How Predictable
Patterns of Human Economic Behavior in the Wild

Katherine (Coco) Krumme
B.S. Yale University 2005

Submitted to the
Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of
Master of Science
at the Massachusetts Institute of Technology

September 2010

© Massachusetts Institute of Technology 2010. All rights reserved.

Author

Coco Krumme
July 30, 2010

Certified

Alex (Sandy) Pentland
Professor of Media Arts and Sciences
MIT Media Lab

Accepted

Pattie Maes
Department Head
Media Arts and Sciences
How Predictable
Patterns of Human Economic Behavior in the Wild

Coco Krumme

Submitted to the
Program in Media Arts and Sciences,
School of Architecture and Planning,
on the 30th of July, 2010 in partial fulfillment
of the requirements for the degree of
Master of Science

Abstract

Shopping is driven by needs (to eat, to socialize, to work), but it is also a driver of where we go. I examine the transaction records of 80 million customers and find that while our economic choices predict mobility patterns overall, at the small scale we transact unpredictably. In particular, we bundle together multiple store visits, and interleave the order in which we frequent those stores. Individual predictability also varies with income level. I end with a description of how merchant composition emerges in US cities, as seen through the lens of credit card swipes.
How Predictable
Patterns of Human Economic Behavior in the Wild

Coco Krumme
September 2010

Advisor
Alex (Sandy) Pentland
Professor of Media Arts and Sciences
MIT Media Lab

Reader
Dan Ariely
Professor of Behavioral Economics
Duke Fuqua School of Business

Reader
Erik Ross
Senior Vice President
Bank of America
Why Thank You

That there is nothing new under the sun bears repeating. That being said, Manuel Cebrian at MIT and Erik Ross at BAC deserve special thanks for their generosity, time, and insight: sine qua non. A number of others offered critical feedback, fellowship, and fun in the coinciding months—and I’ve thanked you each outside this text.

The thesis is set in Optima, a font designed by Hermann Zapf. How could a would-be economist elect anything else?
Perfunctory Quotations

I worked my way up from nothing to a state of extreme poverty – Groucho Marx

Men's ideas are the most direct emanations of their material state – Karl Marx
## Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before. Abstract and Acknowledgements</td>
<td>2</td>
</tr>
<tr>
<td>One. How Predictable</td>
<td>7</td>
</tr>
<tr>
<td>Two. Homo Economicus Naturalis Historia</td>
<td>10</td>
</tr>
<tr>
<td>Three. Especially about the Future</td>
<td>20</td>
</tr>
<tr>
<td>Four. Predictability and its Discontents</td>
<td>27</td>
</tr>
<tr>
<td>Five. Some consumers are more predictable</td>
<td>32</td>
</tr>
<tr>
<td>Six. Material World</td>
<td>35</td>
</tr>
<tr>
<td>Seven. In Sum</td>
<td>38</td>
</tr>
<tr>
<td>Eight. References</td>
<td>40</td>
</tr>
<tr>
<td>Nine. Appendix</td>
<td>42</td>
</tr>
</tbody>
</table>
1
How Predictable
From individuals to merchants to cities

If we are truly what we eat, one might know a man by where he shops for food. Pick a purchase at random from a person's credit card statement, and it's likely (about 40%) to represent a grocery store or restaurant. Then, look at his purchases over time: and time and again he'll return to the same grocery store and fast food joint, and spend more or less the same amount.

Here, I use credit card transaction data as a lens on patterns of human behavior. In short: how predictable is shopping? My findings are threefold. First, given an individual's past purchases, we can predict with high accuracy where he'll go next. Second, shoppers group together, or bundle, purchases, and shopping patterns vary by income level. Finally, the economic landscape of American cities develops in a predictable manner. While individual predictability is limited on the small scale by our tendency to interleave the order in which we visit shops, from 10,000 feet, we're tried and we're true.

I focus here on economic behavior as a means of studying the predictability of people. Shopping is not age-old, and it isn't what makes us human. We've long foraged, hoarded, traded chits and shells, fabricated and forged currencies, and built new worlds on promised notes. But most human behavior is much older: we exchanged rites and rituals long before we traded coins. Economic activity is mere scaffolding atop the richer lives of mind and body: although never quite deemed baseless, shopping is often derided as auxiliary, base, and vain.

At the same time, shopping—in the narrow, modern sense of credit cards swiped—comprises the very structure of modern American life. It's a largely elective behavior, and yet it governs where we go, whom we meet, and how we fill our pantries, closets, garages, and weekend afternoons. We clip coupons, navigate parking lots, and rise well before dawn on national holidays—all in the name of the demigod we call Commerce.

But where and when and why do we shop? How often do the ruts cut out by my economic routines cross those cut by my neighbors? The present analysis uses millions of credit and debit transactions to consider just how predictable Americans are when we shop. The dataset, described below, is a sample of 10,000 individuals drawn from nearly 80 million customers of a major US financial institution, including all credit, debit, and check transactions, as well as cash withdrawals and wire transfers. These are the financial footprints that together comprise the invisible hand.

* But what does it mean to be predictable? Our notions of predictability are grounded, both computationally and colloquially, in the ideas of information theory. Claude Shannon defined entropy as the number of bits per character needed to encode a sequence [Shannon, 1948]. Since, Shannon's entropy has been co-opted by everyone from Thomas Pynchon (in an eponymous short story, Entropy appears in the form of wild winter nights in a tiny Washington apartment) [Pynchon,
1957] to population ecologists (as a measure for the species richness in an ecosystem). Here, we continue the pattern of co-option for our own (commercial) purposes.

Entropy is in some sense a measure of predictability, and people have been obsessed with predictability for a long time. The ability to place reasonable bounds on future affairs is of course critical for agriculture, travel, and trade, but it’s also a natural product of human curiosity.

In line with this obsession, we’ve in turn loved and detested those who claim privileged knowledge about the future. The Assyrians lived by their fortune-tellers and the Greeks by their oracles; later, Leviticus suggests we “stone all mediums and necromancers” as well as “all who whore after them” [Leviticus 19:31]. We might cite both the Wall Street Journal’s financial analysis and @Twittascope – an online horoscope and one of the most followed broadcasts on the twitter platform – as more recent evidence of our devoted interest in the future.

Here, I consider the regularity of individuals. When we speak of a “predictable” person, what is it that we mean? Is it someone who is reliably located: the retiree on the park bench on the regular at the local bar? Does predictability entail being reliably propelled: if I have lunch at the steakhouse, I’m bound to stop off at the Starbucks’s next door; I pick up the dry-cleaning after dropping off the kids? Or is predictability rather the reliability of motivations: it’s not hard to tell where a predictable person will go next?

And when human behavior aggregates, we might wish to measure the predictability of a system. In many instances, collective human action only becomes more messy: a single trade, cascaded, can send a market into tailspin, and as cities grow the behavior of individuals changes—and changes the city—dynamically.

Here, I measure predictability using several metrics, which relate primarily to the first two descriptions of individual predictability: reliability of trajectory and reliability of location. First, I consider a “simple” predictability of location: how many months of data do I need to know 95% of the places you’ll shop over the course of six months? (Part 2) Next, I construct a network of sequential location information: given your presence at store A, how probable is it that you’ll go to store B next? (Part 3)

I then consider informational entropy as a measure of predictability: how diverse is an individual’s shopping behavior over a given window of time, and how much does sequential, versus temporal, variation drive uncertainty? (Part 4) I look at differences in how the rich and the poor shop (Part 5). Finally, I use transaction data as a lens to study the correlates of industrial diversity (and, perhaps, unpredictability) in cities over time: what elements are tied to increasing complexity in the economies of American metropolises? (Part 6)

Of course, any model is only as good as its inputs and assumptions. Here, I define four measures of predictability to compare individuals using a stream of behavior. Transaction data reveals nothing, of course, about decision-making in high definition, nor does it tell us about an individual’s predictability in other domains. It can, however, reveal both how we shop and how our patterns aggregate.

* After all, why should we care about shopping patterns? Beyond mere scientific or prurient curiosity, a number of business cases can be made.
We’d like to know, for example, the extent to which marketing campaigns make a difference in consumer behavior. Are certain people more likely to switch products, or to be loyal to a brand? To what extent are shopping patterns immutable (governed by existing economic habits) and to what extent can we be convinced to shop somewhere else? If we can model a person’s probability of transition from point A (a department store, for example) to one of several points B (say, restaurants), we might offer a coupon at point A to coerce him to a single B.

There are potential applications to economic development as well, at two scales. If we know how spending behavior changes in response to a change in income (or to an exogenous shock, such as a change in a city’s industrial landscape), we might better respond to such shocks. And the dynamics of a city’s industrial development has implications for developers and transportation planners.

* In the present section, we are introduced. Part 2 describes the data and gives context. Part 3 details simple results on the predictability of people. Part 4 introduces measures of entropy and compares shopping patterns to mobility patterns derived from cell phone data. Part 5 considers how income correlates with predictability, and part 6 looks at how human economic behavior assembles in cities. Then I conclude, cite, and append.
2

Homo Economicus: Naturalis Historia

The financial footprints of our species

Transaction data offers a lens on economic activity: economies move because people fritter away their money. To date, there’s been little to link the micro to the macro. Credit card data allows us to observe how people make decisions in the wild, and how individual habits accumulate over time.

While the academic study of economic behavior has been the province of models and, more recently, behavioral experiments, the study of transaction data by banks has largely focused on the segmentation of customers for the purpose of assessing risk. Some individual merchants (notably: grocery stores and casinos) have taken solid stock of their customers’ habits, but lack a global view to compare apples purchased in their stores to those bought from a competitor.

Here, I aim for analysis whose reach is slightly broader than a lab’s, slightly more far-sighted than a bank’s, and slightly more universal than that of a single merchant.

In this section, I present a framework for studying shopping patterns in the wild, and survey the literature on shopping behavior, human mobility, and the development of cities. I also describe the transaction dataset and subset used here, assess potential biases in the data, and characterize our commercial behavior with basic statistics.

Shopping behavior

The conventions of consumption have been studied from the perspective of psychology, physics, anthropology and economics. No matter the disciplinary lens, much about shopping remains the same. Although the Phoenician sea routes long ago gave way to the suburban mall, and Persian coins to AmEx Gold, we continue to congregate for commerce, and to rely on currencies to transact efficiency and fairly. At its heart, of course, consumption is driven by basic needs: for food, shelter, warmth, and transport. We buy things to entertain ourselves and fill our homes (and to organize the things that fill our homes).

Psychologists have studied what motivates shopping. Misery, it’s been found, is not miserly: we are apt to spend more when we’re sad [Cryder et al, 2008]. Studies of online shoppers have shown fun, control, and saving time to be chief motivations in staying home to shop [Aron, 2005].

Others have looked at the element of choice in where we shop, whether logistical (proximity to work), preference (for brand or price), social (the recommendation of a friend), or serendipity (the right place and time) [Arnold, 2003]. And then there’s what we buy: there’s some evidence that spending on luxury items drops, or at least diffuses, as discretionary income falls [Gardyn, 2002].

It is by now a truism that we rarely know what we want (or: preferences are fuzzy), and that more choice is not necessarily more efficient or more conducive to satisfaction [Schwartz, 2005]. Our commercial lives are a generalization of the paradox of too many jams [Iyengar, 2000].
Additionally, we like to think of ourselves as less predictable than we are, as non-adherants to routine. We’re overconfident about our own abilities [Adams 1960] and overestimate our ability to predict the future [Tetlock]. We believe in our own unique proclivity for prediction [Dawes, 1979], even though simple models do better than human judgment in some settings [Kahneman 2009].

The analysis of credit card transactions can contribute scant evidence to theories of small-scale decision-making; instead, the present analysis aims to characterize the broad patterns that compile from small decisions over time: when is shopping driven by routine, and when do we deviate?

**Bounding the predictability of humans**

Literature from the physics community has considered patterns of human mobility and found that activity is governed by repeated and discernable routines. Additionally, theoretical models of social networks and real analysis of market data suggest mechanisms for measuring the part of an individual’s actions due to the choices of others [Salganik, 2006].

Human interaction is constrained by geography and necessity. We drive on the Interstate and not through the national park, we have a layover in Atlanta, we head to the office Monday morning after dropping off the kids at school, and pick them up on the way back home.

A literature on human mobility is emerging from the analysis of mobile phone providers’ logs. Using cell phone towers to pin down the location of a customer at the time of a call, and correcting for intervals without calls, scientists have described the trajectories of individuals over the course of the day. Human dynamics display strong regularity: an individual’s travel distance is time-independent, and people have a high probability of being found at several highly-frequented locations, independent of their average distance traveled and locations frequented [Gonzales 2008].

Beyond our broadest movements, some research suggests that human behavior is subject to rare, high amplitude bursts or crises. Sunstein summarizes the effects of cascades [2005] and Sornette goes so far as to claim their dynamics predictable [2002]. Barabasi attributes the “burstiness” of human behavior to a non-randomly-distributed queuing process, whereby we complete some tasks in rapid succession and let others linger, leading to fat tailed patterns of activity [2005].

Nonetheless, it is possible to bound the predictability of individual trajectories. Song et al [2010] measure the entropies of mobile phone users and find that, given information on the sequential locations of people, it is possible to place an upper bound of 0.93 on predictability: that is, for a mean user, about 7% of his behavior will fall outside of those anticipated by algorithm. Without sequence information, however, the predictability of an individual’s location is widely distributed.

Herein lies one of the starkest contrasts between the patterns that emerge from cell phone data and those from credit card transactions: while our large-scale routines are rote, close-up our economic behavior is flurrysome and chance, as we’ll describe below. Impulsive purchases and the interleaving of store visits define shopping at the small scale (interleaving –essentially, the randomization of subsets of store visits over short time scales—is defined in part 4).

While the results from transaction data to some extent validate those from mobile phone records, they also point to the uniqueness of shopping patterns. A cell phone tower is a waypoint in an individual’s daily trajectory, but a store is a destination, and ultimately, a nexus that drives human social and economic activity.
Accounting for acquisition

Several studies have explored how people choose where to shop. A 1978 study of two adjacent South Carolina communities found that residents in the richer of the two neighborhoods tended to optimize for grocery stores close to other merchants frequented, while poor residents elected stores close to their homes, which they revisited frequently [Lloyd 1978]. Others have considered what happens in markets when investors with different memory lengths come together [LeBaron].

Using transaction data, I aim to connect individual purchases with more general patterns. I create a measure for the *bundling* of shopping trips, and to find how people differ in their propensity to group store visits (Part 4). We can also explore volition by measuring the way patterns form, and then tracking the tendency of an individual to deviate from what he normally does. An understanding of how people switch from their habitual trajectories has implications for a number of fields, from marketing to health to city planning.

We'd like to know, in short, how much of shopping is determined by existing constraints on mobility and habit, and to what extent shopping is governed by “elective” search behaviors, social influence, or preference. I'll begin to answer those questions in the subsequent sections by comparing patterns of shopping with the general mobility patterns seen in call log records, and showing where economic predictability diverges.

*  

Finally, what can transaction data can tell us about a regional economy? Customers of the financial institution in question can represent up to 80% of the population of a city, allowing us to make reasonable predictions about the distribution of consumer activity in various metropolitan areas.

By examining properties of production in cities, such as patents, R&D employment, and new infrastructure, Bettencourt et al note strong scaling relationships that allow disambiguation of growth due to economies of scale from that due to innovation [Bettencourt, 2007]. We might, for example, be interested in how the growth of different industries, as measured by an increase in number of merchants and amount spend at those merchants, impacts city growth.

Another analysis proposes that growth is heralded by a comparative advantage in production: that is, the economic zones that excel are those that have a diversified economy and can produce a relatively specialized product compared to the production of others [Hidalgo, 2009]. Regions that produce only a product that is readily made by others fare poorly. Here, I use this methodology to study American cities, with industries as proxy for diversification (Part 6)

Transaction dataset and methods

The present analysis uses a sub-sample of transaction records drawn from a database of approximately 80 million customers of a major consumer bank (henceforth “the Bank”). Activity is available dating to 2005, and includes information on transaction date, amount, channel (e.g. check, debit, credit), merchant, merchant category code (MCC; described below), and whether the transaction took place on- or offline. Customers are identified by zip code, join date, and year of birth, and are associated with any linked (e.g. joint) accounts.
Transactions total about $30-$35 billion per month and thus can represent significant flows in the US economy. For the metropolitan areas we consider, a range of 28% to 79% of residents hold accounts with this financial institution, with a median of 57.5%.

The sample utilized for the majority of this research comprises the transaction history of 10,000 customers during 6-month periods (April – September) in the years 2006, 2007, 2008, and 2009. All records are captured, including credit and debit purchases, inflows to the account, automatic online payments, paper checks, and cash withdrawals.

The 10,000-person random, anonymized sample is used to study geographic patterns and basic predictability. For our analysis of entropy and by-income variation, we take a second slice of our sub-sample and consider the approximately 2000 of these individuals residing in the mid-Atlantic region.

Individual income is inferred using inflows to an account. To prevent returned purchases and other debits from being counted as income, we consider only those inflows coming tagged with identifiers for employer direct deposit, annuity or disability payments, and Social Security income. What we call “income” actually captures a reasonable lower bound on true income. It is possible that not all of an individual’s true income is captured by our measure: for example, if a person’s earnings are primarily in the form of cash or personal check, or if he deposits only a portion of his salary into his account with this Bank, and routes the remainder to a retirement or stock market account, a spouse’s account, or a personal account at a separate bank. I anticipate that the effect is stronger for wealthier individuals, who tend to have multiple accounts and are generally more sophisticated financially [Federal Reserve 2007]. Therefore we expect these estimates to exhibit amplified dampening as income rises.

Other sources of sampling bias arise from the 10 million American households (the “unbanked”) without bank accounts [Federal Reserve]. This absent slice tends to include recent immigrants to the United States as well as residents of very rural areas and urban centers. Our sample comes from a bank with relatively even distribution across all other income categories, although the wealthiest American consumers tend to be under-represented by this financial institution. We also expect that in filtering for accounts with electronic inflows (of any amount), we are biasing our sample against individuals who are paid exclusively in cash.

In drawing the quintiles for analysis, the distribution of our proxy for income actual falls below the distribution of American household incomes. Our sample has a median account inflow (whether individual or joint) of about $24,000 annually, while the national median was $52,000 in 2008 [American Community Survey]. I surmise that this discrepancy arises from individuals placing only part of their incomes into this account, rather than from a customer base with below-average income.

Information on merchants is provided in the form of a string, which often includes store name and (in the case of chain retailers) number, and occasionally information on location. Some of these strings have been hand-coded and standardized: the aggregate business name is also listed in a separate column.

To categorize merchants, I use the MCC codes established by MasterCard and Visa. A list of MCC categories is included in the Appendix. The distribution of codes is heavily skewed in three categories – there are about 150 codes for individual airlines and 200 for individual hotels, for
example – and I create three new aggregate categories to comprise (1) all airlines, (2) all hotels, and (3) all rental car purchases.

Because the data used here does not include information to pinpoint a store geographically, we are need to build proxies for the locations of merchants

These estimates are based on a gravity model or estimate of the relative tie strength between pairs of location. Originally proposed to explain volumes of migration between cities, the model takes the form

\[
\frac{\text{population}_A \times \text{population}_B}{\text{distance}^2}
\]

where A and B are two locations. Distance can be geographical distance, or the functional communication distance or travel time.

Reilly’s law of retail gravitation is a proposed extension of this model, intended to predict which city a customer will choose to frequent for shopping, based on city size and distance [citation 1931].

The break-point is defined as the instance of indifference between traveling to each of two cities:

\[
\text{BP} = \frac{\text{distance}}{\sqrt{\frac{\text{population}_A}{\text{population}_B}}}
\]

where A is the larger city

Huff redefines a trading as a set of probability contours for a given product and the resident set of consumers.

![Figure 1. Contours of a retail trading area. A similar gravity model can be used to predict the location of a store from the zip codes of the customers making purchases at that store.]

I build on these findings to compare individual stores across a merchant chain, and to show what factors contribute to differently-constituted retail areas.
For the analysis of US metropolitan regions, I select 35 cities based on (1) presence of Bank customers and (2) size and general importance. The list of cities and corresponding customer zip codes is described in the appendix. For each metropolitan area, I examine all of the transaction records associated with customers in the relevant zip codes for three-month periods (April, May, June) for the years 2005 to 2008.

Within cities, certain types of individuals may be more or less likely to hold Bank accounts compared to individuals in other cities, whether due to differential marketing efforts, first mover effects, or specific location of bank branches. We believe our analysis of mobility patterns is sufficiently broad as to render this bias negligible. For the study of cities, I additionally weight samples by a measure of persons represented in Bank data relative to Census population estimates.

All zip code-level income, population and population density estimates come from the Census Bureau.

Data was drawn from the Bank database servers using an SQL client and returned in column format. Initial processing was then conducted in Python, and further statistical analyses and simulations in the software packages R and MATLAB. Visualizations and charts were created using R, MATLAB, cytoscape, and Microsoft Excel.

Individual account details were extracted in anonymized form, and all information presented in this analysis is sufficiently aggregate to preclude the possibility of inferring personal or private information.

**Mean behavior**

Using this sub-sample of individuals, we chart basic statistics describing how Americans spend their shopping time and money. As we see below, the majority of individual visits made (locations at which transactions occur) are to restaurants, grocery stores, and gas stations. Other categories of miscellaneous retail tail these major purchases, including department store, discount, liquor, and barber shop purchases.

Figure 2 shows a breakdown of how Americans spend there time. Of all shopping visits, what percentage is spent at each store type (by MCC industry).
Figure 2. Distribution of visits of 2000 customers to stores, by merchant category code, over 6 months in 2007 (April 2007 to September 2007). Over 50% of visits were to restaurants, gas stations, and grocery stores.

A chart of amount, rather than time, spent tells a slightly different story. Most of our swipes are dedicated to small purchases: food and gasoline but when we consider dollars spent (Figure 3), the miscellaneous categories (one time visits to the vet, a big automobile purchase, a donation to charity) begin to eclipse the common ones. Still, the largest single category into which we put our money and our time is, predictably: food.
A simple analysis of the changes in basic spending categories between 2007 and 2009 (for the same set of individuals) shows a decline in both percentage spent and number of visits in many categories. We see a disconnect between changes in visits and changes in amount spent for two types of shops: the post office, and gas stations (gas prices rose appreciably during this period). Interestingly, visits to and total spend at liquor stores rose by the highest percentage, which we’ll conservatively attribute to an artifact of sampling rather than to a generalized reaction to financial crisis, although there exists some evidence that we spend more on “sin” products where the economy goes south [Fabozzi, 2008].
We can also consider the aggregate weekly patterns of shoppers, to examine on which days of the week different activities are likely to be distributed (Figure 5). Not surprisingly, we’re more likely to eat on a Saturday than on a Monday, to buy home supplies over the weekend, to stop by the liquor store on Friday.

Figure 4. Percentage change in amount spent and number of visits of 2000 customers between 2007 and 2009 across consumer categories.
Figure 5. Distribution by weekday of visits of 2000 customers to stores by merchant category, across 6 months in 2007. Shoppers are most likely to frequent liquor stores on Fridays and gas stations on Thursdays.

We now know where to find people en masse, and that en masse, people are largely predictable. At the same time, individuals can deviate significantly from these averages, and in the next sections we’ll explore the varied commercial trajectories of Americans.
3
Especially about the future
Where and when do we shop?

Yogi Berra reminds us that prediction about the future is especially difficult. In the previous section I examine aggregate patterns of shopping across merchant categories in the present section, I both step backward from and reorient the question of predictability in shopping. I ignore store type and consider each store an equivalent waypoint in a customer’s trajectory, in order to study the more generalized patterns of how our shopping drives where we go.

Throwing darts

Much of consumerism is not conspicuous, but routine. Transaction data suggest that shopping behavior is constrained by some of the same features that govern mobility generally. Shoppers return to familiar stores with remarkable regularity: a Zipf distribution describes the probability that a customer will visit a store at rank $N$ (where $N = 3$ is his third most-frequented store, for example), independent of the total number of stores visited in a 6-month period (Figure 6). These results support those of Gonzalez et al [2008] with location inferred from mobile call logs.

Just as with cell phone data, we find that the range of shoppers resembles a truncated Levy flight, and that the Zipf distribution holds independent of the total number of purchases an individual makes. If one were to throw darts and try to hit a particular shopper, two darts thrown at locations 1 and 2 would have a 35% chance of hitting the target: that is, we return again and again to habitual “feeding grounds.”
Figure 6. Zipf distribution describes probability a shopper will visit location L. This distribution is independent of the total number of locations frequented (here, average probabilities for individuals who visit 20, 30, 40, and 50 locations in the 6-month period are shown). Slope = -4, average standard error = 0.031

However, segmented along demographic lines, the chart looks different, as we’ll see in Part 5.

Unlike the mobile phone logs, our transaction records afford no indication of a merchant’s location: thus was can make no claims about the self-similarity of individuals who travel long or short distances over a day, as do Gonzales et al. However, most of the findings therein assume away all notions of total distance by arguing that the same patterns hold independent of radius of gyration. Future work with transaction data might use a proxy for radius of gyration to validate these results.

**Shopping sprees**

The main difference between mobility as predicted by phones versus credit cards is rooted in how we group together shopping trips. Commerce is often organized around commercial centers. Although Venice is no longer a trading post for anything but novelty masks, its historical spirit of condensed consumerism lives on: we might spend a Saturday meandering between the Gap and Home Depot, the Blockbuster and the Starbucks, without ever leaving the confines of the mall. This bunching of stores serves several purposes: it reduces travel costs for the consumer, it allows merchants to reap certain benefits of the collective, and incites customers who’ve already “made the trip” to succumb to unplanned purchases.

Much thought and more than a little design are brought to bear on the modern day mall [Downs, 1970]. Escalators are placed to maximize circulation and induce browsing. The outside world is eclipsed. The food court is never far away, and bright-eyed salesgirls beckon from the median trinket booths.
When it comes to shopping, the unit of measurement is the trip. Some purchases are made individually – a new car, for example, or a Saturday outing for brunch – but an important subset are bundled together: dinner and a movie, coffee after lunch, the parking lot, dentist, and daycare. Some of these groupings occur in the same sequence every time – few of us possess sufficient bravado to eat dessert first – yet most follow no definite ordering: a trip to the mall, for example, might mean a new ordering of stores visited.

Figure 7. Network linking stores frequented by a single individual over 6 months in 2007. Edge weight represents the transition probability from one location to another. A grocery store serves as the shopper’s main hub, and a natural grocery store as his secondary hub. Actual store names have been replaced with merchant type.

We begin to glimpse the pattern of these transitions from one store to the next – and to guess at motivations – when we visualize a network of a single individual’s chosen purchases. Figure 7 describes the links between stores visited sequentially, with thickness as proxy for the number of
times a transition was made. Node 1, a grocery store, serves as a “hub” of the individual’s shopping, and two restaurants are at the ends of strong spokes: here is someone who tends to eat out, predictably, at one of two locations before or after grocery shopping. Such a network might have interesting commercial applications: the grocery store could offer a coupon to drive traffic to one restaurant over the other, for example.

However, the prediction of the “next step” is only probabilistic: there’s actually a fair bit of interleaving of stores that occurs within the context of a single shopping trip. In the next section, we show that the sequence of purchases has little to no impact on the predictability of shopping behavior: due in large part to this interchangeability of shopping events. Unlike our mobility patterns, which are ordered by daily routines and anchored at work and home, our shopping is predictable only from 10,000 feet: at the single purchase level, impulse and interleaving rule.

**Dependent paths**

By scaling up an individual’s network, we can produce a map of interrelated industries. How prevalent is the habit of hitting the food court after shopping? Of heading to the bar after paying the electricity bill? If I make an exception and go to the organic market, am I less likely to stop at the burger joint on the way home?

In particular, we can select two subsets of people, the “most predictable” and the “least predictable” quintiles, from our dataset, controlling for income. Our precise measures of predictability will be defined in the next section: for now, it is treated as a generic segmentation. We then construct a network with industry (MCC) codes as nodes and links describing a pair of consecutive purchases occurring at a store in each of the two linked industries.

Figure 8 shows the network of industries frequented by the most predictable quintile (further defined in Part 5). As in the individual case, the grocery store (5411) serves as the hub for much consumer activity, followed by restaurants, gas stations, and drug stores. As in the individual network above, we can also detect transition probabilities between two types of merchant locations, aggregated here across the population as a whole.
Figure 8. Network of transitions between stores of different merchant category codes of 2000 individuals in 2007. Edge weight and color represent the transition probabilities from a merchant in one retail category to another. See Appendix for MCC codes.

By adding people to the picture, that is, by constructing a bipartite network linking people to the stores at which they shop together, we might also construct a model for predicting risk of default or other individual metrics.
We also control for income level we can measure the non-wealth dependent difference between high and low entropic individuals in terms of rank-1-store: while the “most predictable” individuals are most likely to be found at a grocery store, we can expect to find their “least predictable” brethren at a gas station or fast food restaurant.

<table>
<thead>
<tr>
<th>Most Predictable</th>
<th>Least Predictable</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Grocery store</td>
<td>1. Fast food</td>
</tr>
<tr>
<td>2. Drug store</td>
<td>2. Gas station</td>
</tr>
<tr>
<td>3. Restaurant</td>
<td>3. Drug Store</td>
</tr>
</tbody>
</table>

Table 1. Top locations of most and least predictable individuals (where we’d hit them if throwing darts).

We find that there is lower overlap for the type of stores frequented communally by Most and Least predictable people. In a 10,000 run simulation, there is a 11.2% chance that a member of the highly predictable set will be found in the same store type as one of the unpredictable set, compared to a mean 31.4% chance that either a Most with coincide with a Most, or a Least with a Least. Different, only partially overlapping planes are cut out by the trajectories of the most and least predictable shoppers.

**Predictability in a nutshell**

A simple measure of an individual’s predictability is the novelty of the shopping locations he chooses over time. What percentage of all stores visited in a six-month window are captured in the first month of visits? After 3 months? After 5 months? Is the rate of new store uptake, or exploration, spiky or constant?

Figure 9 shows patterns of merchant uptake over time for a sample of Bank customers. A consumer who visits 95% of the same stores in month 1 as he does over six months could be considered more predictable than a customer with a great deal of “churn” in his shopping locations. While a single spike may point to an exogenous change, such as a move, a layoff, or a child beginning school, more constant unpredictability may signal an individual who is more likely to explore new merchants or to switch brands, for example.
Figure 9. Average pattern of store uptake for a sample of individuals who shop at ~35 locations over six months. Error bars indicate one standard deviation of “percentage of total stores frequented by month x”. Most people exhibit a smooth uptake of new stores, but there is considerable spikiness in some trajectories e.g. rapid discovery in early or late months.

We’ve considered here where people are, and when, probabilistically, as well as how they begin to go there. In Part 4 I’ll present entropy measures to compare the diversity of individual trajectories, and discuss the two shopping behaviors – bundling and interleaving – that set our commercial activity apart from most others.
4
Predictability and its Discontents
Measuring unpredictability in shopping

Part 3 considered three measures of where a person might be: his probabilistic distribution over a set of shops, the network tying together pairs of subsequently visited locations, and an estimate of “simple predictability,” that is, of his preference for newness.

Here, I take informational entropy as a metric for unpredictability, and compare mobility as described by transaction records to the patterns of mobility that emerge from cell phone data. I show that two features differentiate shopping from general movement through space: the “interleaving” of store visits makes sequence lend less information to predictability in the case of shopping, and “bundling” shopping trips makes some shoppers more entropic than others.

Measures of predictability

As discussed in Part 1, “predictability” takes a number of hues. Perhaps we are interested in an individual’s overall reliability, in minimizing the error around the probability that we will find him at any particular location, in the magnitude or volume of his occasional diversions, or in his propensity for breaking and forming new patterns. Here, we use a single measure that, of course, captures a single facet of predictability.

Entropy as a measure of unpredictability

Informational entropy measures the uncertainty associated with an individual’s trajectory. A shopper with high entropy can be expected to frequent a large number of locations without a repeated sequence of shops; a low entropy shopper might visit the same two stores in quick succession every day at 5pm.

Here, we use three measures of entropy to compare the predictability of different people. We take the random entropy

$$\log_2 N_i,$$

where $N_i$ = number of locations visited by shopper $i$

to be the entropy when the likelihood of shopping at a store is the same for all stores. Likewise, the temporal-uncorrelated

$$\sum p_i(j) \log_2 p_i(j),$$

where $p_i(j)$ is the probability that user $i$ visited location $j$

entropy measures the time-independent diversity: that is, it considers the randomness across daily bins of stores visited without regard to the sequence in which they are visited. Finally, the true entropy

$$S = (1/n \sum \Lambda_i)^{-1} \log_2 n$$
where $\Lambda_i$ is the length of the shortest subsequence not previously appearing, considers the sequences of shops as well as the occurrence of shopping events over time. We use the average kolmogorov complexity to approximate the true entropy.

Song et al show that the distribution of entropies as described by cell phone locations peak around 1,3, and 6 for the true, uncorrelated and random entropies respectively [Song, 2010].

We find that using transaction data that the random and uncorrelated entropies of individuals comprise a distribution akin to that found by Song et al, albeit with higher mean entropies (Figure 11). Moreover, these entropies are largely stable over four years for the same sample of individuals (Figure 10). Yet when we incorporate the sequence of stores frequented, we find the entropy largely unchanged: that is, the diversity of an individual’s shopping behavior is driven by the number of stores he visits and the frequency at which he visits, not by the order in which he shops. This is a major departure from the findings of Song et al: the order of sites visited explains a large portion of these general mobility patterns.

Figure 10. Distribution of random and uncorrelated entropies for a single set of 2000 individuals over 6-month windows (April to September) from 2006-2009. The distribution of individual entropies remains largely constant over time.
Interleaving shopping trips

The discrepancy between the true entropy distribution of mobile phone users and of shoppers can be explained in part by the effect of shuffling or “interleaving” store visits in time: today, I might go to first to the supermarket and then to the post office, but a week hence I may reverse this ordering.

Indeed, when we run Monte Carlo simulations of the effect of novel orderings by randomizing sequence within a day, we find little change in total entropy. However, it’s possible to approach the levels of true entropy seen in the mobile phone data by sorting the order of shops visited (Figure 12) over daily or weekly intervals: that this, the presence of interleaving over the course of one or several days increases the entropy. With cell phone mobility, the sequence of events adds useful information to our estimate of predictability; the small-scale randomness of shopping brings down the true entropy distribution only by uniformly ordering shopping events. If we visited a common grouping of stores in the same, sorted order on each shopping trip, our entropy would be lower.
Figure 12. Simulation with sub-sample sequence data of effect of (a) randomizing and (b) sorting a set of stores visited over a week-long period. Averaging multiple runs of “shuffled” sequence information does not change the entropy significantly, while sorting lowers the entropy.

**Bundling**

While some of our purchases are one-off (buying a dishwasher) and some repeated (paying the electricity bill, buying the company lunch on Fridays), a portion occurs in the context of a shopping trip. We dub purchases that occur in succession on the same day *bundling*: this behavior is evidenced in anecdotal tales of shopping (“I went on a spree to the mall…” or “I had to run errands at the grocery store, post office, hardware store, and then I got a haircut…”). It also lends an explanation for the increase in entropy of some individuals (in particular the wealthy, as shown in the next section) when the number of store visits is held constant.

More specifically, we define bundling as the entropy over time bins (days) of stores visited, independent of the identity of an individual store. So, an individual who does all of his shopping on Saturday will have a higher bundling coefficient than a person who spreads the same number of store visits over the course of a week.

Figure 13. Visual description of bundling. Red is highly bundled, while blue shows a low bundling coefficient.
Under the microscope

When set alongside mobility patterns from cell phones, our consumer behavior looks much the same: we can be found with high probability at one of three or so locations, a result that holds independent of the total number of locations at individual frequents. The distribution of our temporally-uncorrelated entropies is also similar to that seen in the cell phone data. And, as seen in part 2, most of us build up a catalogue of new stores at an even pace.

The differences end at the disaggregate level. Over short timeframes, we shift the order in which we frequent the same stores, and we are inconsistent in the number of stores we visit on any given day. At this resolution, the “elective” element of shopping comes to the fore.

Figure 14. Salvador Dali at large and a distance
All consumers are predictable, but some consumers are more predictable than others

How the other half shops

We know from experience that some take for granted what others can only imagine affording once. The rich and the poor spend their salaries on very different products at very different stores, as if we inhabited altogether distinct spheres delineated by income.

Here, I consider how differences in income are linked to differences in predictability of shoppers. I set aside, for the time being, the obvious distinctions in amount spent and type of store frequented: these have been treated elsewhere. The focus is on the more basic question: do the wealthy and poor behave differently as they shop?

The rich bundle trips

We made an earlier case for the shopping trip as a unit of measurement. Trips take on a variety of stops, from the single run to the corner store, to the coupled Chinese take-out and video rental, to the epic weekend spree. It is this variety and unpredictability at the granular level that defines our economic habits.

Field studies [Lloyd 1978] and anecdote point to differences along socio-economic lines in how people shop. When looking at mobility patterns, however, Song et al find no change in entropy due to income [2010]. The authors in this case consider the median income that corresponds to the home metropolitan area of a cell phone user; here, we use the actual income associated with each individual via account inflows described in part 2.

Holding number of visits and stores constant, we find that the distribution of uncorrelated entropies varies significantly between those in the highest and lowest quintiles based on income. We attribute this discrepancy to a tendency on the part of wealthier shoppers to bundle trips, as well as variance in probability of rich and poor individuals frequenting their top stores: the poor return with higher probability to a single location (Figure 16).

Figure 15. Distribution of uncorrelated entropies of rich and poor individuals holding sequence length constant. Rich individuals show higher entropies. Distributions are significantly different (p=4x10^{-9})
Stepping up and down

The hedonic treadmill thesis has that as our salaries increase, our spending and wants increase in step [Brickman 1971]. We examine the set of shopping transactions to see the effect of gaining or losing income on spending habits, and find that individuals who lost $20-30 k income between 2007 and 2009 saw their entropies decrease by 0.05 on average without any change in average visits. Those whose incomes stayed constant had entropies stay constant (decrease by 0.001 on average, not significantly different from income decline, with $p = 0.2$).

Meet the Predictables

It is tempting to believe that predictability might itself be predicted by geography: that is, that high entropies can be explained away by the jungles of urbanity or the long distances of rural settings. Remarkably, we find little correlation between geography (US region) or population density and entropy: people across the country exhibit similar patterns.

However, we can with a touch of flippancy the “most” and “least” predictable regions, when aggregated to the two-digit zip code level. Garrison Keillor would rejoice to learn that Minnesota (zip codes beginning 55) houses the nation’s most predictable folk, while in Seattle (zip codes beginning 98), people are likeliest to deviate from the norm.
Figure 17. The “most” and “least” predictable zip codes in the United States. The most predictable location is in eastern Minnesota, and the least in Washington state. The most predictable people (regardless of location) have as their most-frequented stores grocery stores and restaurants, while the least predictable people are most likely to be found at gas stations and fast food outlets.

Figure 18. Map of predictability of a sample of US regions by zip code. Large, darker squares represent the most predictable individuals while lighter smaller squares are less predictable individuals.

In the end, we’re all driven by our fairly predictable desires: groceries, prescriptions, and gas. Sometimes we eat out.
6
Material World
The Habitat of the Economic Man

We seldom shop alone. Although many marketplaces have retreated indoors, we still share the aisles of grocery stores with other pushcarts and the patrons behind them. We wait in line at the post office, and battle hoards for the newest electronics (sometimes while still digesting our Thanksgiving turkey). Even online shopping is a many-souled system: if nothing else, prices are subtly but constantly tweaked by the lesser eddies of supply and demand.

Consumerism remains social, much as we try to reduce human interaction in the name of efficiency. At worst, separate shopping events co-occur in the same locations. At best, the mall remains the proverbial agora not just for suburban tweens but for the over-20 set as well.

And economic activity is much of what constitutes our lives in cities. We can exchange a greater variety of goods in central hubs: people drive in from rural outskirts to buy and sell. In this section, we examine the composition of 35 US cities using transaction records as lens into merchant diversity. Do some industries rise and fall alongside others? Which cities offer a comparative advantage in economic opportunities?

The Butcher, the Baker

First, I construct a matrix of pair-wise covariance in industry growth. The industrial landscape of urban America is at once variegated and self-same: we can’t go far without finding a McDonalds, and yet even small towns surprise us with a quirky shop or unknown chain. Why is it that certain sets of products and chains dominate in one region and not the next? How do cities give rise to a manufacturing or commercial specialization that draws in both labor and customers?

We find from these covariances, for example, an average increase (1.94%) in payments to management consultants was correlated with an average decline in payments to trade and vocational schools (0.92%) in cities across the US. We use this raw data to consider how the specialization of one city relative to its peers (an abundance of agricultural services, for example) relates to the overall diversity of commerce present in that city.

Industrial Portraiture

Next, I use a measure of relative comparative advantage to look at the development of American cities. Krugman paints a picture of the emergence of industrial centers as a result of low transportation costs, economies of scale, and existing infrastructure and labor for manufacturing [Krugman, 1981]. Richard Florida argues that commercial diversity is driven in part by a “creative class” of individuals inhabiting a city [2005]. We follow the model proposed by Hidalgo et al [2009] to study the co-occurrence of capability specialization and industrial diversity in American cities.
I aggregate the transactions of all Bank customers for each of 35 cities (see Figure 19 legend or Appendix for list of cities) and catalog the number of stores in each of the MCC industries represented. We then apply Hidalgo’s analysis to our own data and find the same pattern of comparative advantage via spillovers from industrial diversity, seen previously between countries of the world, holds for American cities. The cities with the greatest diversity of available merchant categories also tend to have the greatest comparative advantage from specialization.

For example, San Diego has both the highest merchant diversity relative to other cities, as well as greater relative specialization. Detroit and Saint Louis, on the other hand, have generic merchants and little comparative advantage.
Predictability Quotient

I use the above measure of industrial diversity in conjunction with the earlier measures of individual predictability, as a preliminary indicator for the “complexity” of a city. I find a slight correlation ($r^2 = 0.40$) between the individual and aggregate metrics, suggesting a link between the shopping opportunities that exist and the way people shop. That is, cities with greater merchant diversity also tend to house more unpredictable residents. I combine the two figures to create an index of complexity for US Cities based on (a) the predictability of a city’s residents and (b) its overall commercial comparative advantage.

<table>
<thead>
<tr>
<th>City</th>
<th>Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>boston</td>
<td>21</td>
</tr>
<tr>
<td>albany</td>
<td>23</td>
</tr>
<tr>
<td>atlanta</td>
<td>29</td>
</tr>
<tr>
<td>new york city</td>
<td>30</td>
</tr>
<tr>
<td>oakland/SF</td>
<td>31</td>
</tr>
<tr>
<td>houston</td>
<td>31</td>
</tr>
<tr>
<td>portland</td>
<td>32</td>
</tr>
<tr>
<td>dallas</td>
<td>35</td>
</tr>
<tr>
<td>phoenix</td>
<td>35</td>
</tr>
<tr>
<td>seattle/spokane</td>
<td>37</td>
</tr>
<tr>
<td>san diego</td>
<td>39</td>
</tr>
</tbody>
</table>

Table 2. Predictability Index for 11 US cities: $P = C \times \sqrt{S}$, where $C = \text{city complexity}$ and $S = \text{average entropy of residents}$.

I find a strong correlation between high indices of complexity and high rates of population growth between 2000 and 2007. On average, the cities with the highest indices (Houston, Portland, Dallas, Phoenix, Northern Washington, San Diego) had a growth rate of 17.6%, while the rate for the other cities, with the exception of Atlanta, was less than 8% [US Census].

In short, to maximize the randomness of interactions with your economic habitat, the old adage may hold: go west young man.
This analysis presents a small bridge between individual routines and the broad trends of commerce. Using transaction data as a lens, I find that consumers by and large adhere to predictable patterns of shopping. That is, knowing where you’ve been I can with reasonable accuracy venture where you’ll go next.

I find that routines look similar across a population. That is, the distribution of entropies is thin-peaked, and an individual’s trajectory is fundamentally the same as his neighbor’s, independent of where each person shops or how much he spends. This is a striking result: while there exist stark delineations based on what we buy, how we shop is something shared.

These results validate the mobility patterns gleaned from the locations of cell phone users: there, the same narrow-peaked entropy distributions emerge, as does the fundamental predictability of people. The trajectory of an individual is described by the same function, whether he visits 10 or 100 different locations in a week. Thus, the man who drives to and from work every day is no more predictable than the one who zigzags to client meetings all afternoon.

At the same time, transaction data reveal two deviations from the patterns seen in cell phones. First, the specific sequence adds little information to the temporally-uncorrelated set of store visits. Although an individual is likely to be at one of his two most frequented stores 40% of the time, he is constantly tweaking these broad patterns: perhaps he discovers a new lunch spot or makes a Saturday visit to the beach, or maybe his favorite after-work hangout shuts down.

The resolution of cell phone data (a location is only noted when a call is made) does not permit us to see this kind of small-scale variance. And there is way to classify the location at which a call was made; with credit cards, we know the type of merchant. Thus, the same large patterns (grocery shopping, bill paying, gas) drive our fundamental predictability, but we’re constantly discovering and forgetting places to swipe our cards.

Second, there exist income-based differences in how people distribute their shopping. While the poorest quintile returns more frequently to a top shop, the richest quintile “bundles” multiple shopping trips at once. Moreover, holding income constant, we find that the individuals with the highest entropies frequent different types of stores than their low-entropy counterparts.

Knowing the home zip codes of individuals, we are able to make inferences about the cities in which they live and shop. The diversity of merchants present in a city is strongly correlated with the emergence of specialization in industry: it has been suggested that specialization allows for the inter-industry spillovers that create new kinds of business.

There is also a striking link between the complexity of a city and the choices of consumers. In cities with a high level of relative merchant diversity (controlling for population), a more entropic distribution of shoppers is likely to be found. Much more work is needed to understand if this phenomenon is robust, and if so, why. Also, additional research might look at the dynamics of the merchant composition in cities over several years.

A great many questions remain about the formation of patterns seen in the present review: why is it that people are so predictable? First, it would be fruitful to look at the emergence of regularity for
individuals: how do shopping patterns form, how do shocks (such as income change) perturb patterns, and why do favored stores change over time?

Patterns can be learned, as well, and an analysis of shared habits would be worthwhile. That is, if you and I shop at a similar set of stores, are we likely to adopt the same new stores? Combining the present framework with an analysis of the network of customers and merchants would yield this genre of insights.

It's worth hoping that, as a bridge between the micro- and macro-, fine-grained transaction data will allow us to better chart human economic behavior in the wild, and to build better systems to sustain it.
References

Adams, PA and Adams, JK *Confidence in the recognition and reproduction of words difficult to spell.* American Journal of Psychology 1960


Babin BJ et al *Work and/or fun: measuring hedonic and utilitarian shopping value* Journal of consumer research 1994


Bettencourt L et al *Growth, innovation, scaling, and the pace of life in cities* PNAS 2005

Brand S *Environmental heresies* Conservation in Practice 2006

Brickman and Cambell *Hedonic Relativism and Planning the Good Society* 1971

Childers TC et al *Hedonic and utilitarian motivations for online retail shopping behavior* Journal of Retailing 2001


Cryder CE et al, *Misery is not miserly,* Psychological Science 2008


Dawes RM *The robust beauty of improper linear models in decision making.* American Psychologist 1979


Downs RM *The cognitive structure of an urban shopping center* Environment and Behavior, 1970


Gardyn R *Oh, the good life* American Demographics 2002

Hidalgo CA and Hausman R The Building Blocks of Economic Complexity PNAS 2009


Iyengar, SS and Lepper, MR When choice is demotivating: can one desire too much of a good thing? Journal of Personality and Social Psychology 2000


Klawitter M and Fletschner D Who is banked in low income families? The effects of gender and bargaining power Social Science Research 2010

Krugman PR Intraindustry specialization and the gains from trade Journal of Political Economy 1981

Leviticus 19:31

Lloyd R and D Jennings D, Shopping Behavior and Income: comparisons in an urban environment, Economic Geography 1978


Reilly WJ The Law of Retail Gravitation. 1931


Rohm AJ and Swaminathan V A typology of online shoppers based on shopping motivations Journal of Business Research 2004

Salganik M et al Unpredictability and inequality in an artificial cultural market Science 2006

Schwartz B The Paradox of Choice 2005

Shannon, CE, A Mathematical Theory of Communication Bell System Technical Journal 1948

Song C et al Limits of Predictability in Human Mobility, Science 2010

Sornette D Predictability of catastrophic events: Material rupture, earthquakes, turbulence, financial crashes, and human birth PNAS 2002

Sunstein C Infotopia 2006

Tetlock P Expert Political Judgment 2005

Appendix

a. List of US metropolitan areas

Albany  Durham  Saint Louis
Annapolis  Grand Rapids
Atlanta  Houston  Salt Lake City
Baltimore  Memphis  San Antonio
Boise  Minneapolis  San Diego
Boston  New York City  San Francisco
Buffalo  Oakland  Santa Fe
Charlotte  Phoenix  Seattle
Chicago  Portland  Spokane
Dallas  Providence  Syracuse
Denver  Rochester  Tulsa
Detroit  Sacramento  Wilmington

b. List of Merchant Category Codes

742  Veterinary Services
763  Agricultural Cooperative
780  Landscaping Services
1520  General Contractors
1711  Heating, Plumbing, A/C
1731  Electrical Contractors
1740  Masonry, Stonework, and Plaster
1750  Carpentry Contractors
1761  Roofing/Siding, Sheet Metal
1771  Concrete Work Contractors
1799  Special Trade Contractors
2741  Miscellaneous Publishing and Printing
2791  Related Services
2842  Specialty Cleaning
3000-3299  Airlines
3351-3441  Car Rental
3501-3790  Hotels/Motels/Inns/Resorts
4011  Railroads
4111  Commuter Transport, Ferries
4112  Passenger Railways
4119  Ambulance Services
4121  Taxicabs/Limousines
4131  Bus Lines
4214  Delivery Services
4215  Courier Services
4225  Household Goods, and Storage
4411  Cruise Lines
4457  Boat Rentals and Leases
4468  Marinas, Service and Supplies
4511  Airlines, Air Carriers
4582  Airports, Flying Fields
4722  Travel Agencies, Tour Operators
4723  TUI Travel - Germany
4784  Tolls/Bridge Fees
4789  Elsewhere Classified
4812  Telephone Sales
4814  Telecommunication Services
4816  Computer Network Services
4821  Telegraph Services
4829  Wires, Money Orders
Cable, Satellite, and Other Pay Television and Radio
Utilities
Motor Vehicle Supplies and New Parts
Office and Commercial Furniture Construction Materials (Not Elsewhere Classified)
Photographic, Photocopy, Microfilm Equipment, and Supplies
Computers, Peripherals, and Software Commercial Equipment (Not Elsewhere Classified)
Medical, Dental, Ophthalmic, and Hospital Equipment and Supplies
Metal Service Centers
Electrical Parts and Equipment
Hardware, Equipment, and Supplies Plumbing, Heating Equipment, and Supplies Industrial Supplies (Not Elsewhere Classified)
Precious Stones and Metals, Watches and Jewelry Durable Goods (Not Elsewhere Classified)
Stationary, Office Supplies, Printing and Writing Paper Drugs, Drug Proprietaries, and Druggist Sundries Piece Goods, Notions, and Other Dry Goods Uniforms, Commercial Clothing Commercial Footwear Chemicals and Allied Products (Not Elsewhere Classified)
Home Supply Warehouse Stores
<table>
<thead>
<tr>
<th>Code</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>5719</td>
<td>Miscellaneous Home Furnishing Specialty Stores</td>
</tr>
<tr>
<td>5722</td>
<td>Household Appliance Stores</td>
</tr>
<tr>
<td>5732</td>
<td>Electronics Stores</td>
</tr>
<tr>
<td>5733</td>
<td>Music Stores-Musical Instruments, Pianos, and Sheet Music</td>
</tr>
<tr>
<td>5734</td>
<td>Computer Software Stores</td>
</tr>
<tr>
<td>5735</td>
<td>Record Stores</td>
</tr>
<tr>
<td>5811</td>
<td>Caterers</td>
</tr>
<tr>
<td>5812</td>
<td>Eating Places, Restaurants</td>
</tr>
<tr>
<td>5813</td>
<td>Drinking Places</td>
</tr>
<tr>
<td>5814</td>
<td>Fast Food Restaurants</td>
</tr>
<tr>
<td>5912</td>
<td>Drug Stores and Pharmacies</td>
</tr>
<tr>
<td>5921</td>
<td>Package Stores-Beer, Wine, and Liquor</td>
</tr>
<tr>
<td>5931</td>
<td>Used Merchandise and Secondhand Stores</td>
</tr>
<tr>
<td>5932</td>
<td>Antique Shops</td>
</tr>
<tr>
<td>5933</td>
<td>Pawn Shops</td>
</tr>
<tr>
<td>5935</td>
<td>Wrecking and Salvage Yards</td>
</tr>
<tr>
<td>5937</td>
<td>Antique Reproductions</td>
</tr>
<tr>
<td>5940</td>
<td>Bicycle Shops</td>
</tr>
<tr>
<td>5941</td>
<td>Sporting Goods Stores</td>
</tr>
<tr>
<td>5942</td>
<td>Book Stores</td>
</tr>
<tr>
<td>5943</td>
<td>Stationery Stores, Office, and School Supply Stores</td>
</tr>
<tr>
<td>5944</td>
<td>Jewelry Stores, Watches, Clocks, and Silverware Stores</td>
</tr>
<tr>
<td>5945</td>
<td>Hobby, Toy, and Game Shops</td>
</tr>
<tr>
<td>5946</td>
<td>Camera and Photographic Supply Stores</td>
</tr>
<tr>
<td>5947</td>
<td>Gift, Card, Novelty, and Souvenir Shops</td>
</tr>
<tr>
<td>5948</td>
<td>Luggage and Leather Goods Stores</td>
</tr>
<tr>
<td>5949</td>
<td>Sewing, Needlework, Fabric, and Piece Goods Stores</td>
</tr>
<tr>
<td>5950</td>
<td>Glassware, Crystal Stores</td>
</tr>
<tr>
<td>5960</td>
<td>Direct Marketing - Insurance Services</td>
</tr>
<tr>
<td>5962</td>
<td>Direct Marketing - Travel</td>
</tr>
<tr>
<td>5963</td>
<td>Door-To-Door Sales</td>
</tr>
<tr>
<td>5964</td>
<td>Direct Marketing - Catalog Merchant</td>
</tr>
<tr>
<td>5965</td>
<td>Direct Marketing - Combination</td>
</tr>
<tr>
<td>5966</td>
<td>Catalog and Retail Merchant</td>
</tr>
<tr>
<td>5967</td>
<td>Direct Marketing - Outbound Tele</td>
</tr>
<tr>
<td>5968</td>
<td>Direct Marketing - Inbound Tele</td>
</tr>
<tr>
<td>5969</td>
<td>Direct Marketing - Other</td>
</tr>
<tr>
<td>5970</td>
<td>Artist's Supply and Craft Shops</td>
</tr>
<tr>
<td>5971</td>
<td>Art Dealers and Galleries</td>
</tr>
<tr>
<td>5972</td>
<td>Stamp and Coin Stores</td>
</tr>
<tr>
<td>5973</td>
<td>Religious Goods Stores</td>
</tr>
<tr>
<td>5975</td>
<td>Hearing Aids Sales and Supplies</td>
</tr>
<tr>
<td>5976</td>
<td>Orthopedic Goods - Prosthetic Devices</td>
</tr>
<tr>
<td>5977</td>
<td>Cosmetic Stores</td>
</tr>
<tr>
<td>5978</td>
<td>Typewriter Stores</td>
</tr>
<tr>
<td>5983</td>
<td>Fuel Dealers (Non Automotive)</td>
</tr>
<tr>
<td>5992</td>
<td>Florists</td>
</tr>
<tr>
<td>5993</td>
<td>Cigar Stores and Stands</td>
</tr>
<tr>
<td>5994</td>
<td>News Dealers and Newsstands</td>
</tr>
<tr>
<td>5995</td>
<td>Pet Shops, Pet Food, and Supplies</td>
</tr>
<tr>
<td>5996</td>
<td>Swimming Pools Sales</td>
</tr>
<tr>
<td>5997</td>
<td>Electric Razor Stores</td>
</tr>
<tr>
<td>5998</td>
<td>Tent and Awning Shops</td>
</tr>
<tr>
<td>5999</td>
<td>Miscellaneous Specialty Retail</td>
</tr>
<tr>
<td>6010</td>
<td>Manual Cash Disburse</td>
</tr>
<tr>
<td>6011</td>
<td>Automated Cash Disburse</td>
</tr>
<tr>
<td>6012</td>
<td>Financial Institutions</td>
</tr>
<tr>
<td>6051</td>
<td>Non-Fi, Money Orders</td>
</tr>
<tr>
<td>6211</td>
<td>Security Brokers/Dealers</td>
</tr>
<tr>
<td>6300</td>
<td>Insurance Underwriting, Premiums</td>
</tr>
<tr>
<td>6399</td>
<td>Insurance - Default</td>
</tr>
<tr>
<td>6513</td>
<td>Real Estate Agents and Managers - Rentals</td>
</tr>
<tr>
<td>7011</td>
<td>Hotels, Motels, and Resorts</td>
</tr>
<tr>
<td>7012</td>
<td>Timeshares</td>
</tr>
<tr>
<td>7032</td>
<td>Sporting/Recreation Camps</td>
</tr>
<tr>
<td>7033</td>
<td>Trailer Parks, Campgrounds</td>
</tr>
<tr>
<td>7210</td>
<td>Laundry, Cleaning Services</td>
</tr>
<tr>
<td>7211</td>
<td>Laundries</td>
</tr>
<tr>
<td>7216</td>
<td>Dry Cleaners</td>
</tr>
<tr>
<td>7217</td>
<td>Carpet/Upholstery Cleaning</td>
</tr>
<tr>
<td>7221</td>
<td>Photographic Studios</td>
</tr>
<tr>
<td>7230</td>
<td>Barber and Beauty Shops</td>
</tr>
<tr>
<td>7251</td>
<td>Shoe Repair/Hat Cleaning</td>
</tr>
<tr>
<td>7261</td>
<td>Funeral Services, Crematories</td>
</tr>
<tr>
<td>7273</td>
<td>Dating/Escort Services</td>
</tr>
<tr>
<td>7276</td>
<td>Tax Preparation Services</td>
</tr>
<tr>
<td>7277</td>
<td>Counseling Services</td>
</tr>
<tr>
<td>7278</td>
<td>Buying/Shopping Services</td>
</tr>
<tr>
<td>7296</td>
<td>Clothing Rental</td>
</tr>
<tr>
<td>7297</td>
<td>Massage Parlors</td>
</tr>
<tr>
<td>7298</td>
<td>Health and Beauty Spas</td>
</tr>
<tr>
<td>7299</td>
<td>Miscellaneous General Services</td>
</tr>
<tr>
<td>7311</td>
<td>Advertising Services</td>
</tr>
<tr>
<td>7321</td>
<td>Credit Reporting Agencies</td>
</tr>
<tr>
<td>7333</td>
<td>Commercial Photography, Art and Graphics</td>
</tr>
<tr>
<td>7338</td>
<td>Quick Copy, Repro, and Blueprint</td>
</tr>
<tr>
<td>7339</td>
<td>Secretarial Support Services</td>
</tr>
<tr>
<td>7342</td>
<td>Exterminating Services</td>
</tr>
</tbody>
</table>
7349  Cleaning and Maintenance
7361  Employment/Temp Agencies
7372  Computer Programming
7375  Information Retrieval Services
7379  Computer Repair
7392  Consulting, Public Relations
7393  Detective Agencies
7394  Equipment Rental
7399  Miscellaneous Business Services
7511  Truck Stop
7512  Car Rental Agencies
7513  Truck/Utility Trailer Rentals
7519  Recreational Vehicle Rentals
7523  Parking Lots, Garages
7531  Auto Body Repair Shops
7534  Tire Retreading and Repair
7535  Auto Paint Shops
7538  Auto Service Shops
7542  Car Washes
7549  Towing Services
7622  Electronics Repair Shops
7623  A/C, Refrigeration Repair
7629  Small Appliance Repair
7631  Watch/Jewelry Repair
7641  Furniture Repair, Refinishing
7692  Welding Repair
7699  Miscellaneous Repair Shops
7829  Picture/Video Production
7832  Motion Picture Theaters
7841  Video Tape Rental Stores
7911  Dance Hall, Studios, Schools
7922  Theatrical Ticket Agencies
7929  Bands, Orchestras
7932  Billiard/Pool Establishments
7933  Bowling Alleys
7941  Sports Clubs/Fields
7991  Tourist Attractions and Exhibits
7992  Golf Courses - Public
7993  Video Amusement Game Supplies
7994  Video Game Arcades
7995  Betting/Casino Gambling
7996  Amusement Parks/Carnivals
7997  Country Clubs
7998  Aquariums
7999  Miscellaneous Recreation Services
8011  Doctors
8021  Dentists, Orthodontists
8031  Osteopaths
8041  Chiropractors
8042  Optometrists, Ophthalmologists
8043  Opticians, Eyeglasses
8049  Chiropodists, Podiatrists
8050  Nursing; Personal Care
8062  Hospitals
8071  Medical and Dental Labs
8099  Medical Services
8111  Legal Services, Attorneys
8211  Elementary, Secondary Schools
8220  Colleges, Universities
8241  Correspondence Schools
8244  Business/Secretarial Schools
8249  Vocational/Trade Schools
8299  Educational Services
8351  Child Care Services
8398  Organizations - Fundraising
8641  Civic, Social, Fraternal Associations
8651  Political Organizations
8661  Religious Organizations
8675  Automobile Associations
8699  Membership Organizations
8734  Testing Laboratories
8911  Architectural/Surveying Services
8931  Accounting/Bookkeeping Services
8999  Professional Services
9211  Child Support - Courts of Law
9222  Entities
9223  Bail and Bond Payments (payment to the surety for the bond, not the actual bond paid to the government agency)
9311  Tax Payments - Government Agencies
9399  Government Services (Not Elsewhere Classified)
9402  Postal Services - Government Only
9405  U.S. Federal Government Agencies or Departments
9950  Intra-Company Purchases