



CHIEF
MARKETING
TECHNOLOG
OFFICER
UNIVERSITY

draperBot theory | **whitepaper**

20 September 2015

Sapient
CMTO University

draperBot theory | **whitepaper**

September 20, 2015

Table of Contents

1	why draperBot?	1
2	Making Sense Of It	3
	2.1 The Hypothesis	3
	2.2 Wait, Semiotics?	3
	2.2.1 Semantics	3
	2.2.2 Syntactics	3
	2.2.3 Pragmatics	4
	2.3 The Adjusted (For Reality) Hypothesis	4
3	State of The Bot	5
	3.1 Bringing Graphing The Hypothesis To Life	5
	3.2 Graphing For Gold	6
	3.3 Now, Let's Build Some Brand Profiles	9
	3.4 Bringing the Data Together	10
4	Tomorrow's Journey	12
	4.1 What Does This All Mean?	12
	4.2 Open Source Thinking	12
	4.3 A Framework For More Meaning	12
	4.4 TODO?	12
Appendix		13
	1. Data Tables	13
Bibliography		16



1 why draperBot?

Before we answer that, let's talk globalization for sec...

Modern globalization has been a macro-economic trend that has been developing for decades since the end of World War II in the 1940s starting with the introduction of the multilateral trade agreement, General Agreement on Tariffs and Trade (GATT)^[1]. This agreement and its child organization that formed in 1995, the World Trade Organization (WTO), have been paving the way for trade and commerce across borders to expand for decades now. With such barriers to international trade having progressively opened, many businesses spread their wings in the 1960s and 1970s to grow into what we now call multinational firms, firms that operate smaller clones of their organizations in multiple countries. Companies viewed the ability to spread into more territories almost in the same vein that the European explorers did back in the 15th and 16th centuries. Despite this trend to expand for expansion-sake, some firms, typically those with a more consumer focus instead decided to pioneer solutions around the then uncharted space of global brand development and marketing. Once low cost communication networks and the formation of widespread internet capabilities came to be in the 1990s, more businesses sought to globalize their business and their operations than ever before.

In today's world when companies decide to enter new markets, they are accompanied by an army of business consultants wielding field-tested business models that quantify the potential profits, losses, PR impacts, talent pool challenges, political stability and other such factors entry into a new market might have on the business. While these models provide great insight into understanding whether an emerging market makes logical business sense, they often do not factor the more amorphous aspects of business, in particular consumer behaviour, into the math. But how can the nature of consumer behaviour be quantified into a math equation when each person's emotional reaction to brand interaction is different? Whether by geography, demographic, socio-economic grouping, there is no single answer to the challenge of understanding how consumers will react to a brand or a branded experience. In fact, with this uphill challenge of understanding the consumer, many organizations simply operate as though "global consumers", a consumer that marketers can reach beyond the borders of a country or territory, don't really exist and that they are better off marketing to those in a localized methodology, effectively customizing brands, branded products, packaging, and in some cases the products themselves to cater to the needs of consumers in each market. There is logic to this strategy, but ignoring that there is global consumer and how that influences marketing and branding decisions should be considered.

But however compelling it is to learn about the history of globalization and the modern day narrative of how brands and consumers interact, it is important to understand why this matters. There are troves of papers written on marketing in the global era, so what is special about draperBot; why should you continue reading these words?

The true goal of this project is for us to devise a quantitative framework for understanding consumers regardless of their homeland and overcome some of the marketing challenges caused by the 20th century globalization. For this to become a reality though we must truly understand the 1:1 relationship consumers have with products. But today's most advanced marketing technologies that attempt to solve this challenge are based on tracking one's behavior, and classifying individuals based on this behavior. This is inherently good and bad—It means that our privacy, in the eyes of marketers, is not important. In effect, the demand to sell more more product causes marketers to think the best solution is just to watch our every move, our every step. While the potential insight this can give marketers is huge, perhaps there are other ways that we can strive for this success and draperBot's aim is create a potential alternative. So what is draperBot then?

What if we could build a framework that catalogs the underlying meaning of our marketing communication, messaging, symbols and words? Then instead of storing a behavioral log of individuals, we develop a more

privacy-friendly personality index that combines together with this catalog of potential meaning/significance would give marketers a data-driven predictive tool guiding their choices without the potential big-brother tactics being employed today. Currently companies are storing massive amounts of personal behavior information and whether this information is stolen by hackers for identity theft or used against them like in the case of Chinese citizens, who voicemails have been used to convict them of sedition, the simple fact is that we as marketers still don't seem know that much about our audience. Understanding the impact our messaging or communication will have all across the globe will enable us as marketers to know more than recording the last 100 websites someone has visited. Its almost a case of technology outpacing the people using it, in that we have the ability to track and store behavior, but we still don't know what significance our advertising will have in 40 different countries until after we have spend money on the media.

So draperBot's aim is to become a collection of data-driven methods and tools, really a framework that can understand the meaning of our messaging and provide marketers and maybe, given its open-source nature, others [*hint: I'm thinking of people who might build Star Trek communicators*] with the capability of understanding what will happen when they send their messages, before they send them.

2 Making Sense Of It

2.1 The Hypothesis

So if the goal is to understand the meaning of our communication to the individuals living all across the world, then the hypothesis should be something like this:

<hypothesis>

By using a holistic framework of multi-cultural semiotics we can effectively understand the emotional connection individuals living in all parts of the world would, could and/or should have with brands.

</hypothesis>

2.2 Wait, Semiotics?

You might be asking yourself, what is **semiotics**? Well I certainly invite you to visit the multiple sources of great information on the subject, such as this great [Wikipedia](#) article, but suffice to say it is basically the study of meaning-making. And meaning-making is what we need if we are to set out to understand the impact of our words, symbols, communication in a predictive multi-cultural sense.

So let's unpack semiotics a little before we continue; It is comprised of multiple key areas of study: **semantics**, **syntactics** and **pragmatics**.

2.2.1 Semantics

We've likely all heard someone tell us "That's just semantics...", suggesting that someone is being persnickety about word selection, but semantics actually refers to relation between words and their significance to us, not someone being overly picky about similar words. So next time someone tells you that, you can say "yes, *I actually care what my words mean...*"

Semantics is a large area of study attempting to find the multiple meanings that words have and also their different meanings across different cultures. Simple examples of common semantics are words with multiple meanings like the word crash; 'Crash' can mean auto accident, a drop in the Stock Market, to attend a party without being invited, ocean waves hitting the shore or the sound of a cymbals being struck together. Other common examples are words that have no defined or "real" meaning, like in the phrase "Go for the *gusto*".

So in addition to handling basic definitions and meanings of words, to truly tackle semantics requires understanding context enough to recognize the implied word meaning being used in a particular instant, as well as understand fringe or slang words and their meanings as well.

2.2.2 Syntactics

Syntactics is about the rules we use to arrange our words. These rules can often impact the implied meaning that a set of words brings. Obviously if we mix our words up it can make a sentence sense not make... okay, I did that just to check if you were still reading, but the point here is that word order can make or break the meaning our communication has.

While a broken sentence may be the worst case, changing the order (but not making a phrase nonsensical) can significantly alter the potential meaning that a phrase has to recipient. Again, Wikipedia has devoted an entire article to this, entitled [Syntactic Ambiguity](#), although they also refer to it as amphibology.^[2] Here is one of the examples they reference as demonstrating this linguistic phenomenon:

Flying planes can be dangerous.

Either the act of flying planes is dangerous, or planes that are flying are dangerous.

Clearly understanding how we put words together to form meaning is critical to understanding how recipients of that message will react emotionally.

2.2.3 Pragmatics

Pragmatics is the most complex of the sub-studies within semiotics. Its primary objective is to understand the significance context has in the meaning of words and phrases. Pragmatics in fact pulls together multiple disciplines including speech act theory, conversational implicature, talk in interaction, and other language behaviour theories within philosophy, sociology and anthropology.^[3] The focus areas of pragmatics are the context of the speech (or an utterance as it is often referred to), the intent of the speaker, and any pre-existing knowledge of those involved.

Each discipline involved, such as speech act theory, is its own deep theoretical study with published theories and approaches, making the broader study of pragmatics as a parent entity a lesser explored area. In fact, only one published framework for pragmatics exists from Carlo Dalla Pozza who was an Italian philosopher of science and linguistics.[4] His theory suggests that it is possible to connect semantics with contextual intuition or intent to understand meaning. In addition to Pozza's theory, there is a large volume of research written surrounding speech act theory that informs on the direction that pragmatics as whole is heading.

So suffice it say that the given information and level of research available in the area of pragmatics is developing, and that incorporating it into draperBot might be premature at this time (2015, in you are reading this in the future). Nevertheless, once pragmatics reaches a place of practical maturity, it will be a significant addition to the tools and methods offered in draperBot.

2.3 The Adjusted (For Reality) Hypothesis

So while our ultimate goal here is to understand all meaning aspects of our communication, across multiple languages, as marketers (and maybe others too), through the use of these various branches of semiotics, that is not going to be the first milestone. We will set a more achievable first milestone of trying to first of all build some of the core framework needed to connect these bits of information together and second attacking a small subset of our semiotic Everest mountain of work. So for today, the adjusted hypothesis is:

<hypothesis>

By using a shared semantic framework, we can more effectively understand the emotional connection individuals have with brands.

</hypothesis>

There is another implied nature to this goal, which is that the focus is building this framework to work for any language, but that English will be the first one we test. So where do we begin?

Well in 2010, Josh Kaufman published the book, *The Personal MBA: Master The Art Of Business*, which details out 5 core human drivers.

- Drive to Acquire
- Drive to Bond
- Drive to Learn
- Drive to Defend
- Drive to Feel

If we take Josh Kaufman's drivers and leverage intelligence about people living in different parts of the world, with the way in which brands present their interactions to consumers, then perhaps we can understand if there is in fact a truly global consumer that marketers can reach. While this is limited intelligence by the standards set from our original hypothesis, it still can help us understand how brands connect to the consumer through data-driven methods. Ultimately predicting the potential intent consumers have with a brand could result in businesses not only understanding whether they should enter a new market from a product demand perspective, but how they market their products to the consumers that are in that market.

3 State of The **Bot**

3.1 Bringing Graphing The Hypothesis To Life

So what does a shared semantic framework look like? The 5 core human drivers mentioned in section 2.3 could be used as a shared framework if we connect both brands and a global audience together around them. But the simple fact here is that if you google “core human drivers”, you will find that there are multiple such frameworks that claim to be the quintessential framework for our human needs. Theories like Maslow’s hierarchy of needs, have actually been published and spread for decades, but when you examine the needs listed in Maslow’s theory there are multiple banal needs, like breathing and sleeping, that don’t carry significance for *all* brands and marketers. So this has led to two conclusions around the topic of frameworks. (1)—**Any framework built into draperBot should be a variable that can changed later by the user**; (2)—For the purposes of making short-term success with the adjusted hypothesis, we will assume that the 5 core human drivers Josh Kaufman’s book mentioned will suffice, changing only the last driver from “feel” to “experience” for the purposes of clarity. So the 5 core drivers or human desires being used by draperBot are:

1. Learn
2. Bond
3. Defend
4. Experience
5. Acquire

But 5 words alone will not be able to quantify how a brand is communicating, so the first technical aspect of draperBot is to extend these 5 core drivers using a node concept as outlined in **Figure 1** into something more meaningful.

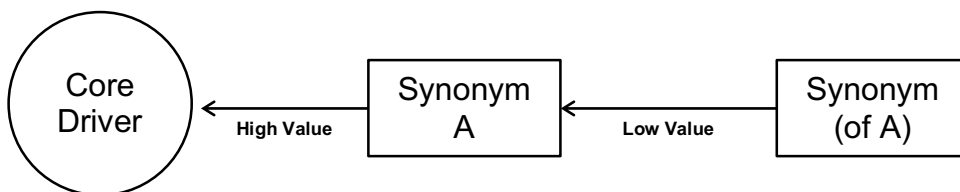


Figure 1

This concept can be implemented easily using a graph database, so draperBot leverages the open sourced graph database, Neo4j, to do just this. And while manually finding and recording the synonyms for 5 words would be reasonable, the goal here is to develop a recursive semantic framework to tease out the more subtle semantic meaning associations between words. With this in mind, a programmatic approach has been adopted using Merriam-Webster’s Dictionary.com API service, which includes a thesaurus entry. By recursively loading synonyms of synonyms of the 5 core desires, a word graph comprised of 4,491 and 22,120 relationships (of varying value) can be built as illustrated in **Figure 2**. Variations in this base word intelligence could be adjusted through adjusting the interpretation of the synonyms coming in from the API service. The supporting source code needed to operate this recursive synonym sequence is available in *build_initial_graph.js* and as was mentioned above—it is worth noting that the core driver framework can be changed in this file to support any other suitable emotional framework the software user sees fit to leverage.

This base word intelligence graph provides a bridge to synthesis other data inputs. For example, to better understand the language that brands use in their communications and marketing, one could take words used in their marketing channels as a data input and compare them to words in the graph for total relationship value to see emotional relativity of the communication. In addition, to understand how an individual’s personal proclivity to

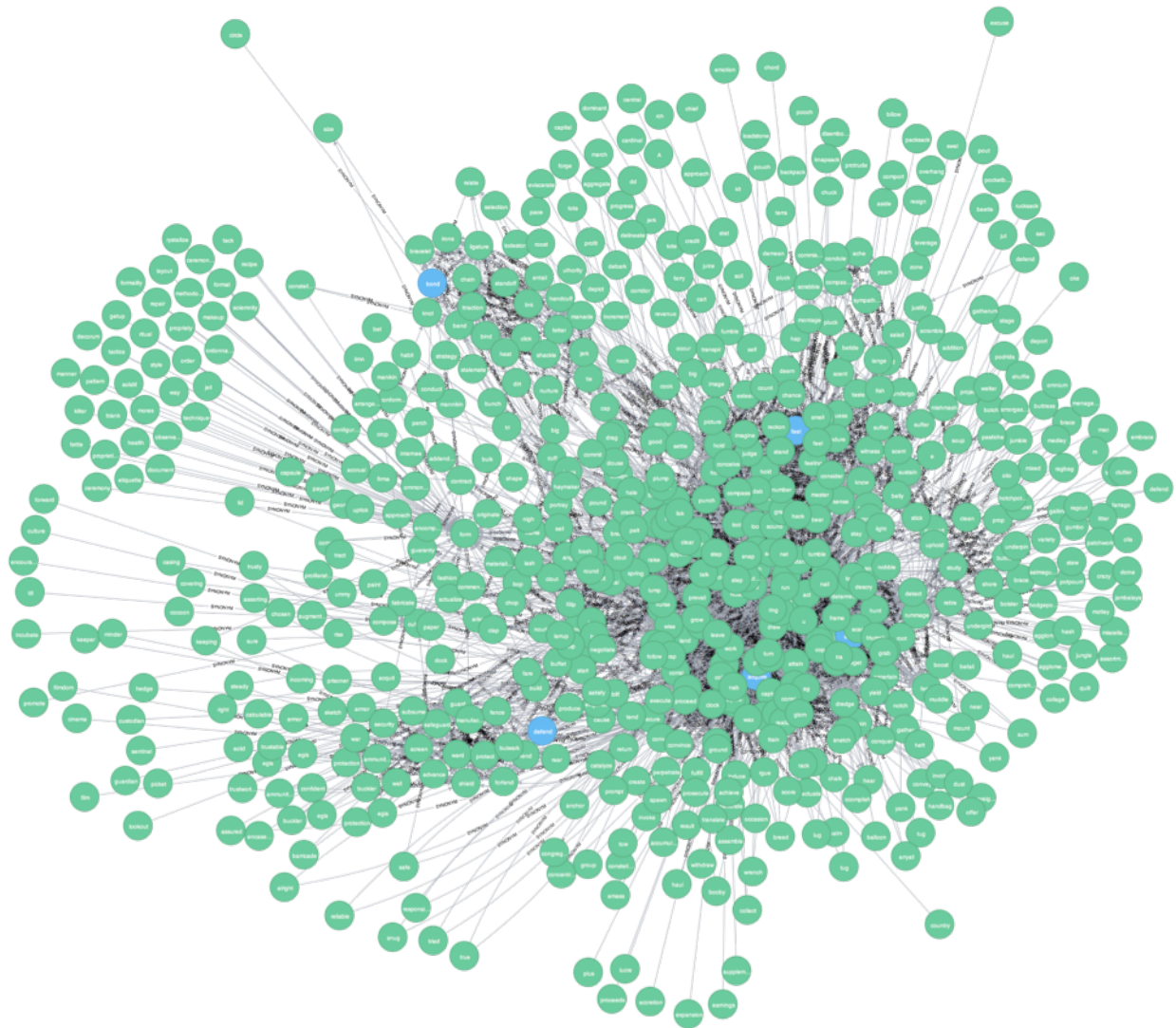


Figure 2

these emotional core desires might net out, words used on their social media posts could be interpreted in similar ways.

3.2 Graphing For Gold

Regardless of the possible ways that this base intelligence graph *could* be used, our first goal with draperBot is to connect consumers and brands together with data that is available to the general public.

With that as the goal, finding generally available data for global consumers is limited to very few data suppliers. The data source that was selected for initial use is from an organization called World Values Survey (WVS)^[5], which is a non-profit organization that canvasses the globe and surveys individuals to understand how the dynamic of cultural beliefs and values might affect the advancement, economic and otherwise, of many regions across the globe. Their survey data consists of over 200 questions probing into the nature of such beliefs and values. As such, many of their questions also answer what an individual's proclivity to the 5 core desires (or any other core driver framework used) would be, and since the questions and answers are made available openly to non-commercial needs, a mapping of these questions (and the answer data), became a basis for understanding the consumer side of the goal to connect brands and consumers at an emotional level. This mapping can be found in the appendix as **A1.Table1**.

The data from WVS is available in a variety of statistical package formats, including the open-source statistical software R, which was selected for draperBot to consistently maintain the open-source nature of this project end-to-end. The data will not be packaged directly with draperBot, as that would be a violation of the agreement WVS maintains with its data users, but the data can be downloaded for free^[6].

Using the R script (*score_setup.r*) provided in the draperBot open-source package, the data can be manipulated into visual representations, including the visual provided in **Figure 3**, which renders all survey respondents answers, which has more than 85,000 participants, to rank the importance of each of the 5 core desires to the respondents. The question to core driver framework mapping is available in Appendix **A1.Table1**.

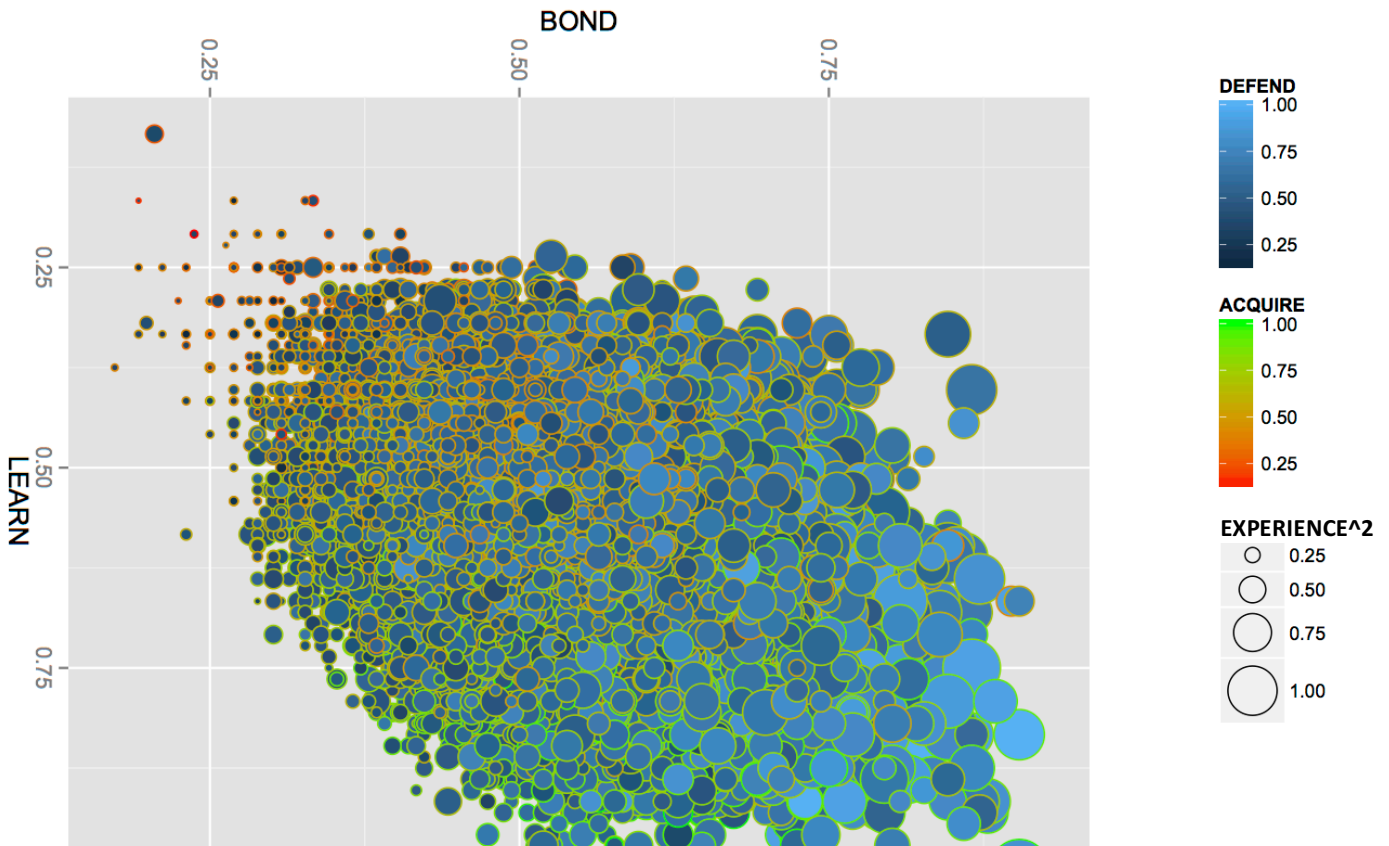


Figure 3

In addition to the R script rendering the data in an all plots sense, the data was also rendered in a comparative (comparing 2 desires at once) perspective across all the desires (not all combinations), as can be seen in **Figure 4**, which illustrates a general middling of the data points, but also that additional investigation, specifically in the desire to acquire is warranted.

The last component available in the R script is the process of clustering the data for better understanding. The statistical method chosen here was K-Mean clustering with a Lloyds model. A variety of model runs were run to understand the best possible grouping of data to chose, eventually selecting a grouping of 6 groups, which was later reduced to 5 groups of data results, as per the inflection (or elbow) found in the K-mean data graph, **Figure 5**.

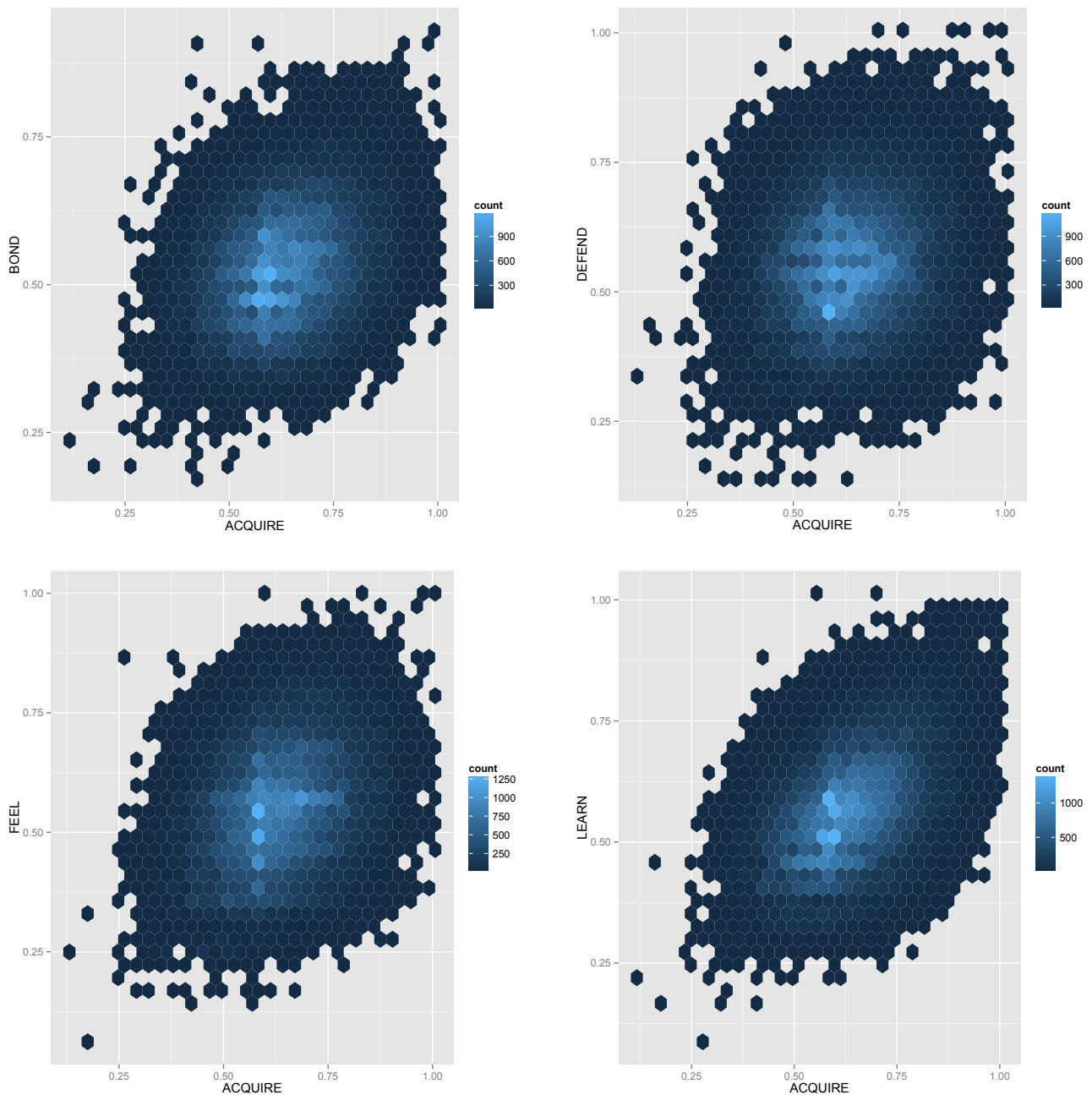


Figure 4

While no strong conclusions are likely to be made from this data analysis there the output of the data clustering results, available in Appendix **A1.Table2**, show that the data is not statistically middling, meaning that the data clusters show a bias to specific core desires, without the mathematics being manipulated to provide this result, indicating that not all people care about these 5 core desires equally. While not a dramatic proof, considering that this proof may be counted as a basic assumption by many, it is mathematical fact nonetheless, as it validates the use of the survey data and the core desire framework working together.

Even more dramatic than any potential proof though is the notion that there is a large set of metadata against each of the survey entries that allows for many more correlations and understandings to be extracted from this data set. For example, the data could be sliced by country of origin, by gender and any other number of fields contained in the WVS data records.

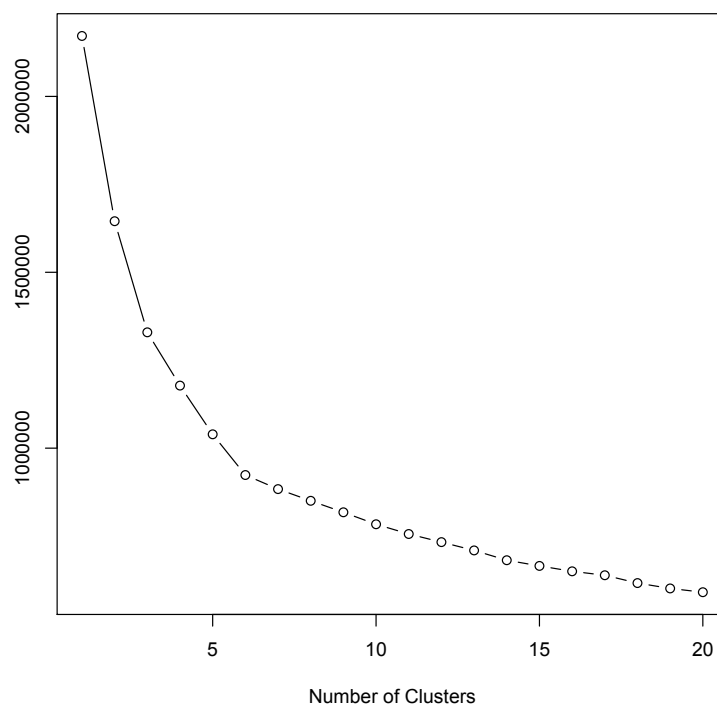


Figure 5

3.3 Now, Let's Build Some Brand Profiles

OK, so we've learned a bit about using data to understand how people might think about the 5 core desires (or any other framework chosen), but we have not discussed yet how we do the same for brands. Similar to the consumer approach, there is a set of supporting source code files available in the "*word_based_meaning*" folder of the project. The brand profile building code set, including a live demo version that runs on draperBot.com, addresses the integration of the base intelligence graph with a set of words from known sources. Selecting these sources is up to the users of the draperBot software, with some key approaches being pre-built into the software. The knowledge gathering approaches pre-built in the code calculate word frequency from brand-owned or brand reference (such as Wikipedia) websites, from recent news articles mentioning the brand name and an optional set of brand keywords, and lastly from popular web results using the brand name and optional keywords.

This approach of pulling together multiple sources of data may actually be something that is both useful and not. From an inclusive perspective, more data creates more accuracy, but in this case the data sources may be opportunities for data classification moments themselves. In fact, one of the highest potential commercial use cases for this portion of the draperBot framework is to evaluate the distinction of emotional brand voice and tone across different channels, rather than read all of them at the same time.

With this in mind, the live demo version being used for draperBot.com (siteeval.js) is a simplified version that only focuses on the approach mentioned of inputting a specified website to be downloaded and calculated word frequency is used to create a share of voice across the 5 core desires, an example of which can be seen in **Figure 6**. More comprehensive (multi-source) brand voice data performed on 9 brands is available in Appendix **A1.Table3**.

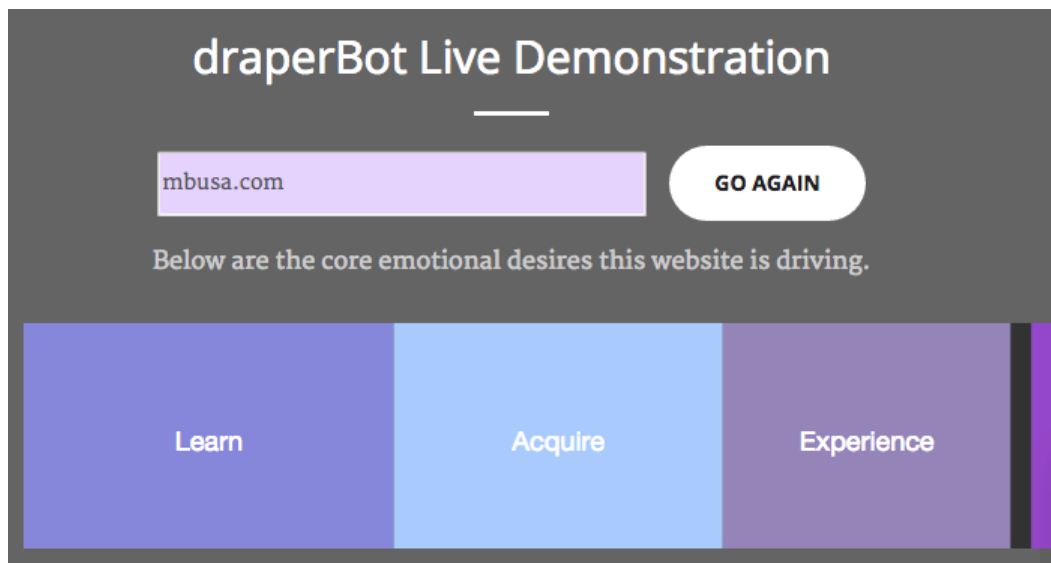


Figure 6

3.4 Bringing the Data Together

So with all this data being generated in the current form of draperBot, are there clear ways to analyze it? Are there conclusions that can be drawn from the data here?

Yes and no. (Of course)

Yes, there are a few currently explored ways to analyze the data.

1. Comparing the brand data to other brands.
2. Comparing the WVS data clusters (or applicable consumer slices) to the brand voice profile.
3. Comparing brand voice from channel to channel (does the website content carry the same tone as our print, email or social?)

Let's be clear, these methods are being explored currently, but there has been too little analysis to share output, success, failure or otherwise. These are experimental methods of data analysis at this point, but I would encourage the users of draperBot to further ideate and explore new ways to leverage the data that is being generated, and share!

So as far as conclusions, it's obvious that conclusions at this point would be premature. But if I haven't disappointed you enough to stop reading this section, I'll share one bit of interesting analysis that was observed while practicing method #2 above (comparing one brand to another). If you compare the data from Apple and Mercedes-Benz in **A1.Table3**, it's clear that these brands represent their brands using starkly different wordsets in their marketing and communication. While you can look at this result and say that makes sense, one sells cars, the other sells electronics (for now), further examination around this result is warranted. The result depicted in , clearly shows that Apple's result is balanced across all 5 core desires almost equally, versus Mercedes-Benz's clear bias towards Acquire, Learn and Experience. It could be argued that both of these brands are successful at communicating, so this examination is not about pointing out failure, but rather that success looks very different for different brands. In Apple's case, they are attempting to appeal to the masses with their products, so they

have balanced their communication. Comparing Mercedes to this and its clear they are targeting more specific audiences, those who are interested and capable of acquisition of their more exclusive products.

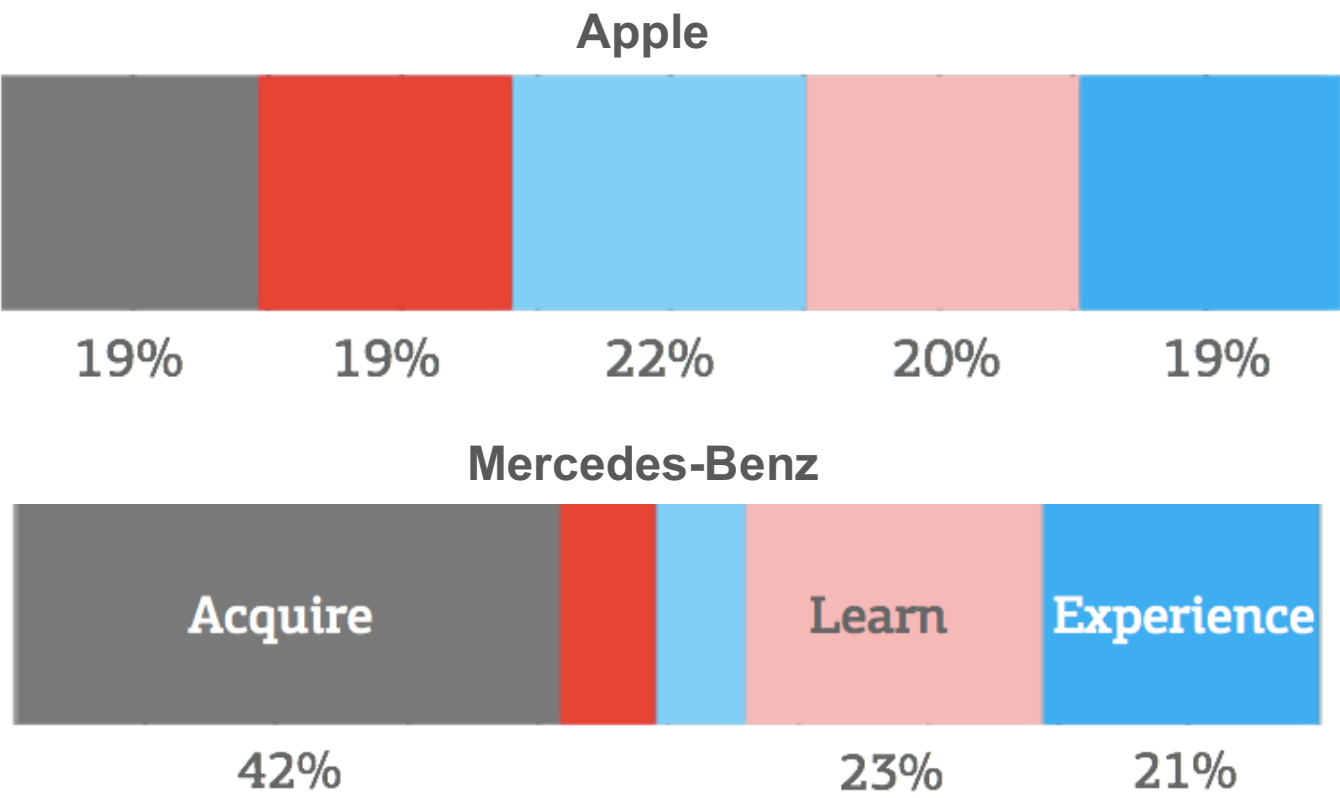


Figure 7

4 Tomorrow's Journey

4.1 What Does This All Mean?

TODO: Draw some conclusions around the value of this mechanism (even in its present form). But elude to the future of the tool needing to harness more meaning making.

4.2 Open Source Thinking

TODO: Talk about need for more meaning making, for this project to continue beyond my efforts. In a sense, we are trying to create an artificial intelligence system to filter our marketing messaging through that will tell us the emotions, the reactions and the acceptance of our messaging before we put it out there. It will take hundreds or thousands of iterations to get the system to where it needs to go for realizing this ultimate goal, but we all have to start somewhere. Call for open source community to take over and turn this into “the” solve for big brother marketing tactics, by creating a supercomputing technology that allows us to bypass massive data collection and truly get inside our heads from a meaning and emotion perspective.

4.3 A Framework For More Meaning

TODO: Talk about the structure of the software design, as to allow more meaning-making to be plugged into this system, more adaptations to the scoring thinking. Reiterate the call to open source volunteers.

4.4 TODO?

Appendix

1. Data Tables

1) Table 1 – Core Driver/ Mappings to WVS Data

WVS Question #	Highest Score	Lowest Score	CoreDriver(s)	Question Detail
V70	1	6	Learn	new ideas and creativity
V71	1	6	Acquire	money, expensive things
V72	1	6	Defend	security in surroundings
V73	1	6	Feel	having a good time, spoiling oneself
V74	1	6	Bond;Feel	Doing good for others
V74B	1	6	Bond;Feel	Caring for those nearby
V75	1	6	Acquire;Bond	Being successful and being recognized for it
V76	1	6	Feel	Adventure, live an exciting life
V77	1	6	Defend	Behave properly, avoid doing anything others think is wrong
V78	1	6	Defend	Environment
V79	1	6	Bond	Tradition, religion
V12	1	2	Defend	Independence
V13	1	2	Learn;Acquire	Hard Work
V14	1	2	Defend	Feeling of Responsibility
V15	1	2	Learn	Imagination
V16	1	2	Bond	Tolerance / respect others
V17	1	2	Acquire	Thrift/ value of money
V18	1	2	Acquire	Determination
V19	1	2	Bond;Feel	Religious faith
V20	1	2	Bond	Unselfishness
V21	1	2	Bond	Obedience
V22	1	2	Feel	Self-expression
V66	1	2	Defend	Under war would you serve

V182	1	4	Learn;Bond	worry for educating my children
V4	1	4	Bond	Family (how important)
V5	1	4	Bond	Friends (how important)
V6	1	4	Learn	Leisure Time
V7	1	4	Defend; Bond	Politics
V8	1	4	Acquire; Learn	Work
V9	1	4	Feel;Bond	Religion

- 2) Table 2 – K-Means Clusters (from WVS R Data file) – amounts listed are percentage that each group answered question towards a specific core desire.

	ACQUIRE	BOND	DEFEND	LEARN	EXPERIENCE
Group 1 - Experience	18.4	22.9	12.7	21.0	25.0
Group 2 - Learn	20.7	21.1	13.5	27.2	17.5
Group 3 - Acquire	25.8	20.1	14.3	18.7	21.1
Group 4 - Acquire + Learn	28.6	14.0	13.8	26.3	17.3
Group 5 - Everything Matters	21.9	16.7	20.1	22.9	18.4

- 3) Table 3 – Brand Analysis, Multi Source, With Optional Keywords – Indexed Score (computed relative to other brands and as a percentage of overall voice)

INDEXED SCORE						
BRAND	Branded Product	acquire	bond	defend	feel	learn
Coca-Cola		19	22.4	10.9	15.7	32
Audi	A4	21.7	9.5	21.4	25.2	22.1
Mercedes-Benz	C class	40.2	8	7.1	22.5	22.1
BMW	3 series	34.6	12.9	14.2	17.7	20.6
Honda	Accord	36.2	2.8	12.3	35.8	12.8
Subaru	WRX	25.4	11.6	17.4	21.8	23.8
Apple	iPhone	18.1	19.8	22.8	19.5	19.7
Ebay		19.2	22.7	15.3	24.7	18.1
Walmart		14	32.3	21.5	16.7	15.5

4) Table 4 – Brand Analysis, Multi Source, With Optional Keywords – Raw Scores

RAW SCORE						
BRAND	Branded Product	acquire	bond	defend	feel	learn
Coca-Cola		47400	4800	1660	9560	15520
Audi	A4	21770	820	1310	6150	4300
Mercedez-Benz	C class	44290	760	480	6040	4720
BMW	3 series	61730	1970	1540	7720	7120
Honda	Accord	41410	280	860	10000	2840
Subaru	WRX	73750	2880	3080	15450	13400
Apple	iPhone	437940	41170	33560	115170	92440
Ebay		55620	5640	2710	17420	10180
Walmart		80400	15950	7550	23530	17330

Bibliography

[1] "General Agreement on Tariffs and Trade". Wikipedia. 23 March 2015.
http://en.wikipedia.org/wiki/General_Agreement_on_Tariffs_and_Trade

[2] "Syntactic ambiguity". Wikipedia. 13 July 2015.
https://en.wikipedia.org/wiki/Syntactic_ambiguity

[3] "Pragmatics". Wikipedia. 13 August 2015.
<https://en.wikipedia.org/wiki/Pragmatics>

[4] "Carlo Dalla Pozza". Wikipedia. 25 June 2015.
https://en.wikipedia.org/wiki/Carlo_Dalla_Pozza

[5] "What We Do". World Values Survey. 6 September 2015.
<http://www.worldvaluessurvey.org/WVSContents.jsp>

[6] "WVS Wave 6 (2010-2014) Download Page". World Values Survey. 4 April 2015.
<http://www.worldvaluessurvey.org/WVSDocumentationWV6.jsp>