis tools (e.g., the tm package in R, the Natural Language Toolkit package in Python) have a built-in part-of-speech tagging tool. If no predefined dictionary exists, or the dictionary is not sufficient for the extraction needed, one could add handcrafted rules to help define entities. However, the list of rules can become long, and the task of identifying and writing the rules can be tedious. If the entity extraction by dictionaries or rules is difficult or if the entities are less defined, machine learning–supervised classification approaches (e.g., conditional random fields [Netzer et al. 2012], hidden Markov models) or deep learning (Timoshenko and Hauser 2019) can be used to extract entities. The limitation of this approach is that often a relatively large hand-coded training data set needs to be generated.

To allow for a combination of words, entities can be defined as a set of consecutive words, often referred to as n-grams, without attempting to extract the relationship between these entities (e.g., the consecutive words "credit card" can create the unigram entities "credit" and "card" as well as the bigram "credit card"). This can be useful if the researcher is interested in using the text as input for a predictive model.

If the researcher wishes to extract entities while understanding the context in which the entities were mentioned in the text (thus avoiding the limitation of the bag-of-words approach), the emerging set of tools of word2vec or word embedding (Mikolov et al. 2013) can be employed. Word2vec maps each word or entity to a vector of latent dimensions called embedding vector based on the words with which each focal word appears. This approach allows the researcher not only to extract words but also to understand the similarity between words based on the similarities between the embedding vectors (or the similarities between the sentences in which each word appears). Thus, unlike the previous approaches discussed thus far, word2vec preserves the context in which the word appeared. While word embedding statistically captures the context in which a word appears, it does not directly linguistically "understand" the relationships among words.

Topic modeling. Entity extraction has two major limitations: (1) the dimensionality of the problem (often thousands of unique entities are extracted) and (2) the interpretation of many entities. Several topic modeling approaches have been suggested to overcome these limitations. Similar to how factor analysis identifies underlying themes among different survey items, topic modeling can identify the general topics (described as a combination of words) that are discussed in a body of text. This text summarization approach increases understanding of document content and is particularly useful when the objective is insight generation and interpretation rather than prediction (e.g., Berger and Packard 2018; Tirunillai and Tellis 2014). In addition, monitoring topics, as opposed to words, makes it easier to assess how discussion changes over time (e.g., Zhong and Schweidel 2019).

Methodologically, topic modeling mimics the data-generating process in which the writer chooses the topic she wants to write about and then chooses the words to express these topics. Topics are defined as word distributions that commonly co-occur and thus have a certain probability of appearing in a topic. A document is then described as a probabilistic mixture of topics.

The two most commonly used tools for topic modeling are LDA (Blei, Ng, and Jordan 2003) and Poisson factorization (PF; Gopalan, Hofman and Blei 2013). The predominant approach prior to LDA and PF was the support-vector-machine latent semantic analysis (LSA) approach. While LSA is simpler and faster to implement than LDA and PF, it requires larger textual corpora and often achieves lower accuracy levels. Other approaches include building an ontology of topics using a combination of human classification of documents as seeding for a machine learning classification (e.g., Moon and Kamakura 2017). Whereas LDA is often simpler to apply than PF, PF has the advantage of not assuming that the topic probabilities must sum to one. That is, some documents may have more topic presences than others, and a document can have multiple topics with high likelihood of occurrence. In addition, PF tends to be more stable with shorter text. Büschken and Allenby (2016) relax the common bag-of-words assumption underlying the traditional LDA model and leverage the within-sentence dependencies of online reviews. LDA2vec is another approach to assess topics while accounting for the sequence context in which the word appears (Moody 2016). In the context of search queries, Liu and Toubia (2018) further extend the LDA approach to hierarchical LDA for cases in which related documents (queries and search results) are used to extract the topics. Furthermore, the researcher can use an unsupervised or seeded LDA approach to incorporate prior knowledge in the construction and interpretation of the topics (e.g., Puranam, Narayan, and Kadiyali 2017; Toubia et al. 2019).

While topic modeling methods often produce very sensible topics, because topics are selected solely based on a statistical approach, the selection of the number of topics and the interpretation of some topics can be challenging. It is recommended to combine statistical approaches (e.g., the perplexity measure, which is a model fit–based measure) and researcher judgment when selecting the number of topics.

Relation extraction. At the most basic level, relationships between entities can be captured by the mere co-occurrence of entities (e.g., Boghrati and Berger 2019; Netzer et al. 2012; Toubia and Netzer 2017). However, marketing researchers are often more interested in identifying textual relationships among extracted entities, such as the relationships between products, attributes, and sentiments. Such relationships are often more relevant for the firm than merely measuring the volume of brand mentions or even the overall brand sentiment. For example, researchers may want to identify whether consumers mentioned a particular problem with a specific product feature. Feldman et al. (2015) and Netzer et al. (2012) provide such examples by identifying the textual relationships between drugs and adverse drug reactions that imply that a certain drug may cause a particular adverse reaction.

Relation extraction also offers a more advanced route to capture sentiment by providing the link between an entity of interest (e.g., a brand) and the sentiment expressed, beyond their mere cooccurrence. Relation extraction based on the bag-of-words approach, which treats the sentence as a bag of unsorted words and searches for word cooccurrence, is limited because the cooccurrence of words may not imply a relationship. For example, the cooccurrence of a drug (e.g., Advil) with a symptom (e.g., headache) may refer to the symptom as a side effect of the drug or as the effect the drug is aiming to alleviate. Addressing such relationships requires identifying the sequence of words and the linguistic relationship among them. There have been only limited applications of such relation extraction in marketing, primarily due to the computational and linguistic complexities involved in accurately making such relational inferences from unstructured data (see, e.g., the diabetes drugs application in Netzer et al. [2012]). However, as the methodologies used to extract entity relations evolve, we expect this to be a promising direction for marketers to take.

The most commonly used approaches for relation extraction are handwritten relationship rules, supervised machine learning approaches, and a combination of these approaches. At the most basic level, the researcher could write a set of rules that describe the required relationship. An example of such a rule may be the co-occurrence of product (e.g., "Ford"), attribute (e.g., "oil consumption"), and problem (e.g., "excessive"). However, such approaches tend to require many handwritten rules and have low recall (they miss many relations) and thus are becoming less popular.

A more common approach is to train a supervised machine learning tool. This could be linguistic agnostic approaches (e.g., deep learning) or natural language processing (NLP) approaches that aim to understand the linguistic relationship in the sentence. Such an approach requires a relatively large training data set provided by human coders in which various relationships (e.g., sentiment) are observed. One readily available tool for NLP-based relationship extraction is the Stanford Sentence and Grammatical Dependency Parser (http://nlp.stanford.edu:8080/parser/). The tool identifies the grammatical role of different words in the sentence to identify their relationship. For example, to assign a sentiment to a particular attribute, the parser first identifies the presence of an emotion word and then, in cases where a subject is present, automatically assesses if there is a grammatical relationship (e.g., in the sentence "the hotel was very nice," the adjective "nice" relates to the subject "hotel"). As with many off-the-shelf tools, the validity of the tool for a specific relation extraction needs to be tested.

Finally, beyond the relations between words/entities within one document, text can also be investigated across documents (e.g., online reviews, academic articles). For example, a temporal sequence of documents or a portfolio of documents across a group or community of communicators can be examined for interdependencies (Ludwig et al. 2013, 2014).

Text Analysis Metrics

Early work in marketing has tended to summarize unstructured text with structured proxies for this data. For example, in online reviews, researchers have used volume (e.g., Godes and Mayzlin 2004; Moe and Trusov 2011); valence, often captured by numeric ratings that supplement the text (e.g., Godes and Silva 2012; Moe and Schweidel 2012; Ying, Feinberg and Wedel 2006); and variance, often captured using entropy-type measures (e.g., Godes and Mayzlin 2004). However, these quantifiable metrics often mask the richness of the text. Several common metrics are often used to quantify the text itself, as we explain next.

Count measures. Count measures have been used to measure the frequency of each entity's occurrence, entities' co-occurrence, or entities' relations. For example, when using dictionaries to evaluate sentiment or other categories, researchers often use the proportion of negative and/or positive words in the document, or the difference between the two (Berger and Milkman 2012; Borah and Tellis 2016; Pennebaker et al. 2015; Schweidel and Moe 2014; Tirunillai and Tellis 2014). The problem with simple counts is that longer documents are likely to include more occurrences of every entity. For that reason, researchers often focus on the proportions of words in the document that belong to a particular category (e.g., positive sentiment). The limitation of this simple measure is that some words are more likely to appear than others. For example, the word "laptop" is likely to appear in almost every review in corpora that is composed of laptop reviews.

Accuracy measures. When evaluating the accuracy of text measures relative to human-coded or externally validated documents, measures of recall and precision are often used. Recall is the proportion of entities in the original text that the text-mining algorithm was able to successfully identify (it is defined by the ratio of true positives to the sum of true positives and false negatives). Precision is the proportion of correctly identified entities from all entities identified (it is defined by the ratio of true positives to the sum of true positives and false positives). On their own, recall and precision measures are difficult to assess because an improvement in one often comes at the expense of the other. For example, if one defines that every entity in the corpora is a brand, recall for brands will be perfect (you will never miss a brand if it exists in the text), but precision will be very low (there will be many false positive identifications of a brand entity).

To create the balance between recall and precision, one can use the F1 measure—a harmonic mean of the levels of recall and precision. If the researcher is more concerned with false positives than false negatives (e.g., it is more important to identify positives than negatives), recall and precision can be weighted differently. Alternatively, for unbalanced data with high proportions of true or false in the populations, a receiver operating characteristics curve can be used to reflect the relationship between true positives and false positives, and the area under the curve is often used as a measure of accuracy.

Similarity measures. In some cases, the researcher is interested in measuring the similarity between documents (e.g., Ludwig et al. 2013). How similar is the language used in two advertisements? How different is a song from its genre? In such cases, measures such as linguistic style matching, similarity in topic use (Berger and Packard 2018), cosine similarity, and the Jaccard index (e.g., Toubia and Netzer 2017) can be used to assess the similarity between the text of two documents.

Readability measures. In some cases, the researcher is interested in evaluating the readability of the text. Readability can reflect the sophistication of the writer and/or the ability of the reader to comprehend the text (e.g., Ghose and Ipeirotis 2011). Common readability measures include the Flesch–Kincaid reading ease and the simple measure of gobbledygook (SMOG) measures. These measures often use metrics such as average number of syllables and average number of words per sentence to evaluate the readability of the text. Readability measures often grade the text on a 1–12 scale reflecting the U.S. school grade-level needed to comprehend the text. Common text-mining packages have built-in readability tools.

The Validity of Text-Based Constructs

While the availability of text has opened up a range of research questions, for textual data to provide value, one must be able to establish its validity. Both internal validity (i.e., does text accurately measure the constructs and the relationship between them?) and external validity (i.e., do the test-based findings apply to phenomena outside the study?) can be established in various ways (Humphreys and Wang 2017). Table 5 describes how the text analysis can be evaluated to improve different types of validity (Cook and Campbell 1979).

Internal Validity

Internal validity is often a major threat in the context of text analysis because the mapping between words and the underlying dimension the research aims to measure (e.g., psychological state and traits) is rarely straightforward and can vary across contexts and textual outlets (e.g., formal news vs. social media). In addition, given the relatively young field of automated text analysis, validation of many of the methods and constructs is still ongoing.

Accordingly, it is important to confirm the internal validity of the approach used. A range of methods can be adopted to ensure construct, concurrent, convergent, discriminant, and causal validity. In general, the approach for ensuring internal validity is to ensure that the text studied accurately reflects the theoretical concept or topic being studied, does so in a way that is congruent with prior literature, is discriminant from other related constructs, and provides ample and careful evidence for the claims of the research.

Construct validity. Construct validity (i.e., does the text represent the theoretical concept?) is perhaps the most important to address when studying text. Threats to construct validity occur when the text provides improper or misleading evidence of the construct. For instance, researchers often rely on existing standardized dictionaries to extract constructs to ensure that their work is comparable with other work. However, these dictionaries may not always fit the particular context. For example, extracting sentiment from financial reports using sentiment tools developed for day-to-day language may not be appropriate. Particularly when attempting to extract complex constructs (e.g., psychological states and traits, relationships between consumers and products, and even sentiment), researchers should attempt to validate the constructs on the specific application to ensure that what is being extracted from the text is indeed what they intended to extract. Construct validity can also be challenged when homonyms or other words do not accurately reflect what researchers think they do.

Strategies for addressing threats to construct validity require that researchers examine how the instances counted in the data connect to the theoretical concept(s) (Humphreys and Wang 2017). Dictionaries can also be validated using a saturation approach, pulling a subsample of coded entries and verifying with a hit rate of approximately 80% (Weber 2005). Another method is to use input from human coders, as is done to support machine learning applications (as previously discussed). For example, one can use Amazon Mechanical Turk workers to label phrases on a scale from "very negative" to "very positive" for sentiment analysis and then use these words to create a

Table 5. Text Analysis Validation Techniques.

Type of Validity	Validation Technique	Description of Method for Validation	References
Internal Validity			
Construct validity	Dictionary validation	After draft dictionary is created, pull 10% of the sample and calculate the hit rate. Measures such as hit rates, precision, and recall can be used to measure accuracy.	Weber (2005)
		Have survey participants rate words included in the dictionary. Based on this data, the dictionary can also be weighted to reflect the survey data.	Brysbaert, Warriner, and Kuperman (2014)[a]
		Have three coders evaluate the dictionary categories. If two of the three coders agree that the word is part of the category, include; if not, exclude. Calculate overall agreement.	Humphreys (2010); Pennebaker, Francis, and Booth (2001)[a]
	Saturation	Pull 10% of instances coded from the data and calculate the hit rate. Adjust word list until saturation reaches 80% hit rate.	Weber (2005)
Concurrent validity	Multiple dictionaries	Calculate and compare multiple textual measures of the same construct (e.g., multiple sentiment measures)	Hartmann et al. (2018)
	Comparison of topics	Compare with other topic models of similar data sets in other research (e.g., hotel reviews)	Mankad et al. (2016)[a]
Convergent validity	Triangulation	Look within text data for converging patterns (e.g., positive/e emotion correlates with known-positive attributes); apply Principle Components Analysis to show convergent groupings of words	Humphreys (2010); Kern et al. (2016)
	Multiple operationalizations	Operationalize constructs with textual and nontextual data (e.g., sentiment, star rating)	Ghose et al. (2012)[a]; Mudambi, Schuff, and Zhang (2014)[a]
Causal validity	Control variables	Include variables in the model that address rival hypotheses to control for these effects	Ludwig et al. (2013)
	Laboratory study	Replicate focal relationship between the independent variable and dependent variable in a laboratory setting	Spiller and Belogolova (2016)[a]; Van Laer et al. (2018)
External Validity			
Generalizability	Replication with different data sets	Compare the results from the text analysis with the results obtained other (possibly non-text-related) data sets	Netzer et al. (2012)
	Predict key performance measure	Include results from text analysis in regression or other model to predict a key outcome (e.g., sales, engagement)	Fossen and Schweidel (2019)
Predictive validity	Holdout sample	Train model on approximately 80%–90% of the data and validate the model with the remaining data. Validation can be done using k-fold validation, which trains the mode on k-1 subsets of the data and predicts for the remaining subset of testing.	Jurafsky et al. (2014)
Robustness	Different statistical measures, unitizations	Use different, but comparable, statistical measures or algorithms (e.g., lift, cosine similarity, Jaccard similarity), aggregate at different levels (e.g., day, month)	Netzer et al. (2012)

[a]Reference appears in the Web Appendix.

weighted dictionary. In many cases, multiple methods for dictionary validation are advisable to ensure that one is achieving both theoretical and empirical fit. For topic modeling, researchers infer topics from a list of cooccurring words. However, these are theoretical inferences made by researchers. As such, construct validity is equally important and can be ascertained using some of the same methods of validation, through saturation and calculating a hit rate through manual analysis of a subset of the data. When using a classification approach, confusion matrices can be produced to provide details on accuracy, false positives, and false negatives (Das and Chen 2007).

Concurrent validity. Concurrent validity concerns the way that the researcher's operationalization of the construct relates to prior operationalizations. Threats to concurrent validity often come when researchers create text-based measures inductively from the text. For instance, if one develops a topic model from the text, it will be based on the data set and may not therefore produce topics that are comparable with previous research. To address these threats, one should compare the operationalization with other research and other data sources. For example, Schweidel and Moe (2014) propose a measure of brand sentiment based on social media text data and validate it by comparing it with brand measures obtained through a traditional marketing research survey. Similarly, Netzer et al. (2012) compare the market structure maps derived from textual information with those derived from product switching and surveys, and Tirunillai and Tellis (2014) compare the topics they identify with those found in *Consumer Reports*. When studying linguistic style (Pennebaker and King 1999), for example, it is beneficial to use robust measures from prior literature where factor analysis and other methods have already been employed to create the construct.

Convergent validity. Convergent validity ensures that multiple measurements of the construct (i.e., words) all converge to the same concept. Convergent validity can be threatened when the measures of the construct do not align or have different effects. Convergent validity can be enhanced by using several substantively different measures (e.g., dictionaries) of the same construct to look for converging patterns. For example, when studying posts about the stock market, Das and Chen (2007) compare five classifiers for measuring sentiment, comparing them in a confusion matrix to examine false positives. Convergent evidence can also come from creating a correlation or similarity matrix of words or concepts and checking for patterns that have face validity. For instance, Humphreys (2010) looks for patterns between the concept of crime and negative sentiment to provide convergent evidence that crime is negatively valenced in the data.

Discriminant validity. Discriminant validity, the degree to which the construct measures are sufficiently different from measures of other constructs, can be threatened when the measurement of the construct is very similar to that of another construct. For instance, measurements of sentiment and emotion in many cases may not seem different because they are measured using similar word lists or, when using classification, return the same group of words as predictors. Strategies for ensuring discriminant validity entail looking for discriminant rather than convergent patterns and boundary conditions (i.e., when and how is sentiment different from emotion?). Furthermore, theoretical refinements can be helpful in drawing finer distinctions. For example, anxiety, anger, and sadness are different kinds of emotion (and can be measured via psychometrically different scales), whereas sentiment is usually measured as positive, negative, or neutral (Pennebaker et al. 2015).

Causal validity. Causal validity is the degree to which the construct, as operationalized in the data set, is actually the cause of another construct or outcome, and it is best ascertained through random assignment in controlled lab conditions. Any number of external factors can threaten causal validity. However, steps can be taken to enhance causal validity in naturally occurring textual data. In particular, rival hypotheses and other explanatory factors for the proposed causal relationship can be statistically controlled for in the model. For example, Ludwig et al. (2013) include price discount in the model when studying the relationship between product reviews and conversion rate to control for this factor.

External Validity

To achieve external validity, researchers should attempt to ensure that the effects found in the text apply outside of the research framework. Because text analysis often uses naturally occurring data that is often of large magnitude, it tends have a relatively high degree of external validity relative to, for example, lab experiments. However, establishing external validity is still necessary due to threats to validity from sampling bias, overfitting, and single-method bias. For example, online reviews may be biased due to self-selection among those who elected to review a product (Schoenmüller, Netzer, and Stahl 2019).

Predictive validity. Predictive validity is threatened when the construct, though perhaps properly measured, does not have the expected effects on a meaningful second variable. For example, if consumer sentiment falls but customer satisfaction remains high, predictive validity could be called into question. To ensure predictive validity, text-based constructs can be linked to key performance measures such as sales (e.g., Fossen and Schweidel 2019) or consumer engagement (Ashley and Tuten 2015). If a particular construct has been theoretically linked to a performance metric, then any text-based measure of that construct should also be linked to that performance metric. Tirunillai and Tellis (2012) show that the volume of Twitter activity affects stock price, but they find mixed results for the predictive validity of sentiment, with negative sentiment being predictive but positive sentiment having no effect.

Generalizability can be threatened when researchers base results on a single data set because it is unknown whether the findings, model, or algorithm would apply in the same way to other texts or outside of textual measurements. Generalizability of the results can be established by viewing the results of text analysis along with other measures of attitude and behavioral outcomes. For example, Netzer et al. (2012) test their substantive conclusions and methodology on message boards of both automobile discussions and drug discussions from WebMD. Evaluating the external validity and generalizability of the findings is key, because the analysis of text drawn from a particular source may not reflect consumers more broadly (e.g., Schweidel and Moe 2014).

Robustness. Robustness can be limited when there is only one metric or method used in the model. Researchers can ensure robustness by using different measures for relationships (e.g., Pearson correlation, cosine similarity, lift) and probing results by relaxing different assumptions. The use of holdout samples and k-fold cross-validation methods can prevent researchers from overfitting their models and ensure that relationships found in the data set will hold with other data as well (Jurafsky et al. 2014; see also Humphreys and Wang 2017). Probing on different "cuts" of the data can also help. Berger and Packard (2018), for example, compare lyrics from different genres, and Ludwig et al. (2013) include reviews of both fiction and nonfiction books.

Finally, researchers should bear in mind the limitations of text itself. There are thoughts and feelings that consumers, managers, or other stakeholders may not express in text. The form of communication (e.g., tweets, annual reports) may also shape the message; some constructs may not be explicit enough to be measured with automated text analysis. Furthermore, while textual information can often involve large samples, these samples may not be representative. Twitter users, for example, tend to be younger and more educated (Smith and Anderson 2018). Those who contribute textual information, particularly in social media, may represent polarized points of view. When evaluating cultural products or social media, one should consider the system in which they are generated. Often viewpoints are themselves filtered through a cultural system (Hirsch 1986; McCracken 1988) or elevated by an algorithm, and the products make it through this process may share certain characteristics. For this reason, researchers and firms should use caution when making attributions on the basis of a cultural text. It is not necessarily a reflection of reality (Jameson 2005) but rather may represent ideals, extremes, or institutionalized perceptions, depending on the context.

Future Research Agenda

We hope this article encourages more researchers and practitioners to think about how they can incorporate textual data into their research. Communication and linguistics are at the core of studying text in marketing. Automated text analysis opens the black box of interactions, allowing researchers to directly access what is being said and how it is said in marketplace communication. The notion of text as indicative of meaning-making processes creates fascinating and truly novel research questions and challenges. There are many methods and approaches available, and there is no space to do all of them justice. While we have discussed several research streams, given the novelty of text analysis, there are still ample opportunities for future research, which we discuss next.

Using Text to Reach Across the Marketing Discipline

Returning to how text analysis can unite the tribes of marketing, it is worth highlighting a few areas that have mostly been examined by one research tradition in marketing where fruitful cross-pollination between tribes is possible through text analysis. Brand communities were first identified and studied by researchers coming from a sociology perspective (Muñiz and O'Guinn 2001). Later, qualitative and quantitative researchers further refined the concepts, identifying a distinct set of roles and status in the community (e.g., Mathwick, Wiertz, and De Ruyter 2007). Automated text analysis allows researchers to study how consumers in these communities interact at scale and in a more quantifiable manner—for instance, examining how people with different degrees of power use language and predict group outcomes based on quantifiably different dynamics (e.g., Manchanda, Packard, and Pattabhitamaiah 2015). Researchers can track influence, for example, by investigating which types of users initiate certain words or phrases and which others pick up on them. Research could examine whether people begin to enculturate to the language of the community over time and predict which individuals may be more likely to stay or leave on the basis of how well they adapt to the group's language (Danescu-Niculescu-Mizil et al. 2013; Srivastava and Goldberg 2017). Quantitative or machine learning researchers might capture the most commonly discussed topics and how these dynamically change over the evolution of the community. Interpretive researchers might examine how these terms link conceptually, to find underlying community norms that lead members to stay. Marketing strategy researchers might then use or develop dictionaries to connect these communities to firm performance and to offer directions for firms regarding how to keep members participating across different brand communities (or contexts).

The progression can flow the other way as well. Outside of a few early investigations (e.g., Dichter 1966), word of mouth was originally studied by quantitative researchers interested in whether interpersonal communication actually drove individual and market behavior (e.g., Chevalier and Mayzlin 2006; Iyengar, Van den Bulte, and Valente 2011). More recently, however, behavioral researchers have begun to study the underlying drivers of word of mouth, looking at why people talk about and share some stories, news, and information rather than others (Berger and Milkman 2012; De Angelis et al. 2012; for a review, see Berger [2014]). Marketing strategy researchers might track the text of word-of-mouth interactions to predict the emergence of brand crises or social media firestorms (e.g., Zhong and Schweidel 2019) as well as when, if, and how to respond (Herhausen et al. 2019).

Consumer–firm interaction is also a rich area to examine. Behavioral researchers could use the data from call centers to better understand interpersonal communication between consumers and firms and record what drives customer satisfaction (e.g., Packard and Berger 2019a; Packard, Moore, and McFerran 2018). The back-and-forth between customers and agents could be used to understand conversational dynamics. More quantitative researchers should use the textual features of call centers to predict outcomes such as churn and even go beyond text to examine vocal features such as tone, volume, and speed of speech. Marketing strategy researchers could use calls to understand how customer-centric a company is or assess the quality, style, and impact of its sales personnel.

Finally, it is worth noting that different tribes not only have different skill sets but also often study substantively different types of textual communication. Consumer-to-consumer communication is often studied by researchers in consumer behavior, whereas marketing strategy researchers more often tend to study firm-to-consumer and firm-to-firm communication. Collaboration among researchers from the different subfields may allow them to combine these different sources of textual data.

There is ample opportunity to apply theory developed in one domain to enhance another. Marketing strategy researchers, for example, often use transaction economics to study business-to-business relationships through agency theory, but these approaches may be equally beneficial when studying consumer-to-consumer communications.

Broadening the Scope of Text Research

As noted in Table 1, certain text flows have been studied more than others. A large portion of existing work has focused on consumers communicating to one another through social media and online reviews. The relative availability of such data has made it a rich area of study and an opportunity to apply text analysis to marketing problems.[3] Furthermore, for this area to grow, researchers need to branch out. This includes expanding (1) data sources, (2) actors examined, and (3) research topics.

Expand data sources used. Offline word of mouth, for example, can be examined to study what people talk about and conversational dynamics. Doctor–patient interactions can be studied to understand what drives medical adherence. Text items such as yearbook entries, notes passed between students, or the text of speed dating conversations can be used to examine relationship formation, maintenance, and dissolution. Using offline data requires carefully transcribing content, which increases the amount of effort required but opens up a range of interesting avenues of study. For example, we know very little about the differences between online recommendations and face-to-face recommendations, where the latter also include the interplay between verbal and nonverbal information. Moreover, in the new era of "perpetual contact" our understanding of cross-message and cross-channel implications is limited. Research by Batra and Keller (2016) and Villarroel Ordenes et al. (2018) suggests that appropriate sequencing of messages matters; it might similarly matter across channels and modality. Given the rise of technology-enabled realities (e.g., augmented reality, virtual reality, mixed reality), assistive robotics, and smart speakers, understanding the roles and potential differences between language and nonverbal cues could be achieved using these novel data sources.

Expand dyads between text producers and text receivers. There are numerous dyads relevant to marketing in which text plays a crucial role. We discuss just a few of the areas that deserve additional research.

Considering consumer–firm interactions, we expect to see more research leveraging the rich information exchanged between consumers and firms through call centers and chats (e.g., Packard and Berger 2019a; Packard, Moore, and McFerran 2018). These interactions often reflect inbound communication between customers and the firm, which can have important implications for the relationship between parties. In addition, how might the language used on packaging or in brand mission statements reflect the nature of organizations and their relationship to their consumers? How might the language that is most impactful in sales interactions differ from the language that is most useful in customer service interactions? Research could also probe how the impact of such language varies across contexts. The characteristics of language used by consumer packaged goods brands and pharmaceuticals brands in direct-to-consumer advertising likely differ. Similarly, the way in which consumers process the language used in disclosures in advertisements for pharmaceuticals (e.g., Narayanan, Desiraju, and Chintagunta 2004) and political candidates (e.g., Wang, Lewis, and Schweidel 2018) may vary.

Turning to firm-to-firm interactions, most conceptual frameworks on business-to-business (B2B) exchange relations emphasize the critical role of communication (e.g., Palmatier, Dant, and Grewal 2007). Communicational aspects have been linked to important B2B relational measures such as commitment, trust, dependence, relationship satisfaction, and relationship quality. Yet research on actual, word-level B2B communication is very limited. For example, very little research has examined the types of information exchanged between salespeople and customers in offline settings. The ability to gather and transcribe data at scale points to important opportunities to do so. As for within-firm communication, researchers could study informal communications such as marketing-related emails, memos, and agendas generated by firms and consumed by their employees.

Similarly, while a great deal of work in accounting and finance has begun to use annual reports as a data source (for a review, see Loughran and McDonald [2016]), marketing researchers have paid less attention to this area to study communication with investors. Most research has used this data to predict outcomes such as stock performance and other measures of firm valuation. Given recent interest in linking marketing-related activities to firm valuation (e.g., McCarthy and Fader 2018), this may be an area to pursue further. All firm communication, including required documents such as annual reports or discretionary forms of communication such as advertising and sales interactions, can be used to measure variables such as market orientation, marketing capabilities, marketing leadership styles, and even a firm's brand personality.

There are also ample research opportunities in the interactions between consumers, firms, and society. Data about the broader cultural and normative environment of firms, such as news media and government reports, may be useful to shed light on the forces that shape markets. To understand how a company such as Uber navigates resistance to market change, for example, one might study transcripts of town hall meetings and other government documents in which citizen input is heard and answered. Exogenous shocks in the forms of social movements such as #metoo and #blacklivesmatter have affected marketing communication and brand image. One potential avenue for future research is to take a cultural

[3] While readily available data facilitates research, there are downsides to be recognized, including the representatives of such data and the terms of service that govern the use of this data.

branding approach (Holt 2016) to study how different publics define, shape, and advocate for certain meanings in the marketplace. Firms and their brands do not exist in a vacuum, independent of the society in which they operate. Yet limited research in marketing has considered how text can be used to derive firms' intentions and actions at the societal level. For example, scholars have shown how groups of consumers such as locavores (i.e., people who eat locally grown food; Thompson and Coskuner-Balli 2007), fashionistas (Scaraboto and Fischer 2012), and bloggers (McQuarrie, Miller, and Phillips 2012) shape markets. Through text analysis, the effect of the intentions of these social groups on the market can then be measured and better understood.

Another opportunity for future research is the use of textual data to study culture and cultural success. Topics such as cultural propagation, artistic change, and the diffusion of innovations have been examined across disciplines with the goal of understanding why certain products succeed while others fail (Bass 1969; Boyd and Richerson 1986; Cavalli-Sforza and Feldman 1981; Rogers 1995; Salganik, Dodds, and Watts 2006; Simonton 1980). While success may be random (Bielby and Bielby 1994; Hirsch 1972), another possibility is that cultural items succeed or fail on the basis of their fit with consumers (Berger and Heath 2005). By quantifying aspects of books, movies, or other cultural items quickly and at scale, researchers can measure whether concrete narratives are more engaging, whether more emotionally volatile movies are more successful, whether songs that use certain linguistic features are more likely to top the Billboard charts, and whether books that evoke particular emotions sell more copies. While not as widely available as social media data, more and more data on cultural items has recently become available. Data sets such as the Google Books corpus (Akpinar and Berger 2015), song lyric websites, or movie script databases provide a wealth of information. Such data could enable analyses of narrative structure to identify "basic plots" (e.g., Reagan et al. 2016; Van Laer et al. 2019).

Key Marketing Constructs (That Could Be) Measured with Text

Beginning with previously developed ways of representing marketing constructs can help some researchers address validity concerns. This section details a few of these constructs to aid researchers who are beginning to use text analysis in their work (see the Web Appendix). Using prior operationalization of a construct can ensure concurrent validity—helping build the literature in a particular domain—but researchers should take steps to ensure that the prior operationalization has construct validity with their data set.

At the individual level, sentiment and satisfaction are perhaps some of the most common measurements (e.g., Büschken and Allenby, 2016; Homburg, Ehm, and Artz 2015; Herhausen et al. 2019; Ma, Baohung, and Kekre 2015; Schweidel and Moe 2014) and have been validated in numerous contexts. Other aspects that may be extracted from text include the authenticity and emotionality of language, which have also been explored through robust surveys and scales or by combining multiple existing measurements (e.g., Mogilner, Kamvar, and Aaker 2011; Van Laer et al. 2019). There are also psychological constructs, such as personality type and construal level (Kern et al. 2016; Snefjella and Kuperman 2015), that are potentially useful for marketing researchers and could also be inferred from the language used by consumers.

Future work in marketing studying individuals might consider measurements of social identification and engagement. That is, researchers currently have an idea of positive or negative consumer sentiment, but they are only beginning to explore emphasis (e.g., Rocklage and Fazio 2015), trust, commitment, and other modal properties. To this end, harnessing linguistic theory of pragmatics and examining phatics over semantics could be useful (see, e.g., Villarroel et al. 2017). Once such work is developed, we recommend that researchers carefully validate approaches proposed to measure such constructs along the lines described previously.

At the firm level, constructs have been identified in firm-produced text such as annual reports and press releases. Market orientation, advertising goals, future orientation, deceitful intentions, firm focus, and innovation orientation have all been measured and validated using this material (see Web Appendix Table 1). Work in organizational studies has a history of using text analysis in this area and might provide some inspiration and validation in the study of the existence of managerial frames for sensemaking and the effect of activists on firm activities.

Future work in marketing at the firm level could further refine and diversify measurements of strategic orientation (e.g., innovation orientation, market-driving vs. market-driven orientations). Difficult-to-measure factors deep in the organizational culture, structure, or capabilities may be revealed in the words the firm, its employees, and external stakeholders use to describe it (see Molner, Prabhu, and Yadav [2019]). Likewise, the mindsets and management style of marketing leaders may be discerned from the text they use (see Yadav, Prabhu, and Chandy [2007]). Firm attributes that are important outcomes of firm action (e.g., brand value) could also be explored using text (e.g., Herhausen et al. 2019). In this case, there is an opportunity to use new kinds of data. For instance, internal, employee-based brand value could be measured with text on LinkedIn or Glassdoor. Finally, more subtle attributes of firm language, including conflict, ambiguity, or openness, might provide some insight into the effects of managerial language on firm success. For this, it may be useful to examine less formal textual data of interactions such as employee emails, salesperson calls, or customer service center calls.

Less work in marketing has measured constructs on the social or cultural level, but work in this vein tends to focus on how firms fit into the cultural fabric of existing meanings and norms. For instance, institutional logics and legitimacy have been measured by analyzing media text, as has the rise of brand publics that increase discussion of brands within a culture (Arvidsson and Caliandro 2016).

At the cultural level, marketing research is likely to maintain a focus on how firms fit into the cultural environment, but it may also look to how the cultural environment affects consumers. For instance, measurement of cultural uncertainty, risk, hostility, and change could benefit researchers interested in the effects of culture on both consumer and firm effects as well as the effects of culture and society on government and investor relationships. Measuring openness and diversity through text are also timely topics to explore and might inspire innovations in measurement, focusing on, for example, language diversity rather than the specific content of language. Important cultural discourses such as language around debt and credit could also be better understood through text analysis. Measurement of gender- and race- related language could be useful in exploring diversity and inclusion in the way firms and consumers react to text from a diverse set of writers.

Opportunities and Challenges Provided by Methodological Advances

Opportunities. As the development of text analysis tools advances, we expect to see new and improved use of these tools in marketing, which can enable scholars to answer questions we could not previously address or have addressed only in a limited manner. Here are a few specific method-driven directions that seem promising.

First, the vast majority of the approaches used for text analysis in marketing (and elsewhere) rely on bag-of-words approaches, and thus, the ability to capture true linguistic relationships among words beyond their cooccurrence was limited. However, in marketing we are often interested in capturing the relationship among entities. For example, what problems or benefits did the customer mention about a particular feature of a particular product? Such approaches require capturing a deeper textual relationship among entities than is commonly used in marketing. We expect to see future development in these areas as deep learning and NLP-based approaches enable researchers to better capture semantic relationships.

Second, in marketing we are often interested in the latent intention or latent states of writers when creating text, such as their emotions, personality, and motivations. Most of the research in this area has relied on a limited set of dictionaries (primarily the LIWC dictionary) developed and validated to capture such constructs. However, these dictionaries are often limited in capturing nuanced latent states or latent states that may manifest differently across contexts. Similar to advances made in areas such as image recognition, with the availability of a large number of human-coded training data (often in the millions) combined with deep learning tools, we hope to see similar approaches being taken in marketing to capture more complex behavioral states from text. This would require an effort to human-code a large and diverse set of textual corpora for a wide range of behavioral states. Transfer learning methods commonly used in deep learning tools such as conventional neural nets can then be used to apply the learning from the more general training data to any specific application.

Third, there is also the possibility of using text analysis to personalize customer–firm interactions. Using machine learning, text analysis can also help personalize the customer interaction by detecting consumer traits (e.g., personality) and states (e.g., urgency, irritation) and perhaps eventually predicting traits associated with value to the firm (e.g., customer lifetime value). After analysis, firms can then tailor customer communication to match linguistic style and perhaps funnel consumers to the appropriate firm representative. The stakes of making such predictions may be high, mistakes costly, and there are clearly contexts in which using artificial intelligence impedes constructing meaningful customer–firm relationships (e.g., health care; Longoni, Bonezzi, and Morewedge 2019).

Fourth, while our discussion has focused on textual content, text is just one example of unstructured data, with audio, video, and image being others. Social media posts often marry text with images or videos. Print advertising usually overlays text on a carefully constructed visual. Although television advertising may not include text on the screen, it may have an audio track that contains text that progresses simultaneously with the video.

Until recently, text data has received the most attention, mainly due to the presence of tools to extract meaningful features. That said, tools such as Praat (Boersma 2001) allow researchers to extract information from audio (e.g., Van Zant and Berger 2019). One of the advantages of audio data over text data is that it provides richness in the form of tone and voice markers that can add to the actual words expressed (e.g., Xiao, Kim, and Ding 2013). This enables researchers to study not just what was said, but how it was said, examining how pitch, tone, and other vocal or paralinguistic features shape behavior.

Similarly, recent research has developed approaches to analyze images (e.g., Liu, Xuan et al. 2018), either characterizing the content of the image or identifying features within an image. Research into the impact of the combination of text and images is sparse (e.g., Hartmann et al. 2019). For example, images can be described in terms of their colors. In the context of print advertising, textual content may be less persuasive when used in conjunction with images of a particular color palette, whereas other color palettes may enhance the persuasiveness of text. Used in conjunction with simple images, the importance of text may be quite pronounced. But, when text is paired with complex imagery, viewers may attend primarily to the image, diminishing the impact of the text. If this is the case, legal disclosures that are part of an advertisement's fine print may not attract the audience's attention.

Analogous questions arise as to the role that text plays when incorporated into videos. Research has proposed approaches to characterize video content (e.g., Liu et al. 2018). In addition to comprising the script of the video, text may also appear visually. In addition to the audio context in which text appears, its impact may depend on the visuals that appear simultaneously. It may also be the case that its position within a video, relative to the start of the video, may moderate its effectiveness. For example, emotional text content that is spoken later in a video may be less persuasive for several reasons (e.g., the audience may have ceased paying attention by the time the text is spoken). Alternatively, the visuals with which the audio is paired may be more compelling to viewers, or the previous content of the video may have depleted a viewer's attentional resources. As our discussion of both images and videos suggests, text is but one component of marketing communications. Future research must investigate its interplay with other characteristics, including not only the content in which it appears but also when it appears (e.g., Kanuri, Chen, and Sridhar 2018), and in what media.

Challenges. While there are a range of opportunities, textual data also brings with it various challenges. First is the interpretation challenge. In some ways, text analysis seems to provide more objective ways of measuring behavioral processes. Rather than asking people how much they focused on themselves versus others when sharing word of mouth, for example, one can count the number of first-person (e.g., "I") and second-person (e.g., "you"; Barasch and Berger 2014) pronouns, providing what seems more like ground truth. But while part of this process is certainly more objective (e.g., the number of different types of pronouns), the link between such measures and underlying processes (i.e., what it says about the word-of-mouth transmitter) still requires some degree of interpretation. Other latent modes of behavior are even more difficult to count. While some words (e.g., "love") are generally positive, for example, how positive they are may depend heavily on idiosyncratic individual differences as well as the context.

More generally, there is challenge and opportunity in understanding the context in which textual information appears. While early work in this space, particularly research using entity extraction, asked questions such as how much emotion is in a passage of text, more accurate answers to that question take must take context into account. A restaurant review may contain lots of negative words, for example, but does that mean the person hates the food, the service, or the restaurant more generally? Songs that contain more second person-pronouns (e.g., "you") may be more successful (Packard and Berger 2019b), but to understand why, it helps to know whether the lyrics use "you" as the subject or object of the sentence. Context provides meaning, and the more one understands not just which words are being used but also how they are being used, the easier it will be to extract insight. Dictionary-based tools are particularly susceptible to variation in the context in which the text appears, as dictionaries are often created in a context-free environment to match multiple contexts. Whenever possible, it is advised to use a dictionary that was created for the specific context of study (e.g., the financial sentiment tool developed by Loughran and McDonald [2016]).

As mentioned previously, there are also numerous methodological challenges. Particularly when exploring the "why," hundreds of features can be extracted, making it important to think about multiple hypothesis testing (and use of Bonferroni and other corrections). Only the text used by the text creator is available, so in some sense there is self-selection. Both the individuals who decide to contribute and the topics people decide to raise in their writing may suffer from self-selection. Particularly when text is used to measure (complex) behavioral constructs, validity of the constructs needs to be considered. In addition, for most researchers, analyzing textual information requires retooling and learning a whole new set of skills.

Data privacy challenges represent a significant concern. Research often uses online product reviews and sales ranking data scraped from websites (e.g., Wang, Mai, and Chiang 2013) or consumers' social media activity scraped from the platform (e.g., Godes and Mayzlin 2004; Tirunillai and Tellis 2012). Although such approaches are common, legal questions have started to arise. LinkedIn was unsuccessful in its attempt to block a startup company from scraping data that was posted on users' public profiles (Rodriguez 2017). While scraping public data may be permissible under the law, it may conflict with the terms of service of those platforms that have data of interest to researchers. For example, Facebook deleted accounts of companies that violated its data-scraping policies (Nicas 2018).[4] Such decisions raise important questions about the extent to which digital platforms can control access to content that users have chosen to make publicly available.

As interest in extracting insights from digitized text and other forms of digitized content (e.g., images, videos) grows, researchers should ensure that they have secured the appropriate permissions to conduct their work. Failure to do so may result in it becoming more difficult to conduct such projects. One potential solution is the creation of an academic data set, such as that made available by Yelp (https://www.yelp.com/dataset), which may contain outdated or scrubbed data to ensure that it does not pose any risk to the company's operations or user privacy.

The collection and analysis of digitized text, as well as other user-created content, also raises questions around users' expectations for privacy. In the wake of the European Union's General Data Protection Regulation and revelations about Cambridge Analytica's ability to collect user data from Facebook, researchers must be mindful of the potential abuses of their work. We should also consider the extent to which we are overstepping the intended use of user-generated content. For example, while a user may understand that actions taken on Facebook may result in their being targeted with specific advertisements for brands with which they have interacted, they may not anticipate the totality of their Facebook and Instagram activity being used to construct psychographic profiles that may be used by other brands. Understanding consumers' privacy preferences with regard to their online behaviors and the text they make available could provide important guidance for practitioners and researchers alike. Another rich area for future research is the advancement of the precision with which marketing can be implemented while minimizing intrusions of privacy (e.g., Provost et al. 2009).

[4] Facebook's terms of service with regard to automated data collection can be found at https://www.facebook.com/apps/site_scraping_tos_terms.php.

Concluding Thoughts

Communication is an important facet of marketing that encompasses communication between organizations and their partners, between businesses and their consumers, and among consumers. Textual data holds details of these communications, and through automated textual analysis, researchers are poised to convert this raw material into valuable insights. Many of the recent advances in the use of textual data were developed in fields outside of marketing. As we look toward the future and the role of marketers, these recent advancements should serve as exemplars. Marketers are well positioned at the interface between consumers, firms, and organizations to leverage and advance tools to extract textual information to address some of the key issues faced by business and society today, such as the proliferation of misinformation, the pervasiveness of technology in our lives, and the role of marketing in society. Marketing offers an invaluable perspective that is vital to this conversation, but it will only be by taking a broader perspective, breaking theoretical and methodological silos, and engaging with other disciplines that our research can reach its largest possible audience to affect the public discourse. We hope this framework encourages a reflection on the boundaries that have come to define marketing and opens avenues for future groundbreaking insights.

Editors

Christine Moorman and Harald van Heerde

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Akpinar, Ezgi, and Jonah Berger (2015), "Drivers of Cultural Success: The Case of Sensory Metaphors," *Journal of Personality and Social Psychology*, 109 (1), 20–34.

Alessa, Ali, and Miad Faezipour (2018), "A Review of Influenza Detection and Prediction Through Social Networking Sites," *Theoretical Biology and Medical Modelling*, 15 (1), 1–27.

Anderson, Eric T., and Duncan I. Simester (2014), "Reviews Without a Purchase: Low Ratings, Loyal Customers, and Deception," *Journal of Marketing Research*, 51 (3), 249–69.

Arsel, Zeynep, and Jonathan Bean (2013), "Taste Regimes and Market-Mediated Practice," *Journal of Consumer Research*, 39 (5), 899–917.

Arvidsson, Adam, and Alessandro Caliandro (2016), "Brand Public," *Journal of Consumer Research*, 42 (5), 727–48.

Ashley, Christy, and Tracy Tuten (2015), "Creative Strategies in Social Media Marketing: An Exploratory Study of Branded Social Content and Consumer Engagement," *Psychology & Marketing*, 32 (1), 15–27.

Barasch, Alixandra, and Jonah Berger (2014), "Broadcasting and Narrowcasting: How Audience Size Affects What People Share," *Journal of Marketing Research*, 51 (3), 286–99.

Bass, Frank M. (1969), "A New Product Growth for Model Consumer Durables," *Management Science*, 15 (5), 215–27.

Batra, Rajeev, and Kevin L. Keller (2016), "Integrating Marketing Communications: New Findings, New Lessons, and New Ideas," *Journal of Marketing*, 80 (6), 122–45.

Berger, Jonah (2014), "Word of Mouth and Interpersonal Communication: A Review and Directions for Future Research," *Journal of Consumer Psychology*, 24 (4), 586–607.

Berger, Jonah, and Chip Heath (2005), "Idea Habitats: How the Prevalence of Environmental Cues Influences the Success of Ideas," *Cognitive Science*, 29 (2), 195–221.

Berger, Jonah, Yoon Duk Kim, and Robert Meyer (2019), "Emotional Volatility and Cultural Success," working paper.

Berger, Jonah, and Katherine L. Milkman (2012), "What Makes Online Content Viral?" *Journal of Marketing Research*, 49 (2), 192–205.

Berger, Jonah, Wendy W. Moe, and David A. Schweidel (2019), "What Makes Stories More Engaging? Continued Reading in Online Content," working paper.

Berger, Jonah, and Grant Packard (2018), "Are Atypical Things More Popular?" *Psychological Science*, 29 (7), 1178–84.

Berman, Ron, Shiri Melumad, Colman Humphrey, and Robert Meyer (2019), "A Tale of Two Twitterspheres: Political Microblogging During and After the 2016 Primary and Presidential Debates," *Journal of Marketing Research*, 56 (6), doi:10.1177/0022243719861923.

Bielby, William, and Denise Bielby (1994), "'All Hits Are Flukes': Institutionalized Decision Making and the Rhetoric of Network Prime-Time Program Development," *American Journal of Sociology*, 99 (5), 1287–1313.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003), "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, 3, 993–1022.

Boersma, Paul (2001), "Praat, a System for Doing Phonetics by Computer," *Glot International*, 5 (9/10), 341–45.

Boghrati, Reihane, and Jonah Berger (2019) "Quantifying 60 Years of Misogyny in Music," working paper.

Bollen, Johan, Huina Mao, and Xiaojun Zeng (2011), "Twitter Mood Predicts the Stock Market," *Journal of Computational Science*, 2 (1), 1–8.

Borah, Abhishek, and Gerard J. Tellis (2016), "Halo (Spillover) Effects in Social Media: Do Product Recalls of One Brand Hurt or Help Rival Brands?" *Journal of Marketing Research*, 53 (2), 143–60.

Boyd, Robert, and Peter Richerson (1986), *Culture and Evolutionary Process*. Chicago: University of Chicago Press.

Büschken, Joachim, and Greg M. Allenby (2016), "Sentence-Based Text Analysis for Customer Reviews," *Marketing Science*, 35 (6), 953–75.

Cavalli-Sforza, Luigi Luca, and Marcus W. Feldman (1981), *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton, NJ: Princeton University Press.

Chen, Zoey, and Nicholas H. Lurie (2013), "Temporal Contiguity and Negativity Bias in the Impact of Online Word of Mouth," *Journal of Marketing Research*, 50 (4), 463–76.

Chevalier, Judith A., and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345–54.

Cohn, Michael A., Matthias R. Mehl, and James W. Pennebaker (2004), "Linguistic Markers of Psychological Change Surrounding September 11, 2001," *Psychological Science*, 15 (10), 687–93.

Cook, Thomas D., and Donald Thomas Campbell (1979), *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.

Danescu-Niculescu-Mizil, Christian, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts (2013), "No Country for Old Members: User Lifecycle and Linguistic Change in Online Communities," in *Proceedings of the 22nd International Conference on World Wide Web*. New York: Association for Computing Machinery, 307–18.

Das, Sanjiv, and Mike Y. Chen (2007), "Yahoo! for Amazon: Sentiment Extraction from Small Talk on the Web," *Management Science*, 53 (9), 1375–88.

De Angelis, Matteo, Andrea Bonezzi, Alessandro M. Peluso, Derek Rucker, and Michele Costabile (2012), "On Braggarts and Gossips: A Self-Enhancement Account of Word-of-Mouth Generation and Transmission," *Journal of Marketing Research*, 49 (4), 551–63.

Dichter, E. (1966), "How Word-of-Mouth Advertising Works," *Harvard Business Review*, 44 (6), 147–66.

Dodds, Peter Sheridan, Harris Kameron Decker, Isabel M. Kloumann, Catherine A. Bliss, and Christopher M. Danforth (2011), "Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter," *PLoS ONE*, 6 (12), e26752.

Dowling, Grahame R., and Boris Kabanoff (1996), "Computer-Aided Content Analysis: What Do 240 Advertising Slogans Have in Common?" *Marketing Letters*, 7 (1), 63–75.

Eliashberg, Jehoshua, Sam K. Hui, and Z. John Zhang (2007a), "From Story Line to Box Office: A New Approach for Green-Lighting Movie Scripts," *Management Science*, 53 (6), 881–93.

Eliashberg, Jehoshua, Sam K. Hui, and Z. John Zhang (2007b), "Assessing Box Office Performance Using Movie Scripts: A Kernel-Based Approach," *IEEE Transactions on Knowledge & Data Engineering*, 26 (11), 2639–48.

Feldman, Ronen, Oded Netzer, Aviv Peretz, and Binyamin Rosenfeld (2015), "Utilizing Text Mining on Online Medical Forums to Predict Label Change Due to Adverse Drug Reactions," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery, 1779–88.

Fiss, Peer C., and Paul M. Hirsch (2005), "The Discourse of Globalization: Framing and Sensemaking of an Emerging Concept," *American Sociological Review*, 70 (1), 29–52.

Fossen, Beth L., and David A. Schweidel (2019), "Social TV, Advertising, and Sales: Are Social Shows Good for Advertisers?" *Marketing Science*, 38 (2), 274–95.

Gandomi, Amir, and Murtaza Haider (2015), "Beyond the Hype: Big Data Concepts, Methods, and Analytics," *International Journal of Information Management*, 35 (2), 137–44.

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou (2018), "Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes," *Proceedings of the National Academy of Sciences*, 115 (16), E3635–44.

Gebhardt, Gary F., Francis J. Farrelly, and Jodie Conduit (2019), "Market Intelligence Dissemination Practices," *Journal of Marketing*, 83 (3), 72–90.

Ghose, Anindya, and Panagiotis G. Ipeirotis (2011), "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," *IEEE Transactions on Knowledge and Data Engineering*, 23 (10), 1498–1512.

Godes, David, and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545–60.

Godes, David, and José C. Silva (2012), "Sequential and Temporal Dynamics of Online Opinion," *Marketing Science*, 31 (3), 448–73.

Goffman, Erving (1959), "The Moral Career of the Mental Patient," *Psychiatry*, 22 (2), 123–42.

Gopalan, Prem, Jake M. Hofman, and David M. Blei (2013), "Scalable Recommendation with Poisson Factorization," (accessed August 19, 2019), https://arxiv.org/abs/1311.1704.

Hancock, Jeffrey T., Lauren E. Curry, Saurabh Goorha, and Michael Woodworth (2007), "On Lying and Being Lied To: A Linguistic Analysis of Deception in Computer-Mediated Communication," *Discourse Processes*, 45 (1), 1–23.

Hartmann, Jochen, Mark Heitmann, Christina Schamp, and Oded Netzer (2019), "The Power of Brand Selfies in Consumer-Generated Brand Images," working paper.

Hartmann, Jochen, Juliana Huppertz, Christina Schamp, and Mark Heitmann (2018), "Comparing Automated Text Classification Methods," *International Journal of Research in Marketing*, 36 (1), 20–38.

Hennig-Thurau, Thorsten, Caroline Wiertz, and Fabian Feldhaus (2015), "Does Twitter Matter? The Impact of Microblogging Word of Mouth on Consumers' Adoption of New Movies," *Journal of the Academy of Marketing Science*, 43 (3), 375–94.

Herhausen, Dennis, Stephan Ludwig, Dhruv Grewal, Jochen Wulf, and Marcus Schögel (2019), "Detecting, Preventing, and Mitigating Online Firestorms in Brand Communities," *Journal of Marketing*, 83 (3), 1–21.

Hill, Vanessa, and Kathleen M. Carley (1999), "An Approach to Identifying Consensus in a Subfield: The Case of Organizational Culture," *Poetics*, 27 (1), 1–30.

Hirsch, Arnold R. (1986) "The Last 'Last Hurrah'," *Journal of Urban History*, 13 (1), 99–110.

Hirsch, Peer M. (1972) "Processing Fads and Fashions: An Organization-Set Analysis of Cultural Industry Systems," *American Journal of Sociology*, 77 (4), 639–59.

Holt, Douglas (2016), "Branding in the Age of Social Media," *Harvard Business Review*, 94 (3), 40–50.

Homburg, Christian, Laura Ehm, and Martin Artz (2015), "Measuring and Managing Consumer Sentiment in an Online Community Environment," *Journal of Marketing Research*, 52 (5), 629–41.

Huang, Karen, Michael Yeomans, Alison W. Brooks, Julia Minson, and Francesca Gino (2017), "It Doesn't Hurt to Ask: Question-

Asking Increases Liking," *Journal of Personality and Social Psychology*, 113 (3), 430–52.

Humphreys, Ashlee (2010), "Semiotic Structure and the Legitimation of Consumption Practices: The Case of Casino Gambling," *Journal of Consumer Research*, 37 (3), 490–510.

Humphreys, Ashlee, and Kathryn A. LaTour (2013), "Framing the Game: Assessing the Impact of Cultural Representations on Consumer Perceptions of Legitimacy," *Journal of Consumer Research*, 40 (4), 773–95.

Humphreys, Ashlee, and Rebecca Jen-Hui Wang (2017), "Automated Text Analysis for Consumer Research," *Journal of Consumer Research*, 44 (6), 1274–1306.

Hutto, Clayton J., and Eric Gilbert (2014), "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," in *Proceedings of the Eighth International Conference on Weblogs and Social Media*. Palo Alto, CA: Association for the Advancement of Artificial Intelligence.

Iyengar, Raghuram, Christopher Van den Bulte, and Thomas Valente (2011), "Opinion Leadership and Social Contagion in New Product Diffusion," *Marketing Science*, 30 (2), 195–212.

Jameson, Fredric (2005), *Archaeologies of the Future: The Desire Called Utopia and Other Science Fictions*. New York: Verso.

Jurafsky, Dan, Victor Chahuneau, Bryan R. Routledge, and Noah A. Smith (2014), "Narrative Framing of Consumer Sentiment in Online Restaurant Reviews," *First Monday*, 19 (4), https://firstmonday.org/ojs/index.php/fm/article/view/4944/3863.

Kanuri, Vamsi K., Yixing Chen, and Shrihari (Hari) Sridhar (2018), "Scheduling Content on Social Media: Theory, Evidence, and Application," *Journal of Marketing*, 82 (6), 89–108.

Kern, Margaret L., Gregory Park, Johannes C. Eichstaedt, H. Andrew Schwartz, Maarten Sap, and Laura K. Smith (2016), "Gaining Insights from Social Media Language: Methodologies and Challenges," *Psychological Methods*, 21 (4), 507–25.

Kübler, Raoul V., Anatoli Colicev, and Koen Pauwels (2017) "Social Media's Impact on Consumer Mindset: When to Use Which Sentiment Extraction Tool," *Marketing Science Institute Working Paper Series* 17-122-09.

Kulkarni, Dipti (2014), "Exploring Jakobson's 'Phatic Function' in Instant Messaging Interactions," *Discourse & Communication*, 8 (2), 117–36.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015), "Deep Learning," *Nature*, 521 (7553), 436–44.

Lee, Thomas Y., and Eric T. Bradlow (2011), "Automated Marketing Research Using Online Customer Reviews," *Journal of Marketing Research*, 48 (5), 881–94.

Li, Feng, and Timon C. Du (2011), "Who Is Talking? An Ontology-Based Opinion Leader Identification Framework for Word-of-Mouth Marketing in Online Social Blogs," *Decision Support Systems*, 51 (1), 190–97.

Liu, Jia, and Olivier Toubia (2018), "A Semantic Approach for Estimating Consumer Content Preferences from Online Search Queries," *Marketing Science*, 37 (6), 855–1052.

Liu, Liu, Daria Dzyabura, and Natalie Mizik (2018), "Visual Listening In: Extracting Brand Image Portrayed on Social Media," working paper.

Liu, Xuan, Savannah Wei Shi, Thales Teixeira, and Michel Wedel (2018), "Video Content Marketing: The Making of Clips," *Journal of Marketing*, 82 (4), 86–101.

Ljung, M. (2000), "Newspaper Genres and Newspaper English," in *English Media Texts Past and Present: Language and Textual Structure*, Friedrich Ungerer, ed. Philadelphia: John Benjamins Publishing, 131–50.

Longoni, Chiara, Andrea A. Bonezzi, and Carey K. Morewedge (2019), "Resistance to Medical Artificial Intelligence," *Journal of Consumer Research*, published online May 3, DOI: https://doi.org/10.1093/jcr/ucz013.

Loughran, Tim, and Bill McDonald (2016). "Textual Analysis in Accounting and Finance: A Survey," *Journal of Accounting Research*, 54 (4), 1187–1230.

Ludwig, Stephan, Ko De Ruyter, Mike Friedman, Elisabeth C. Bruggen, Martin Wetzels, and Gerard Pfann (2013), "More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates," *Journal of Marketing*, 77 (1), 87–103.

Ludwig, Stephan, Ko De Ruyter, Dominik Mahr, Elisabeth C. Bruggen, Martin Wetzels, and Tom De Ruyck (2014), "Take Their Word for It: The Symbolic Role of Linguistic Style Matches in User Communities," *MIS Quarterly*, 38 (4), 1201–17.

Ludwig, Stephan, Tom Van Laer, Ko De Ruyter, and Mike Friedman (2016), "Untangling a Web of Lies: Exploring Automated Detection of Deception in Computer-Mediated Communication," *Journal of Management Information Systems*, 33 (2), 511–41.

Ma, Liye, Sun Baohung, and Sunder Kekre (2015), "The Squeaky Wheel Gets the Grease—An Empirical Analysis of Customer Voice and Firm Intervention on Twitter," *Marketing Science*, 34 (5), 627–45.

Manchanda, Puneet, Grant Packard, and A. Pattabhitamaiah (2015), "Social Dollars: The Economic Impact of Consumer Participation in a Firm-Sponsored Online Community," *Marketing Science*, 34 (3), 367–87.

Mathwick, Charla, Caroline Wiertz, and Ko De Ruyter (2007), "Social Capital Production in a Virtual P3 Community," *Journal of Consumer Research*, 34 (6), 832–49.

McCarthy, Daniel, and Peter Fader (2018), "Customer-Based Corporate Valuation for Publicly Traded Noncontractual Firms," *Journal of Marketing Research*, 55 (5), 617–35.

McCombs, Maxwell E., and Donald L. Shaw (1972), "The Agenda-Setting Function of Mass Media," *Public Opinion Quarterly*, 36 (2), 176–87.

McCracken, Grant (1988), *Qualitative Research Methods: The Long Interview*. Newbury Park, CA: SAGE Publications.

McQuarrie, Edward F., Jessica Miller, and Barbara J. Phillips (2012), "The Megaphone Effect: Taste and Audience in Fashion Blogging," *Journal of Consumer Research*, 40 (1), 136–58.

Melumad, Shiri J., J. Jeffrey Inman, and Michael Tuan Pham (2019), "Selectively Emotional: How Smartphone Use Changes User-Generated Content," *Journal of Marketing Research*, 56 (2), 259–75.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean (2013), "Efficient Estimation of Word Representations in Vector Space," (accessed August 19, 2019), https://arxiv.org/abs/1301.3781.

Moe, Wendy W., and David A. Schweidel (2012), "Online Product Opinions: Incidence, Evaluation, and Evolution," *Marketing Science*, 31 (3), 372–86.

Moe, Wendy W., and Michael Trusov (2011), "The Value of Social Dynamics in Online Product Ratings Forums," *Journal of Marketing Research*, 48 (3), 444–56.

Mogilner, Cassie, Sepandar D. Kamvar, and Jennifer Aaker (2011), "The Shifting Meaning of Happiness," *Social Psychological and Personality Science*, 2 (4), 395–402.

Molner, Sven, Jaideep C. Prabhu, and Manjit S. Yadav (2019), "Lost in the Universe of Markets: Toward a Theory of Market Scoping for Early-Stage Technologies," *Journal of Marketing*, 83 (2), 37–61.

Moody, Christopher E. (2016), "Mixing Dirichlet Topic Models and Word Embeddings to Make lda2vec," (accessed August 19, 2019), https://arxiv.org/abs/1605.02019.

Moon, Sangkil, and Wagner A. Kamakura (2017), "A Picture Is Worth a Thousand Words: Translating Product Reviews into a Product Positioning Map," *International Journal of Research in Marketing*, 34 (1), 265–85.

Moorman, Christine, Harald J. van Heerde, C. Page Moreau, and Robert W. Palmatier (2019a), "Challenging the Boundaries of Marketing," *Journal of Marketing*, 83 (5), 1–4.

Moorman, Christine, Harald J. van Heerde, C. Page Moreau, and Robert W. Palmatier (2019b), " *JM* as a Marketplace of Ideas," *Journal of Marketing*, 83 (1), 1–7.

Muñiz, Albert, Jr., and Thomas O'Guinn (2001), "Brand Community," *Journal of Consumer Research*, 27 (4), 412–32.

Narayanan, Sridhar, Ramarao Desiraju, and Pradeep K. Chintagunta (2004), "Return on Investment Implications for Pharmaceutical Promotional Expenditures: The Role of Marketing-Mix Interactions," *Journal of Marketing*, 68 (4), 90–105.

Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko (2012), "Mine Your Own Business: Market-Structure Surveillance Through Text Mining," *Marketing Science*, 31 (3), 521–43.

Netzer, Oded, Alain Lemaire, and Michal Herzenstein (2019), "When Words Sweat: Identifying Signals for Loan Default in the Text of Loan Applications," *Journal of Marketing Research*, published electronically August 15, 2019, doi:10.1177/0022243719852959.

Nicas, Jack (2018), "Facebook Says Russian Firms 'Scraped' Data, Some for Facial Recognition," *The New York Times* (October 12), https://www.nytimes.com/2018/10/12/technology/facebook-russian-scraping-data.html.

Nisbett, Richard E., and Timothy D. Wilson (1977), "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review*, 84 (3), 231–59.

Opoku, Robert, Russell Abratt, and Leyland Pitt (2006), "Communicating Brand Personality: Are the Websites Doing the Talking for the Top South African Business Schools?" *Journal of Brand Management*, 14 (1/2), 20–39.

Ott, Myle, Claire Cardie, and Jeff Hancock (2012), "Estimating the Prevalence of Deception in Online Review Communities," in *Proceedings of the 21st International Conference on World Wide Web*. New York: Association for Computing Machinery, 201–10.

Packard, Grant, and Jonah Berger (2017), "How Language Shapes Word of Mouth's Impact," *Journal of Marketing Research*, 54 (4), 572–88.

Packard, Grant, and Jonah Berger (2019a), "How Concrete Language Shapes Customer Satisfaction," working paper.

Packard, Grant, and Jonah Berger (2019b), "Thinking of You: How Second Person Pronouns Shape Cultural Success," working paper.

Packard, Grant, Sarah G. Moore, and Brent McFerran (2018), "(I'm) Happy to Help (You): The Impact of Personal Pronoun Use in Customer–Firm Interactions," *Journal of Marketing Research*, 55 (4), 541–55.

Palmatier, Robert W., Rajiv P. Dant, and Dhruv Grewal (2007), "A Comparative Longitudinal Analysis of Theoretical Perspectives of Interorganizational Relationship Performance," *Journal of Marketing*, 71 (4), 172–94.

Peladeau, N. (2016), *WordStat: Content Analysis Module for SIMSTAT*. Montreal, Canada: Provalis Research.

Pennebaker, James W. (2011), "The Secret Life of Pronouns," *New Scientist*, 211 (2828), 42–45.

Pennebaker, James W., Roger J. Booth, Ryan L. Boyd, and Martha E. Francis (2015), *Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX: Pennebaker Conglomerates.

Pennebaker, James W., and Laura A. King (1999), "Linguistic Styles: Language Use as an Individual Difference," *Journal of Personality and Social Psychology*, 77 (6), 1296–1312.

Smith, Aaron, and Monica Anderson (2018), "Social Media Use in 2018," Pew Research Center (March 1), http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/.

Pollach, Irene (2012), "Taming Textual Data: The Contribution of Corpus Linguistics to Computer-Aided Text Analysis," *Organizational Research Methods*, 15 (2), 263–87.

Provost, Foster, Brian Dalessandro, Rod Hook, Xiaohan Zhang, and Alan Murray (2009), *"Audience Selection for On-Line Brand Advertising: Privacy-Friendly Social Network Targeting,"* in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: Association for Computing Machinery, 707–16.

Puranam, Dinesh, Vishal Narayan, and Vrinda Kadiyali (2017), "The Effect of Calorie Posting Regulation on Consumer Opinion: A Flexible Latent Dirichlet Allocation Model with Informative Priors," *Marketing Science*, 36 (5), 726–46.

Ransbotham, Sam, Nicholas Lurie, and Hongju Liu (2019), "Creation and Consumption of Mobile Word of Mouth: How Are Mobile Reviews Different?" *Marketing Science*, published online January 28, DOI: https://doi.org/10.1287/mksc.2018.1115.

Reagan, Andrew J., Lewis Mitchell, Dilan Kiley, Christopher M. Danforth, and Petter Sheridan Dodds (2016), "The Emotional Arcs of Stories Are Dominated by Six Basic Shapes," *EPJ Data Science*, 5 (1), 1–12.

Rocklage, Matthew D., and Russell H. Fazio (2015), "The Evaluative Lexicon: Adjective Use as a Means of Assessing and Distinguishing Attitude Valence, Extremity, and Emotionality," *Journal of Experimental Social Psychology*, 56 (January), 214–27.

Rocklage, Matthew D., Derek D. Rucker, and Loran F. Nordgren (2018), "The Evaluative Lexicon 2.0: The Measurement of Emotionality, Extremity, and Valence in Language," *Behavior Research Methods*, 50 (4), 1327–44.

Rodriguez, Salvador (2017), "U.S. Judge Says LinkedIn Cannot Block Startup from Public Profile Data," *Reuters* (August 14), https://

www.reuters.com/article/us-microsoft-linkedin-ruling/u-s-judge-says-linkedin-cannot-block-startup-from-public-profile-data-idUSKCN1AU2BV.

Rogers, Everett M. (1995). *Diffusion of Innovations*, 4th ed. New York: The Free Press.

Rosa, Jose Antonio, Joseph F. Porac, Jelena Runser-Spanjol, and Michael S. Saxon (1999), "Sociocognitive Dynamics in a Product Market," *Journal of Marketing*, 63 (Special Issue), 64–77.

Rude, Stephanie, Eva-Maria Gortner, and James Pennebaker (2004), "Language Use of Depressed and Depression-Vulnerable College Students," *Cognition & Emotion*, 18 (8), 1121–33.

Salganik, Matthew J., Peter Dodds, and Duncan Watts (2006), "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market," *Science*, 311 (5762), 854–56.

Scaraboto, Daiane, and Eileen Fischer (2012), "Frustrated Fashionistas: An Institutional Theory Perspective on Consumer Quests for Greater Choice in Mainstream Markets," *Journal of Consumer Research*, 39 (6), 1234–57.

Schoenmüller, Verena, Oded Netzer, and Florian Stahl (2019), "The Extreme Distribution of Online Reviews: Prevalence, Drivers and Implications," Columbia Business School Research Paper.

Schweidel, David A., and Wendy W. Moe (2014), "Listening in on Social Media: A Joint Model of Sentiment and Venue Format Choice," *Journal of Marketing Research*, 51 (4), 387–402.

Simonton, Dean Keith (1980), "Thematic Fame, Melodic Originality, and Musical Zeitgeist: A Biographical and Transhistorical Content Analysis," *Journal of Personality and Social Psychology*, 38 (6), 972–83.

Snefjella, Bryor, and Victor Kuperman (2015), "Concreteness and Psychological Distance in Natural Language Use," *Psychological Science*, 26 (9), 1449–60.

Srivastava, Sameer B., and Amir Goldberg (2017), "Language as a Window into Culture," *California Management Review*, 60 (1), 56–69.

Stewart, David W., and David H. Furse (1986), *TV Advertising: A Study of 1000 Commercials*. Waltham, MA: Lexington Books.

Tausczik, Yla R., and James.W. Pennebaker (2010), "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods," *Journal of Language and Social Psychology*, 29 (1), 24–54.

Tellis, Gerard J., Deborah J. MacInnis, Seshadri Tirunillai, and Yanwei Zhang (2019), "What Drives Virality (Sharing) of Online Digital Content? The Critical Role of Information, Emotion, and Brand Prominence," *Journal of Marketing*, 83 (4), 1–20.

Thompson, Craig J., and Gokcen Coskuner-Balli (2007), "Countervailing Market Responses to Corporate Co-Optation and the Ideological Recruitment of Consumption Communities," *Journal of Consumer Research*, 34 (2), 135–52.

Timoshenko, Artem, and John R. Hauser (2019), "Identifying Customer Needs from User-Generated Content," *Marketing Science*, 38 (1), 1–20.

Tirunillai, Seshadri, and Gerard J. Tellis (2012), "Does Chatter Really Matter? Dynamics of User-Generated Content and Stock Performance," *Marketing Science*, 31 (2), 198–215.

Tirunillai, Seshadri, and Gerard J. Tellis (2014), "Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of Big Data Using Latent Dirichlet Allocation," *Journal of Marketing Research*, 51 (4), 463–79.

Toubia, Olivier, Garud Iyengar, Renée Bunnell, and Alain Lemaire (2019), "Extracting Features of Entertainment Products: A Guided LDA Approach Informed by the Psychology of Media Consumption," *Journal of Marketing Research*, 56(1), 18–36.

Toubia, Olivier, and Oded Netzer (2017), "Idea Generation, Creativity, and Prototypicality," *Marketing Science*, 36 (1), 1–20.

Tsai, Jeanne L. (2007), "Ideal Affect: Cultural Causes and Behavioral Consequences," *Perspectives on Psychological Science*, 2 (3), 242–59.

Van Laer, Tom, Jennifer Edson Escalas, Stephan Ludwig, and Ellis A. Van den Hende (2018), "What Happens in Vegas Stays on TripAdvisor? Computerized Analysis of Narrativity in Online Consumer Reviews," *Journal of Consumer Research*, 46 (2), 267–85.

Van Zant, Alex B., and Jonah Berger (2019), "How the Voice Persuades," working paper, Rutgers University.

Villarroel Ordenes, Francisco, Dhruv Grewal, Stephan Ludwig, Ko De Ruyter, Dominik Mahr, and Martin Wetzels (2018), "Cutting Through Content Clutter: How Speech and Image Acts Drive Consumer Sharing of Social Media Brand Messages," *Journal of Consumer Research*, 45 (5), 988–1012.

Villarroel Ordenes, Francisco, Stephan Ludwig, Ko De Ruyter, Dhruv Grewal, and Martin Wetzels (2017), "Unveiling What Is Written in the Stars: Analyzing Explicit, Implicit, and Discourse Patterns of Sentiment in Social Media," *Journal of Consumer Research*, 43 (6), 875–94.

Vosoughi, Soroush, Deb Roy, and Sinan Aral (2018), "The Spread of True and False News Online," *Science*, 359 (6380), 1146–51.

Wang, Xin, Feng Mai, and Roger H.L. Chiang (2013), "Database Submission—Market Dynamics and User-Generated Content About Tablet Computers," *Marketing Science*, 33 (3), 449–58.

Wang, Yanwen, Michael Lewis, and David A. Schweidel (2018), "A Border Strategy Analysis of Ad Source and Message Tone in Senatorial Campaigns," *Marketing Science*, 37 (3), 333–55.

Weber, Klaus (2005), "A Toolkit for Analyzing Corporate Cultural Toolkits," *Poetics*, 33 (3/4), 227–52.

Wies, Simone, Arvid Oskar Ivar Hoffmann, Jaako Aspara, and Joost M.E. Pennings (2019), "Can Advertising Investments Counter the Negative Impact of Shareholder Complaints on Firm Value?" *Journal of Marketing*, 83 (4), 58–80.

Xiao, Li, Hye-Jin Kim, and Min Ding (2013), "An Introduction to Audio and Visual Research and Applications in Marketing," in *Review of Marketing Research*, Vol. 10, Naresh Malhotra, ed. Bingley, UK: Emerald Publishing, 213–53.

Xiong, Ying, Moonhee Cho, and Brandon Boatwright (2019), "Hashtag Activism and Message Frames Among Social Movement Organizations: Semantic Network Analysis and Thematic Analysis of Twitter During the #MeToo Movement," *Public Relations Review*, 45 (1), 10–23.

Yadav, Manjit S., Jaideep C. Prabhu, and Rajesh K. Chandy (2007), "Managing the Future: CEO Attention and Innovation Outcomes," *Journal of Marketing*, 71 (4), 84–101.

Ying, Yuanping, Fred Feinberg, and Michel Wedel (2006), "Leveraging Missing Ratings to Improve Online Recommendation Systems," *Journal of Marketing Research*, 43 (3), 355–65.

Zhong, Ning, and David A. Schweidel (2019), "Capturing Changes in Social Media Content: A Multiple Latent Changepoint Topic Model," working paper, Emory University.