

Improving the Language of Influence

D.B. Skillicorn
School of Computing
Queen's University
Kingston, Canada

C. Leuprecht
School of Policy Studies
Queen's University
Kingston, Canada

Abstract—We show that the language used by U.S. presidential candidates over the past twenty years has an underlying temporal structure associated with electoral success, with the most influential language used by incumbents in their second campaign and the least by losers in a first-cycle open campaign. Influential language is characterized by increased positivity, complete absence of negativity, increased abstraction, and lack of reference to the opposing candidate(s). The way in which language use changes suggests that it is the result of changing self-perception rather than a deliberate strategy. This has implications for the language of influence as deployed by violent extremist groups, suggesting that both success at convincing an audience to participate in violent extremism and the presence of competing groups trying to make similar arguments improve the quality of the influencing language they use.

I. INTRODUCTION

U.S. presidential election campaigns are the most well-funded, well-studied, and well-motivated attempts to convince a wide-spectrum audience to take a particular action. As such, they provide a template for increasing our understanding of the language of influence and, because the outcomes are apparent, its effectiveness [1, 6].

There are two widely-held channels of influence: rational choice (the best argument has the most influence); and rhetoric (the best interpretation of reality has the most influence) [2]. Rational argument founders when the audience collectively holds different world-views (general election campaigns) but may be more significant for a homogeneous audience (primary campaigns). However, rhetoric is generally thought to be the more significant channel. For example, rhetorical skill enables a candidate to present himself with a stronger and better persona than his real character, and to deal with the (modern) problem of multiple audiences, the proximate and the universal, who must simultaneously be influenced. This leads naturally to strategic ambiguity as a way to expand appeal [10].

In this paper, we examine the language patterns used in U.S. presidential elections from 1992 to 2012. We show that, behind the language used by challengers and

incumbents, and successful and unsuccessful candidates, there is a linear scale of linguistic influence. In any given campaign, the candidate whose language is higher on the scale wins; presidents campaigning for re-election use more influential language in their re-election campaign than they did in their first campaign; and the gap between winner and loser tends to remain roughly constant, suggesting that a challenger to an incumbent is drawn into using more influential language than that of the losing candidate in an open race. The increase in strength of influence happens quite quickly, either at the beginning of the first term or at the beginning of the second campaign, suggesting that it is driven by changes in self-perception rather than deliberate strategy or increasing experience. Success breeds success for winning candidates, but also raises the game of challengers.

In an intelligence setting, this implies that violent extremist groups, who are also trying to convince others to act, can and probably will get better at it with experience. This may require exogenous drivers of two kinds: success itself, which increases self-perceived significance, playing back into more powerful language; and competition in the influence marketplace which raises the quality for all participants. We compare presidential campaign speech language with the language of *Inspire* magazine [4, 9], which aims to influence an overlapping demographic using a similar set of channels. At present, *Inspire*'s language patterns lie between first and second campaign quality, and have been static over the several years that the magazine has been produced. Our results suggest that significant success with the magazine's goal of recruiting lone-wolf extremists might cause the quality of influence language to improve rapidly, as might the appearance of a similar channel from another source.

II. EXPERIMENTAL METHOD

We collected campaign speeches for the U.S. presidential election cycles from 1992 for the candidates shown in Table I. In each case speeches only in the campaign year (that is, from January until November)

Year	Winner	Loser
1992	Clinton	
1996	Clinton	Dole
2000	Bush	Gore
2004	Bush	Kerry
2008	Obama	McCain
2012	Obama	Romney

TABLE I
CONTENDERS IN PRESIDENTIAL ELECTION CYCLES

were included; a random selection that included at least one from each month was used where possible.

The frequencies of each word used in each speech were extracted using a parts-of-speech aware tool. This resulted in a matrix with one row per speech, one column per word (with its associated part-of-speech tag), and entries that are the frequencies of each word in each document. The entries in each row were divided by the row sum to normalize for different speech lengths, turning word frequencies into word rates. For technical reasons, the entries in each column were then normalized to z-scores, that is the column mean was subtracted from each entry and the entry divided by the column standard deviation. The effect of z-scoring is to center the data in each column around zero, and make the spreads of different columns roughly comparable.

Each row of this matrix can be interpreted as a location in a vector space spanned by the columns, which correspond to the words. In this space, two documents that are similar, in the sense that they contain approximately the same words with approximately the same frequencies, will be placed at locations that are close. The dimensionality of this space is determined by the number of distinct words, which is large; but most words appear in only a few documents and so play little role in determining document-document similarity. There is little loss of accuracy, and much gain in clarity, in projecting this space into one of much lower dimensionality. In what follows, we will project the space into only two dimensions, making it possible to render it for inspection. The best choice for the axes depends on the structure of the data. This can be determined by using a Singular Value Decomposition (SVD) [5] which transforms the data into coordinates with respect to axes along which the uncorrelated variation is greatest.

We extracted several matrices from the speech data, using different subsets of possible markers: the 5000 most frequent words overall, nouns, verbs, adjectives, adverbs, pronouns, and determiners. For each of these categories, two-dimensional plots of the speeches were examined.

Also, for each word in each of the categories, the change in its rate of use from first to second campaigns

was calculated. This process was done by hand; it is difficult to automate because the observed frequencies are a mixture of functional (but almost certainly subconscious) choices and simple linguistic habits. For example, the word “great” was used at moderate levels by Bush in both campaigns, but shows a marked increase in use in second campaigns for the other candidates. Therefore, we consider it to have increased.

III. DATA AND RESULTS

We project the data matrix corresponding to all speeches and a particular set of words into two dimensions and show the positions corresponding to each speech as a point. By labelling the points with properties of the corresponding speeches (cycle, party, outcome and so on), we can see immediately whether there are higher-level patterns. We show the plot only for all frequent words, but the discovered structure is the same for all other word classes except verbs, which have a much more disparate structure.

A. Frequent words

Figure 1 shows the variation among speeches over all campaigns based on the 5000 overall most frequent words. Winning candidates give speeches towards the upper left end of the figure, and tend to move further towards that style or pattern with time. Losing candidates give speeches towards the lower right end of the figure. Thus the spectrum from lower right to upper left defines the scale of influential language.

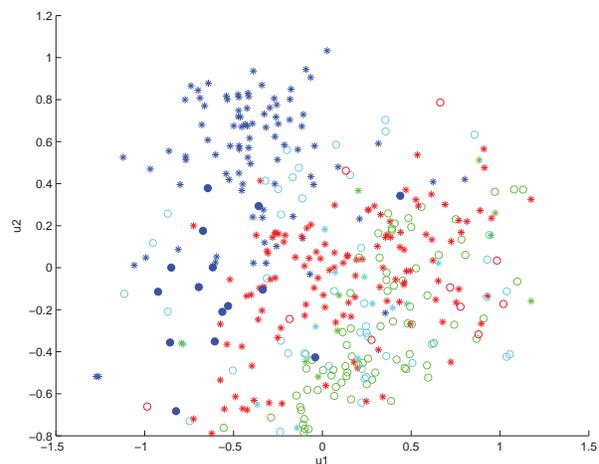


Fig. 1. Language use for the 5000 most frequent words: red: successful in first campaign; blue: successful in re-election campaign; green: failed in first campaign; cyan: failed against an incumbent; circles: Republicans; stars: Democrats. Successful re-election campaigns (blue) use different language from successful first campaigns (red), but seem to cause related changes in the language of challengers as well (cyan).

The mean language used in all four settings is shown in Figure 2 on the same scale as Figure 1. The points represent the average position of the speeches corresponding to the four different categories: winners of second terms, losers against an incumbent, winners of first terms, and losers in a first-term contest. There is a linear progression from lower right to upper left, with incumbent winners at the top and open-contest losers at the lower right. Interestingly, the mean positions for Bush and Gore in their 2000 campaigns (not shown) are almost exactly the same, reinforcing the closeness of that election. The gap between winners and losers increases somewhat in the second campaign cycle.

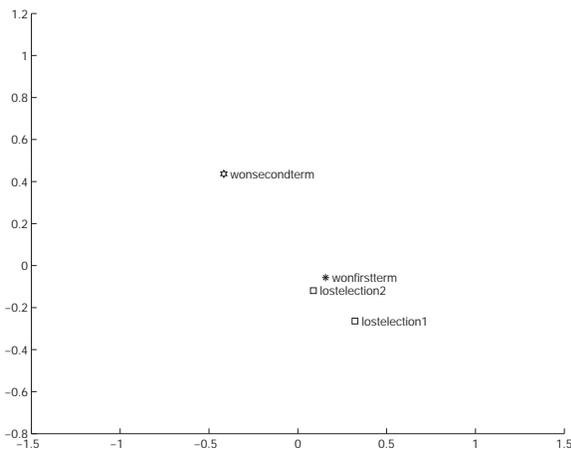


Fig. 2. Mean language pattern for those who won and lost in both cycles. Winning incumbents always change their language from their first campaign; their competitors seem to adapt to match. Relative language patterns are predictive of success across all of these campaigns.

A further illustration of the change between first and second campaign for successful candidates is shown in Figure 3. There are systematic variations in the choice of words from among the 500 most frequent, and this discrepancy continues further down the ranking. Second campaigns use common words at lower rates than first campaign; as a result, they tend to use a wider range of words.

In Figure 1 there is considerable variation, even from day to day, in the language patterns of a given candidate. This variation is orthogonal to that of influential language, that is from lower left to upper right in the figure. This variation derives from the number of different words used in each speech. In most documents, length is not strongly correlated with number of different words present; for example, a document might cover a number of different topics but might also be a detailed discussion of a single topic. This turns out not to be the case for political speeches. Longer speeches seem always to be used to cover more topics, all at about the same level

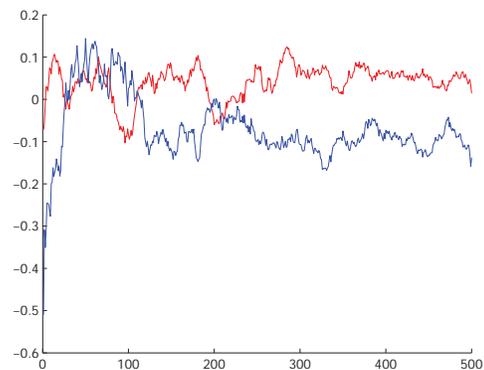


Fig. 3. Differences in the normalized frequencies of the top 500 words by rank: red: first campaign; blue: second campaign

of detail. Since campaign speeches tend to occur in a few conventional lengths – fundraisers, campaign stump speeches, or extended (televised) speeches – and length is strongly correlated with the number of different words used, there is considerable short-term variation.

B. Modelling the Difference

There is clearly a systematic difference between first and second campaign cycle speeches. We now examine the language patterns that drive this systematic difference, which is also connected to the difference between success and failure.

In almost every case, the same speechwriting team was involved in both campaigns, so the observed changes are not the result of changes in personnel leading to changes in style. (It is, in any case, unlikely that personnel changes over twenty years would produce such consistent language changes.)

Table II shows the high-frequency words in several categories that increase or decrease between first and second campaigns for successful candidates. Observing the words in each category suggests a higher-level view of how the language patterns are changing.

Overall, the differences of second-campaign language are characterized by:

- An increase in words with positive connotations;
- The complete absence of words with negative connotations, even to the extent of reduced use of contractions with an implicit negative such as “don’t”;
- An increase in abstract words and a decrease in concrete words, both in the domain of nouns (“life” rather than “policies”) and determiners (“some” rather than “this”);

There is also a decrease in the rate of use of “he” suggesting a reduced focus on opponents (who were all

Freq words	Increase	to, government, good, way, great, thank, security, which, any, as, congress, well, its, life, am, best, most, percent, right (adj.), better, important, power, better, freedom, free, find, growth, friends, raise, progress, wants
	Decrease	that, this, what, or, that's, if, because, up, it's, so, go, like, give, why, others, together, can't, down, same, over, Washington, you're, companies, four, election, change, we'll, about, right, college, class, folks, another, young, won't, I'll, everybody, stand, change, few, policies, billion, while, means, invest, how, provide, less, end, fighting, finally, start, months, willing, giving, old, once, no, hard
Nouns	Increase	world, government, security, life, state, power, freedom, growth, friends, progress, ways, success, others, strength, choices
	Decrease	time, future, Washington, thing, millions, reason, peace, medicare, months, challenge
Verbs	Increase	is, are, be, do, doing, lost
	Decrease	don't, didn't, doesn't, giving, wants, isn't, breaks, struggling, investing, began
Adjectives	Increase	good, great, best, right, better, important, free, greater, personal, corporate, such, outstanding, low, individual
	Decrease	military, willing, old, common, main, dangerous, bankrupt, ordinary, immediate
Adverbs	Increase	even, well, only, always, better, often
	Decrease	now, up, down, still, about, instead, finally, actually, maybe, harder, far
Pronouns	Increase	them, its
	Decrease	it's, we'll, everybody, I'll, you'd
Determiners	Increase	some, many, other, any
	Decrease	this, that, that's, last, same, few, less

TABLE II
WORDS CHANGE FROM FIRST TO SECOND SUCCESSFUL CAMPAIGNS, BY WORD TYPE AND IN DECREASING ORDER OF OVERALL FREQUENCY.

male in these election cycles). Overall, second campaign language is more variable than first campaign language.

These differences do not seem to have been deliberately exploited by speechwriters. First, if these changes had been made consciously, we would expect to see a steady progression *across* election cycles, whereas what we observe is a fresh start for each new candidate. Second, the changes in language happen relatively rapidly. For example, Obama uses consistent language patterns in his first campaign and throughout his first term, changing to the new pattern within the space of a few months in late 2011 and early 2012 [8]. Thus it seems unlikely that speechwriting teams become incrementally better at their craft or grope towards this dimly-perceived influential language.

This suggests that the changes we describe happen because of a change in self-perception by the candidate and campaign, the most obvious change being a track record of election (and so influence) success [11]. From this perspective, the increase in positive language results from increased confidence; and the increase in abstract language from a sense of being above the fray, and a qualitatively different kind of contender from the opponent.

The differences in Table II match a generally agreed view of influence – that it is better to attract because of positive qualities than either to paint opponents negatively or attempt to influence by reasoning [3, 13]. What

is surprising is how difficult it seems to be to convince potential influencers that this strategy is the right one; or perhaps it shows how strong the forces that pull towards negativity and argument, even rational argument, are. Nevertheless, Table II is a recipe for effective influence. Not all of the differences are easy to exploit: changing function word usage is difficult, because so much of their use is subconscious. Speechwriters can tell a candidate not to say “don’t” and leave it off the teleprompter, but this will not necessarily prevent its appearance.

C. Content

It is noteworthy that there are few content words at the top of the frequent word list, and those that are present are quite generic. Of course, different election cycles focus on different topics, but here we examine the use of nouns as surrogates for content.

Figure 4 shows the spectrum of variation among the nouns. They form a rough triangle with national life nouns at the upper right, economic nouns at the lower right, and more-generic nouns at the left. This suggests that most topics are actually perennials that are talked about in most election cycles. There are differences: the economic pole of this triangle was much less in play in the 2012 election cycle than in the 2008 one, for example.

Overall, noun usage aligns with the language of influence, but only in the sense that influential language is

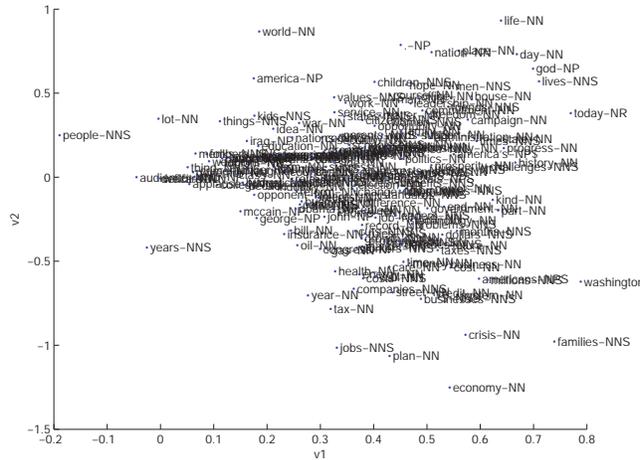


Fig. 4. Variation in noun use aligned with the variation in speeches. associated with the left end of the triangle – the more generic words. In other words, focusing on content, even patriotic content, tends to lower influence.

IV. JIHADIST INFLUENCE

We now relate these results about language of influence in politics to influence in another domain – violent extremism.

Inspire magazine is a pdf jihadist magazine, written in English, and made widely available on the Internet. Eleven issues have appeared, the first in the middle of 2010 and the eleventh in mid-2013. The first eight issues were edited by Samir Khan, with substantial assistance from the charismatic preacher, Anwar al-Awlaki. After their deaths in 2011, Issue 8 and Issue 9 appeared close together in early 2012. Issue 8 had evidently been largely written before their deaths [12]; Issues 9, 10, and 11 are the product of three different, so far unknown, editors.

Figure 5 shows the variations in the language for both successful first and second campaigns, and issues of *Inspire* magazine. The direction of improving quality runs from the lower right to the upper left, with the distinction between first and second campaign language quite obvious.

The issues of *Inspire* magazine use language that lies between that of successful first and second campaigns. There are some strong idiosyncrasies that cause the horizontal displacement. The issues of *Inspire* use three words – “which”, “its” and “such” – at much higher rates than the campaign speeches. These words have little content, either factual or emotional, and so appear to represent simple differences in style. For example, “which” is used correctly in the magazine to introduce optional dependent clauses, but is vanishingly rare in speeches by Bush and Obama, and only slightly more common

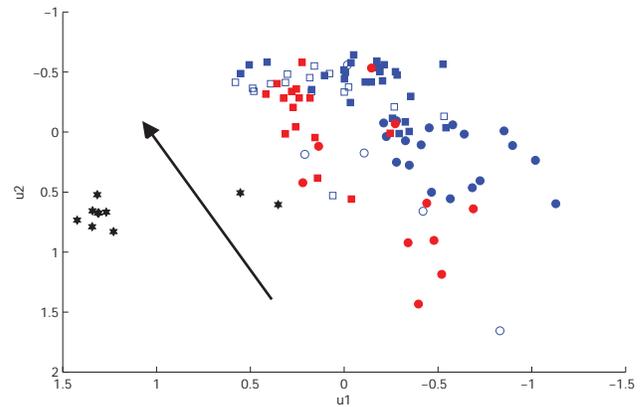


Fig. 5. Successful first and second campaign speeches versus Inspire Magazine with arrow indicating direction of increasingly influential language. Red – Bush, solid blue – Obama, open blue – Clinton, black – Inspire; circles – first campaign, squares – second campaign

in those by Clinton. It is plausible that speechwriters are trained to remove such optional clauses in political speeches. Similarly “its” occurs at high rates in the magazine because it tends to reference abstract nouns heavily (“America”, “Western world”, “globalisation”), another habit that is probably discouraged for speechwriters.

The consistent differences between *Inspire* and campaign speeches with regard to levels of positive language are:

- Used at lower rates: “to”, “good”, “well”, “do”;
- Used at rates similar to first campaigns: “great”, “best”, “right” (adj.), “better”, “doing”, “some”;
- Used at rates similar to second campaigns: “most”;
- Used at rates above campaigns: “any”, “life”, “is”, “be”, “even”.

Thus *Inspire* fails to be as positive as second-campaign speeches primarily because it does not use positive adjectives nearly as much and because it uses fewer short action verbs and directional prepositions.

Another way to consider content is to see the extent to which campaigns use language considered as conservative or liberal. We use language models extracted from [7].

Figure 6 shows the variation across campaign speeches and *Inspire* using liberal language. The variation from left to right is almost entirely driven by the use of “health” and “care” which are associated with Obama’s 2008 campaign, but not much with others. The vertical dimension shows that variation in intensity of liberal language, from top to bottom. As expected, *Inspire* has very low levels of such language. There is, however, very little differentiation between Republican and Democratic presidential campaigns; both Bush and

Obama actually reduce their use of liberal language in their second campaigns.

The Lakoff model for conservative language contains many words with negative connotations. While they might appear in content-filled policy discussions, they are hardly present in campaign speeches (with the exception of “freedom” and “strong”) and there is little differentiation between speeches.

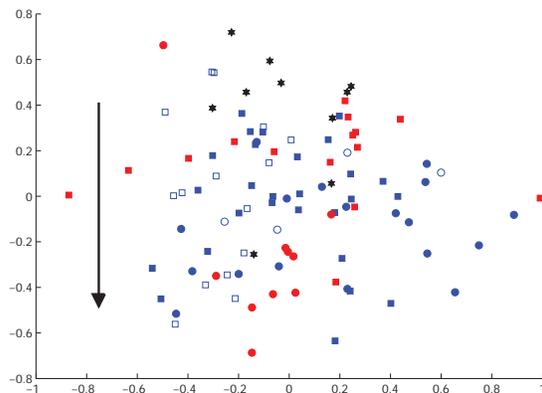


Fig. 6. Variation in speeches based on a model of liberal language. Red – Bush, solid blue – Obama, open blue – Clinton, black – Inspire; circles – first campaign, squares – second campaign

V. DISCUSSION AND CONCLUSIONS

Patterns of word usage are predictive of success in both first and second cycle election campaigns. Incumbents are able to move towards more successful patterns in their second campaigns; but they seem to cause their opponents also to change their language patterns in response.

The changes that characterize this improved language have the following properties:

- It does not depend on changes in content words, so it is not a function of improved arguments or even of improved rhetoric as usually understood. Winners do not have better ideas or present them more cleverly.
- The greatest variation is always caused by variations in the small words: function words, but also, for nouns, short nouns, for verbs, short and auxiliary verbs, and for adverbs, short and common adverbs.
- Words with negative connotations decrease dramatically: explicitly negative adjectives and adverbs, but also less obvious negative words such as contractions containing “not” (“didn’t”), some nouns (“challenge”), and some adverbs (“harder”).
- Words with positive connotations are used at high rates, especially positive adjectives (“good”,

“great”).

- Less attention paid to competitors, reflected in reductions in the rate of third-person pronouns.
- There is more variability in word choices overall.

Nouns vary the least of all parts of speech, presumably because the same issues surface and must be talked about in all election campaign. There tends to be a decrease in the use of economic nouns in second campaigns.

Campaign speeches show that the language of influence improves, although presumably primarily as the result of status changes rather than deliberate effort. Thus channels such as *Inspire*, which exhibit quite stable levels of influencing language, may strengthen if they have success convincing lone-wolf terrorists to carry out attacks in greater numbers than they have so far.

The language of influence also has an effect on competitors. Improvement in one drives matching improvement in competitors. Thus the analysis of the impact of a channel such as *Inspire* depends to some extent on the other channels with which it competes. (A new competitor, *Azan* magazine, has produced two issues so far. It will be interesting to see if this competition has an affect on the language of either publication.)

REFERENCES

- [1] W.L. Benoit. *Communication in Political Campaigns*. Peter Lang, 2007.
- [2] S. Condor, C. Tileagă, and M. Billig. Political rhetoric. In *Oxford Handbook of Political Psychology*. Oxford University Press, 2013.
- [3] J.M. Druckman, M.J. Kifer, and M. Parkin. Campaign communications in U.S. congressional elections. *American Political Science Review*, 103(3):343–366, 2009.
- [4] S.A. Ford. Inspiring a narrative: A content analysis of al Qaeda’s english-language *Inspire* magazine. Master’s Research Essay, Carleton University, 2012.
- [5] G.H. Golub and C.F. van Loan. *Matrix Computations*. Johns Hopkins University Press, 3rd edition, 1996.
- [6] K.L. Hacker, W.R. Zakahi, M.J. Giles, and S. McQuitty. Components of candidate images: Statistical analysis of the issue-persona dichotomy in the presidential campaign of 1996. *Communication Monographs*, 67(3):227–238, 2000.
- [7] G. Lakoff. *Moral Politics: How Liberals and Conservatives Think*. University of Chicago Press, 2nd edition, 2002.
- [8] J. Olson, Y. Ouyang, J. Poe, A. Trantham, and R.W. Waterman. The teleprompter presidency: Comparing Obama’s campaign and governing rhetoric. *Social Science Quarterly*, 93(5):1404–1423, 2012.
- [9] G. Ramsay. Targeting, rhetoric and the failure of grassroots jihad. *Journal of Terrorism Research*, 3(1), 2012.
- [10] K. Shepsle. The strategy of ambiguity: Uncertainty and electoral competition. *The American Political Science Review*, 66(2):555–568, June 1972.
- [11] D.B. Skillicorn and C. Leuprecht. The mental state of influencers. In *Foundations of Open-Source Intelligence FOSINT 2012*, 2012.
- [12] D.B. Skillicorn and E.F. Reid. Language use in *Inspire* magazine. In *IEEE International Conference on Intelligence and Security Informatics*, pages 239–244, 2013.
- [13] D. Stevens. Tone versus information: Explaining the impact of negative political advertising. *Journal of Political Marketing*, 11(4):322–352, 2012.