Social Cognitive Mechanisms of Category-Based Impression Formation in Markets: The Role of Valenced Categories

***This is a working draft, please do not cite or circulate without permission***

Omar Lizardo
olizardo@nd.edu

and

Melissa Fletcher Pirkey
mpirkey@nd.edu
Abstract

We address the limitations of the standard category-based impression formation model of product/producer evaluation in markets by developing the notions of category valence, valenced expectancy violations, and identity relevance. In addition to dealing with the usual case of negative evaluations for producers that violate expectations in a negative direction, our revised model can account for cases of routine negative evaluations in cases of fit to negatively valenced categories. In addition, we are able to deal with cases in which of anomalously high(low) evaluations of producer/product claim negatively(positively) valenced categories who violate expectations in a positive(negative) direction, thus properly incorporating evaluation intensity into the theory. We discuss the implications of our model for audience/producer interactions, the reproduction of boundaries in markets, and the mechanisms generating brand loyalty.
A key insight in recent organizational theory is that the interaction of producers and audiences (including customers, regulators, critics, and other intermediaries) in market settings is structured in significant ways by categorical information (bundles of attributes presumed to hang together) and by the labels associated with these categories (Hannan, Carroll & Polos 2007; Hannan 2010; Negro, Kocak & Hsu 2010). According to this growing line of theory and research, producers and their products are subject to categorization processes, via the assignment of category labels, on the part of relevant audiences. A market category is said to exist when audiences reach consensus as to what set of producer/products fall under a given label and as to what set of category attributes are covered by the label (Hannan et al 2007). Audiences in their turn use category labels for two main purposes.

First, markets, like other institutionalized domains, are built around the establishment of cognitive routines aimed at providing predictability and stability (Hannan et al 2007; Zuckerman 1999; Zuckin and DiMaggio 1990). Accordingly, audiences use category labels in order to organize their expectations regarding the anticipated attributes of a particular set of producers/products to which the label is attached. These expectations are built into the category schemas associated with the label. Schemas, in turn, are structured representations of the set of attributes (and relations between these attributes) typical of the category. For instance, the category label “fast food restaurant” evokes a structured representation consisting of expected attributes (e.g. simple menu, fast service, relatively low prices, etc.) that establish what audiences expect to (and presumably are likely to) encounter when they interact with an organization claiming the category label. This is what has been referred to in the literature on social cognition as a category-based expectancy process (Kernahan, Bartholow, & Bettencourt 2000).
Second, the generation and dissemination of product and producer assessments and ratings is critical for the social organization of many fields, and markets are no exception (Espeland and Sauder 2007; Lamont 2012; Zuckerman 2012). Accordingly, audiences use the relative degree of fit of producer/product attributes with those associated with the label claimed by that producer to generate product evaluations (positive or negative, acceptable or unacceptable). These evaluations are crucial for the viability of specific producer/products. Producers and products that receive positive evaluations are able to thrive and establish market dominance, pushing less well-received producers to the periphery. Producers, in their turn, attempt to position their offerings within market segments where they can maximize positive product reception. The construction of producer/product evaluations from their fit (or lack thereof) to category-based information is what has been referred in the social cognition literature as a category-based impression formation process (Fiske et al 1987).

**LINKING CATEGORY EXPECTANCY AND IMPRESSION FORMATION IN MARKETS**

In the formulation provided by Hannan, Carroll and Polos (2007; hereafter “the standard model”), the impression formation (evaluation) is tightly linked to category expectancy process (fit to category schemas). In their model, the intrinsic appeal of a producer/product is a function of the relative degree of fit of producer/product attributes with those of the category specified by the label. Valence of evaluation is, then, a function of intrinsic appeal. Producers/products that meet default audience expectations by maximizing category-consistent attributes and minimizing category-inconsistent ones (i.e., producer/products that maximize degree of category membership or which come close to the category prototype) are predicted to have more intrinsic
appeal, garnering positive evaluations from relevant audiences (Hannan 2010; Hsu et al 2009; Negro et al 2010).

In the standard model, producers face a tradeoff constraint in that the ability to maximize category consistent attributes decreases with the number of category labels assigned to a producer (Hannan 2010). This means that on average, producers assigned a single (or small number of labels) are able to (on average) maximize category consistent attributes and degree of fit into the category. The model thus predicts that category specialists will tend to garner more positive evaluations from relevant market intermediaries than producer/products who are assigned a multiplicity of labels---namely, category generalists or straddlers. There is now a fairly extensive series of studies that provide empirical support for this basic prediction² (Hannan 2010; Hsu et al 2009; Leung & Sharkey 2014; Negro & Leung 2013).

**Social Cognitive Mechanisms in the Standard Model**

In this section we make explicit the social-cognitive process postulated in the standard model. The sequence begins when audiences identify producers using the category labels claimed by those producers or assigned to them by relevant industry experts, mass media outlets, or general cultural knowledge. This typification process helps audiences retrieve category attributes from long-term memory, in particular the most salient (default) or central attributes associated with the category (those that have the most weight in determining category membership). Audiences then engage in a heuristic process of matching observed producer/product attributes with default category attributes retrieved from long-term memory in order to judge degree of fit. Successful categorization is a function of degree of fit: Producer/products that fit category schemas are more likely to be successfully categorized as those who do not.
Degree of fit is itself a function of the number of category-consistent and category-inconsistent attributes displayed by the producer/product in question. Category consistent attributes increase fit, while category inconsistent attributes minimize it. Here we enter a branching point: The expected valence of the evaluation process (e.g. positive versus negative evaluation) depends on ease of categorization. Successful categorization leads to more favorable product reception, while failed categorization results in audience confusion and less favorable reception (Hannan et al. 2007).

Once the relevant socio-cognitive mechanisms are specified, it becomes clear that in the standard model the deleterious effects of multiple category membership are indirect: Multiple category membership decreases intrinsic appeal because it increases the likelihood that we will observe category inconsistent attributes in any one category; this prevents fit to the category schema, hampering easy categorization (Hsu et al 2009). Note however that there may exist a multiplicity of alternative mechanisms (or mechanism sequences) that can lead to the same outcome by affecting ease of categorization (we highlight some of those below). In this respect, category straddling—the preferred antecedent factor in recent research on categories in markets (Hannan 2010)—may be a common but by no means exclusive mechanism linking categories to evaluations in markets.

The Standard Model as a Cold Cognition Model

A key implication of the foregoing is that the Standard Model, for all intents and purposes, relies on “cold cognition” imagery (Hodgkinson & Healey 2011). Evaluations (which are at the very least a partially affective process) are the result of a cognitive process and not the inverse (Kunda 1990). Emotional valence plays little or no role in the impression formation process. Evaluation (positive or negative) is the outcome of ease of successful categorization;
confusion in failed categorization (a cognitive mechanism) is the cause (Olson, Roese, & Zanna 1996).

This feature of the Standard Model runs counter to a stream of research that links category-based expectancies directly to affect and arousal-based mechanisms. According to this work, met expectancies are usually accompanied by positive (global) affect, while violated expectancies lead to arousal and in some cases intense (negative) emotions (Biernat, Vescio & Billings 1999: 526). In addition, affective responses to expectancy violations are made quickly, with relatively little effort, and have been shown to mediate subsequent attributional judgments regarding the likely explanation for these anomalies (Kernahan et al 2000).

In this sense, the most common form of impression formation involves some level of affective valence (Fiske et al 1987; Thagard 2008). Affective valences are tightly coupled to cognitive representations of the target, attributes, and relations between attributes of a given producer/product (Hsu & Elsbach 2013). This means that the modal impression formation situation involves some level of affect, making models that presume “hot” (affectively driven) cognition more appropriate for theorizing the link between categorization processes and subsequent evaluative judgments (Kunda 1990). In this sense, evaluation valence may be as much of a product of difficulty/ease of categorization as it is of the affective valence that the attributes of the producer being evaluated have for the evaluator.

Note that it is possible that while not entirely explicit, there is an underlying “hot cognition” model in the Standard Model. Of the various theoretical efforts linking cognition, evaluation, and affect in the literature, the one that would fit the basic theoretical structure of the Standard Model would be one that linked difficulty in categorization (e.g. produced by the presence of category inconsistent attitudes) to negative affect. In this case unpredictability and
uncertainty regardless of the affective valence of the attributes that produce it, are inherently unpleasant (Olson et al. 1996). This unpleasantness then drives the devaluation of hard to categorize (e.g. category-straddling) producer/products. EVT on the other hand works with an underlying model of the psychophysiology of expectancy violation that predicts that the immediate affective response to expectancy violations is not the same for positive and negative valenced violations. Recent research bearing on this very issue (e.g. Bartholow, Fabiani, Grafton & Bettencourt 2001: 202) suggest that the EVT model and not the undifferentiated negative affect model is consistent with the psychophysiological evidence.

**Nature of the Theoretical Problem Addressed in the Paper**

This paper addresses the theoretical limitations that stem from this last feature of the Standard Model. We argue that while category-based impression formation is the default mode of engagement between audiences and producers in product markets, not all instances of category fit lead to positive producer evaluations, and not all instances of lack of fit lead to negative evaluations. We show that incorporating affective valence into the Standard Model allows for the integration of a wider range of social cognitive mechanisms linking audience expectations, producer categories and product evaluations in markets. The integration of these mechanisms leads to an elaboration of the basic model that it is able to account for variety of phenomena that are otherwise not properly handled in the current formulation.

Our basic proposal is that when evaluations depend on the affective valence and the identity relevance associated with both the attributes and the overall category label used by audiences to categorize producers, then category fit may not always lead to positive evaluations, and lack of fit may not uniformly result in negative evaluations. More specifically, incorporating valenced expectations allows us to deal with the phenomenon of anomalously high evaluations.
of producers who violate expectations by offering products of unexpectedly high quality given the average for their category ("pleasant surprises") and with the case of initially well-esteemed producers who are harshly devalued when they violate expectations in a negative direction ("unpleasant surprises"). Finally, incorporating category valence explains in a parsimonious way systematic differences in the evaluation of certain producer/products that result from their fit to categories. Of most relevance is the case of routine negative evaluations of producers who exhibit high levels fit to equally negative valenced categories.

THE ROLE OF VALENCE EXPECTATIONS

Affective Tags

Recent research in cognitive science has shown that cognition is inherently affective; rather than first recognizing and categorizing an object ("what that is") and then figuring out how we feel or stand in relation to it ("how I feel about it") objects appear to elicit automatic evaluations after single exposure at time-scales that are faster than slower cognitive processes dealing with assignment to high level categories (Damasio 1994). The world of objects thus appears to us as already partitioned into positive and negative qualities, and this partitioning affects cognition in action in systematic ways.

Accordingly, one way to incorporate the role of affective valence into the Standard Model is to think of retrieved category attributes and perceived (or experienced) producer/product attributes as affectively tagged (Thagard 2008: 19). The affective tag of a retrieved category attribute or a perceived product attribute may be positive, negative, or neutral. In this sense, category labels generate expectations for both presence/absence of different category attributes (as in the Standard Model) and for the expected valence of that attribute for
any given producer/product. This means that both fit to category schemas, and lack of fit to category schemas generate *valenced expectation satisfaction* (in the former case) and *valenced expectancy violations* (in the latter case).

Before we specify the conditions under which we may observe these phenomena, we introduce two concepts that emerge from incorporating affective valence to a model of category-based impression formation.

**Category Valence**

First, we define *category valence* as the “net” valence of a given category across all of its (retrieved) attributes. When a majority of (retrieved) attributes have positive affective tags, the category may be said to have positive valence for that audience member (or group). When the majority of retrieved attributes have negative affective tags on the other hand, the category has negative net valence. If there are roughly the same number of positively and negatively tagged attributes, the category has neutral valence. The net valence of a given producer/product category may be computed in relation to a large social group (e.g. residents of a given nation-state or multi-national region) or among smaller subgroups within the larger group (e.g. men, women, college educated persons, etc.).

The idea that categories are inherently valenced is consistent with recent research that shows that organizational categories are perceived by audiences as having “status,” such that a category may have more or less status relative to another. In that respect, the concept of category of valence is related---although not quite identical---to Sharkey’s (2014) concept of “category status.” The resemblance comes from the fact that, unless evaluators form part of a counternormative subculture, high status categories will generally be positively valenced, while low status categories will tend to be negatively valenced. The two concepts differ, however, in
their sensitivity to fluctuations in audience or subgroup identification. While a category’s status is likely to remain relatively stable over time and across audiences, the category valence of a given producer/product may change over time, or shift from positive to negative among a given subgroup at the same time as it goes from neutral to positive among other subgroups. Most brand-image management on the part of marketers and advertisers is designed to generate enduring associations between a focal producer/products with as many positively valenced attributes as possible. “Brand essence,” “brand soul,” and “emotional” schools of marketing are based on a similar principle (Holt 2004).

For the most part, we should expect that if a given producer/product is still available as an offering it is because its category attributes have a positive net valence for at least one subgroup (producer/products with negative net valence for everybody are driven off the market). The special case of producer/products that have near universal positive net valence is usually covered under (case) studies of so-called “iconic brands” (e.g. Disney, Apple, Coke, etc; see Holt 2004). However, it is clear that certain producer/products do have negative net valence for at least some substantially large segment of possible evaluators, even as they have positive valence among relatively small group of aficionados (e.g. Heavy Metal Bands).

Valenced Expectancy Violations

Second, we define a valenced expectancy violation as the case in which the attribute of a given producer product fails to match the valenced expectation for a given audience member given the category schema typically elicited by that producer label. Thus, a producer/product that displays an observed positive attribute when, in line with the associated category schema, a neutral or negative one (e.g. an extremely clean gas station bathroom) was expected, or a producer/product sporting a negative attribute when the schema specified a positive one (e.g. the
flawed Apple Maps app in iOS 6), has engaged in a valenced expectancy violation. The first case can be referred to as a positive expectancy violation, while the last case is a negative expectancy violation. Our basic claim is that in contrast to purely “cold” violations based on categorization difficulty, valenced violations may have more extreme (and counter-intuitive) consequences for the ultimate evaluation of a given producer product in markets.

**Expectancy Violation Theory**

To elaborate the basic Standard Model into a valenced-expectation violation model, we draw on Expectancy Violation Theory (EVT). EVT is a social cognition model that links valenced expectancy violation processes to the impressions and evaluative judgments likely to be generated by categorized objects (Bettencourt, Dill, Greathouse, Charlton, & Mulholland 1997; Biernat et al 1996; Jackson, Sullivan & Hodge 1993; Jussim, Coleman and Lerch 1987; Kernahan et al 2000). The basic idea of EVT is that “when an individual’s characteristics violate…[category-based] expectations, evaluations should become more intense in the direction of the expectancy violation” (Jussim, et al 1987: 537, our italics).

Under this formulation evaluation *valence* (the focus of the Standard Model), and also evaluation *intensity* vary according to whether unexpected attributes displayed by a candidate category member have a valence that deviates from those set by the overall category. Positive expectancy violations on the part of members of negatively tagged categories should receive (relatively) *intense positive evaluations* (in relation to the baseline established by members of positively valenced categories who meet expectations), while negative expectancy violations on the part of members of positively tagged categories should receive *intensely negative evaluations* (in relation to the baseline established by evaluations of category-consistent exemplars).
In social cognition research, for instance, it has been shown that members of ethnoracial categories (e.g. African-Americans) whose net category valence is negative in relation to members of the majority category but who display category-inconsistent attributes of positive valence (e.g. high grades, professional credentials, high IQ scores, etc.) are evaluated *more positively* than members of the ethnoracial majority category (e.g. Whites) who display the same positive attributes. In the same way, members of categories with net positive valence (in the ethnoracial case, Whites, or Asians) are evaluated more harshly when they display category inconsistent attributes (e.g. low grades, working class jobs, low IQ scores) than members of minority categories with net negative valence who display the same attributes (Bettencourt et al 1997; Jackson et al 1993; Jussim et al 1987; Kunda & Thagard 1996).

**VALENCED EXPECTANCY VIOLATIONS IN MARKETS**

The above analysis suggests that there should be two types of (substantively interesting) valenced expectancy violations in markets. Expectancy violations may be *felicitous* (pleasant surprises) or they may be *unwelcome* (unpleasant surprises). This formulation is consistent with psychological research on the emotion of “surprise.” In contrast to other emotions, which carry with them an explicit valence, surprise appears to be orthogonal to valence, since it is capable of being positive or negative. The affective tone of surprise is instead determined contextually, and is bound to the specific qualities of the experiences, situations, and objects that elicit it (Carver & Scheier 1998: 136-137).

A category violation is a pleasant surprise when the category-attributes elicited by a given category label have (in relation to a given audience member) a net negative valence (low expectations) and the category inconsistent attribute displayed by the producer/product has
positive valence. Thus, a restaurant that serves better food than expected given its default category assignment as a “fast food joint” engages in a form of felicitous category violation. Following expectancy violation theory, we would expect such a producer to receive more positive evaluations than a producer with a more positive label (“gourmet burger joint”) that served the same quality burger.

A category violation is an unpleasant surprise when the initial category assigned to the producer/product by an audience member has positive valence (high expectations) and the observed category attribute has negative valence. Thus, a beer that initially meets the category-attribute specifications for being considered a member of the “microbrew” category but that is later found out to be produced by a contract-brewer (Carroll & Swaminathan 2000), will receive more negative evaluations than that of a producer who made the same quality beer but did not display category-inconsistent attributes (e.g. a family-owned brewpub).

**A Social Cognitive Elaboration of the Standard Model**

In this section, we develop a social cognitive elaboration of the Standard Model that incorporates valence in categories and category attributes. This elaboration allows for unpleasant and felicitous category violations, as well as “routine” positive and negative evaluations even in the absence of category-violations. The process model shows how valuation or devaluation process may result from either good or bad fit to valenced category schemas. The first part of the process (elicitation of the category-driven impression formation) is the same as in the Standard Model. The key difference is in the first branching point, where audiences not only retrieve a set of (default) category attributes and evaluate for presence/absence, but also retrieve the valence of each attribute along with some composite valence for the category as a whole. Net category valence, as noted above, comprises both relative prevalence of negative and positive attributes,
and how each attribute is weighted by the relevant audience (see below on identity relevance and self-categorization processes). The retrieval of valenced categories sets off a valenced expectancy process, which generates a causal sequence that may result in higher (or lower as we will see below) than expected evaluations (given the predictions of the Standard Model) in the presence of either positive or negative expectancy violations.

**Routine negative (positive) evaluations without category violations.** Take for instance the case of “routine” negative (or positive) evaluations. If a critic has a net valenced negative category-based expectation for films that claim the “romantic comedy” label, then that critic will expect that watching a film that meets the category codes for the label to be a generally unpleasant experience. Films that match that category (and thus confirm expectations) will receive a (routine) negative evaluation. Routine negative evaluations result from the fact that there is a structured difference in the average evaluations given to producer/products in different market segments (e.g. high cuisine versus fast food) is consistent with the notion that category labels differ in their overall valence even before any judgment of fit is produced by the evaluator (Sharkey 2014). Market categories (like categories in other non-market settings) are ranked such that some categories enjoy higher (default) worth and prestige than others (net positive category valence) generating a (quasi) hierarchical categorical order. For instance, in the world of fine cuisine, chefs who specialize in French food routinely receive higher evaluations from critics than chefs who work on Italian or other regional/national cuisines (Leszchiner 2007). This means that valences for some categories may not begin from a “neutral” baseline, but may be “preset” to either a negative or positive dial in the absence of countervailing information.

**The pleasant surprise.** The valenced expectancy violation model predicts that when baseline expectations are negative but a given category member generates a positive expectancy
violation (audiences encounter a positively valenced attribute when they were expecting a negatively valenced one), evaluations of that product/producer will tend to be higher than usual even in relation to well-behaved (high degree of fit) members of that category. For instance, films that deviate from baseline negative (or neutral) expectations, by evincing category inconsistent attributes with positive valence (e.g. the complex plotting and three-dimensional characters associated with independent cinema in the romantic comedy *Shakespeare in Love*), will receive *intensely positive evaluations* (in relation to films with similarly weighted category attributes that are members of a category that elicits net positive expectations).

This may explain the “critics’ darling” effect associated cultural products that contain unexpected positive category inconsistent attributes *in relation* to the valenced expectations set off by their initial category assignment. The expectancy violation model can also account for the sometimes “irrationally” high evaluation (and subsequent customer allegiance) of certain producers who offer above average products in market segments characterized by producers who generate (on average) net negative expectations for product quality, such as the fast food industry (see below on “shining exceptions”). In this way, a given producer product may benefit from displaying positive valenced expectancy violations when that producer/product belongs to a category segment that has (on average) more negative net category valence in relation to other category segments in that industry. This is particularly likely in industry segments in which categories are (implicitly or explicitly) ranked, such that some categories have (default) valences that are more positive than other categories.

**The unpleasant surprise.** An analogous expectancy violation process may obtain for producer/products who belong to category labels whose attributes generate net positive expectations but who display negatively valenced attributes instead. Here, valenced expectancy
violations are punished harshly, the obverse of the “pleasant surprise” effect discussed above. Here, (e.g. highly competitive industries where expected product quality is high and valenced expectations even higher), even a relatively minor negative expectancy violation can result in harsh devaluation. Under these conditions, even relatively small negative expectation will result in disproportionately harsh devaluations of the producer/product.

This mechanism may account for the volatile evaluation dynamics of such hyper-competitive worlds as haute cuisine and Fashion in Paris and New York (Crane 2000; Leschziner 2014;). This mechanism may also account for the “tyranny” of high expectations that besets follow-up offerings by previously successful producers. Once a producer/product acquires a generally positive reputation subsequent offerings (or modifications of the product) are judged, not against a neutral category valence baseline but against one containing mostly positive default expectancies on category attributes.

This phenomenon has been observed with film sequels following a successful blockbuster (e.g. the Star Wars prequels), follow-up offerings by writer directors who begin with a smash hit (e.g. M. Night Shyamalan), or with spin off products on the part of a previously successful entrant into a new industry (even after accounting for learning effects). Note that this these sharp devaluations, may (and do) occur in the context of relatively routine positive evaluations received by established players who produce products that meet the generally positive expectations of audiences in these high-quality, high-competition industries.

It is important to note that for both pleasant and unpleasant surprises, the expectancy violation model predicts the intensity (or extremity) of the (positive and negative, respectively) evaluation to be higher than that of either routine valenced evaluations or expectancy violations whose valence matches that of the category as a whole. For instance, when members of high
status categories (which tend to have higher than average positive valence), are found to engage in deviant behavior—a negative expectancy violation—they may be met with disproportionately negative (punitive) reactions on the part of the relevant stakeholders. Sometimes the very act of category spanning is itself counted as such a negative expectancy violation (Negro and Leung 2013). This is especially likely to happen if the routine mechanisms that serve to maintain “pluralistic ignorance” (and thus the capacity of other high status peers to ignore the deviant behavior) are disrupted (Adut 2005).

Following Jussim et al (1987: 537), we propose that the attributional mechanisms of augmentation and discounting may be operative here. Both of these mechanisms are a type of contrast effect (Kunda & Thagard 1996: 293-294). An additional negative (positive) attribute for a category that contains mostly negative (positive) attributes “stands out less” than an unexpectedly negative (positive) attribute in a mostly positively (negatively) valenced category. In the former case we would expect (positive or negative) evaluations of a moderate intensity, while in the latter case we would expect evaluations of high intensity (Jussim et al 1987). Podolny (2005: 79) proposes such a discounting mechanism when he notes that when an evaluator compares two corporate actors who display similar levels of performance, she may discount the one with ties to higher status partners (who presumably had it easier) and evaluate the one with ties to low status partners more favorably (since that actor had obstacles to overcome). Finally, in the case of no expectancy violation, we would still expect to observe affect (valenced in accordance with the valence of the category) but of only mild intensity (Biernat et al 1999: 527).
IDENTITY-RELEVANCE

The Standard Model does not have a satisfactory way of incorporating structured variance in the schemata drawn upon by members of different audience subgroups, beyond observing that the “intrinsic appeal” of given offerings vary across socio-demographic niches (Hannan, Carroll & Polos 2003). This observation (while generally correct) fails to account for the specific socio-cognitive mechanisms that account for variation in evaluations across social position (Lizardo 2009). Here we propose an account that links the notions of category valence elaborated above, with a social cognition model that incorporates the “identity relevance” of certain producer/products (Carroll & Swaminathan 2000) in the form of in and out-group self-categorization processes. This is essentially an elaboration of the process-model described above, in which category valence becomes a function of the extent to which a category is affectively linked to important self-categories (social identities). We will see that this model accounts in a satisfactory way for a variety of phenomena that lie beyond the scope of the restricted Standard Model. Before getting to that it is important to introduce one last set of concepts that will play a key role in the revised formulation.

Category Identity Relevance

We define the identity relevance of a given product/producer category as the extent to which a given evaluator affectively links (in a positive or negative direction) the category label of that producer/product to the social categories with which he or she identifies. Producer/products with a net positive valence within the ingroup are said to have positive identity relevance and are more likely to be evaluated positively (Aaker 1999; Batra et al 2012; Malar et al 2011), while producers/products with a net negative valence within the ingroup have negative identity relevance. Producers or products with negative identity relevance are more likely to be...
evaluated negatively and may be used to further define the consumer’s identity through opposition (Muniz & O’Guinn 2001). Producer/products that are not seen as having any affective connection with social categories relevant for self-identity are identity irrelevant.

**Attribute Identity Relevance**

Audiences not only consider categories as a whole as relevant to their social identities as certain category attributes may also stand out as having (either positive or negative) identity relevance. In general, attributes that are perceived to be unique or especially central to product categories with positive identity relevance, will also have positive identity relevance as standalone features (e.g. craft production methods as an attribute of the microbrewery category). These are especially likely to be category attributes that are not shared with either negatively relevant or irrelevant producer/product categories.

In the same way, category attributes that are seen as specific or unique to producer/product categories with negative identity relevance will also have negative identity relevance; we should expect that attributes with negative identity relevance are perceived as not being shared with producer/product categories with positive identity relevance. Finally, attributes that are shared by producer/product categories of opposite valence (e.g. the fact that both high-end and fast-food restaurants are expected to have chairs and tables) will usually lack identity relevance.

**Self-Categorization Processes**

Under this formulation, the relevance of a given product label to audience identity depends not just on the ability to successfully categorize a given producer/product, but also on whether audiences link (positively or negatively) the producer/product category labels to social categories relevant to the self (Brewer 1999). Thus, from a social cognitive perspective, identity
relevance entails a *self-categorization* process (Turner & Hogg 1987), whereby audiences
determine whether a given producer offering is “for me” (positive connection between the
producer/product category and the self-category) or “not for me” (negative connection)
(Bourdieu 1984).

For instance, beer enthusiasts establish the relevance of a given producer label (e.g.
“microbrewery”) as “for me” when they perceive a positive connection between an available
producer/product category and a salient self-category (“knowledgeable beer drinker”). The same
identity-relevance process determines their rejection of “mass produced” beer label. Attributes
associated with this last category (technology intensive production, bland taste, low alcohol
content, etc.) are seen as incompatible with the knowledgeable beer drinker self-categorization.
Consumers who do not self-categorize as part of the “beer enthusiast” audience category on the
other hand, may consider “mass produced” beer as perfectly compatible with other relevant
social categories used in self-categorization (e.g. “working class,” “football fan”). These same
consumers may even perceive obscure microbrew labels as “not for me” (negative identity
relevance) and thus subject to devaluation.

**Motivated Evaluations**

The identity relevance of certain product categories motivates persons to engage in
certain forms of categorization, with the evaluations that emerge as a result being equally
motivated. When identity relevance is at play, categorization processes may be the product of a
form of *motivated cognition* (Kunda 1990), which is a subtype of affect-laden, hot cognitive
processes (DiMaggio 2002; Thagard 2008). For instance, persons attempt to emphasize the
positive attributes of identity relevant producers/products. They may also be more lenient when
these producer/products display negative expectancy violations (e.g. by providing mild rather
than moderate or high intensity negative evaluations; see e.g. Sharkey 2014) In the same way, negative evaluations are easier when the producer/product category has either negative identity relevance (because attention will focus on either producer/product category negative attributes or negative expectancy violations) or identity irrelevant.

**Positive Evaluation After Successful Categorization**

This is, as we have seen, the modal outcome predicted by the Standard Model (Hannan 2010). Adding self-categorization to a pure categorical conformity model allows us to account for the fact that, like in the valenced expectancy model, very different mechanism sequences may lead to this outcome (equifinality). First, a product may have low (or no) identity relevance, and yet have a high degree of fit to the category schema. In this case, the producer/product is expected to receive a positive evaluation of mild affective valence (Biernat et al 1999). This is essentially a special case of both the original Standard Model, and the our valenced expectancy elaboration of that model. Second, we may observe the case of a producer/product category that is deemed to have positive identity relevance, and the observed attributes match those of the elicited category-schema (no expectancy violation). In this case, we would expect to also observe a positive evaluation but, in contrast to the identity irrelevant case, this evaluation should be of moderate intensity (because the retrieval identity relevant product/producers generates additional positive affect).

**Negative Evaluation After Successful Categorization**

In the Standard Model, successful categorization via the display of category-consistent attributes never leads to decreased appeal. This clashes against both intuition and a large swath of research in consumer culture studies (see, Arnould & Thompson 2005 and Holt 2002 for reviews), cultural capital theory (e.g. Bourdieu 1984), and social cognitive research on
categorization and stereotypes (Jackson et al 1993; Jussim et al 1987; Kunda & Thagard 1996; Linville & Jones 1980) that shows that rejection of certain producers, products, and social categories is the result of successful, not unsuccessful, categorization. This is a key difference between a categorical conformity model that excludes valenced self-categorization processes, like the Standard Model and one that takes valenced self-categorization into consumer segments into account.

The self-categorization model predicts that only a single mechanism sequence is capable of generating this effect however (which may account why it has been hard to detect in the modal case): This is where the producer/product has negative identity relevance (“not for me”) and it fits its category schema. In this case the model predicts that the producer/product will receive a negative evaluation of high intensity. For instance, as noted above, a beer enthusiast will reject a mass produced beer that displays the entire panoply of category attributes elicited by the label (no expectation violation), precisely because it is easy to assign that producer to that category. Those producer/products that have negative identity relevance will be evaluated negatively even if (or because) their attributes match expected category attributes. The key is to specify that the evaluation valence is driven by the self-categorization/identity process not by the relative ease or difficulty of categorization (Carroll & Swaminathan 2000).

What cognitive mechanism underlies the relatively high expected intensity of negative evaluations in the presence of adequate fit to category schemas for producer products with negative identity relevance? One plausible mechanism is what has been identified in the social cognition literature as a complexity-extremity process (Linville & Jones 1980). The basic idea here is that the complexity of the category schema elicited by a producer/product label is negatively correlated with the extremity (but not necessarily the valence) of the evaluation.
Category-schema complexity is correlated with the number of attributes that are retrieved from long-term memory. A more complex schema is thus more likely to contain a mixture of both positively and negatively tagged attributes, while a simple schema is more likely to contain all negative attributes.

Where does schema complexity come from? It is well known that the relative complexity of a categorical representation increases with experience (e.g. social contact with) in the relevant domain. Insofar as audience enthusiasts have more sustained experience with producer/products that have positive identity relevance, it is safe to assume that positive identity relevant category schemas are will be relatively complex. Because they tend to avoid them, audiences will have less of a chance to develop complex schemas for products with negative identity relevance; in essence producer/product categories will in this case resemble the “stereotypes” (e.g. relatively simple and rigid categories) that are the subject of most social cognition research. Complex category schemas should be more likely to (on average) contain a more diverse mixture of positive, neutral, and even negative attributes. Simple category schemas on the other hand, are likely to contain a relatively small number of attributes sharing the same valence; which, in the case of negatively identity relevant categories, should be predominantly negative.

Thus, when faced with the task of evaluating producer/products that have positive identity relevance, audiences will tend to retrieve relatively complex category schemas. The retrieval of relatively complex category schemas reduces the expected extremity of evaluations. Producer/products that have negative identity relevance, should elicit relatively simpler category schemas (less category attributes). When audiences judge candidates within a more restricted set of evaluative dimensions, high intensity (more extreme) evaluations (e.g. most attributes have a negative affective tag) are more likely. This would explain for instance, why producer/products
that have negative identity relevance (“not for me”) tend to be not just rejected, but vehemently and categorically rejected by audiences (Carroll & Swaminathan 2000).

**Negative Expectancy Violations with Negative Identity Relevance**

The identity-relevance model also allows for the complexity-extremity process to interact with the expectancy-violation and identity-relevance mechanisms in interesting (and somewhat counter-intuitive) ways. These dynamics are not necessarily symmetric across levels of evaluative valence. For instance, consider the case of a product that has negative identity relevance and displays a negative expectancy violation. In this case, we would expect the two mechanisms to reinforce one another producing an *intensely negative evaluation*. For instance, if a fine food connoisseur who seldom frequents fast food restaurants (negative identity relevance) experiences a negative expectancy violation during a rare visit (e.g. she gets slow service) we would expect that the intensity of the negative evaluation to be fairly intense.

**Positive Expectancy Violations with Positive Identity Relevance**

Compare this to the case of an audience member who experiences a positive expectancy violation for a producer/product with positive identity relevance. In this case, the complexity-extremity mechanism (which results in the elicitation of a more complex, high-dimensional category schema) works at cross-purposes in relation to the expectancy violation one. Here, we would expect the expectancy effect to have a much weaker positive effect on the resulting evaluation than the negative expectancy effect had on the evaluation of the negative identity relevant producer/product, resulting in a positive evaluation of moderate intensity. One test implication of these consideration is that, on average, intensely negative evaluations should be more (statistically) common than intensely positive ones in various consumer markets.

**Expectancy Violations with the Different Valence as the Identity Relevance of the Category**
The cases in which the valence of the expectancy violation is the opposite of the valence of default category valence exemplify an even more complex set of dynamics. Our proposal is that in these cases, the ultimate valence (and intensity) of the evaluation will depend on the identity relevance of the category attribute that is responsible for the violation. When the attribute is identity relevant, we should observe a recategorization process (Kunda and Thagard 1996: 294), whereby the particular producer/product is recast as a deviant subtype of the original kind. When the attribute is not identity relevant, we should simply observe a (mild) evaluation in the direction of the expectancy violation. The latter is the case for actors who are assigned to (positively valenced) categories of generalized high esteem, when the sources of such high esteem do not come from a motivated self-categorization process.

We predict a mild intensity in this case because we presume that the category-valence mechanism (which pushes evaluations in the direction of the valence) works to dampen the expectancy violation effect. Here our model makes predictions that are indistinguishable from a pure “category status” model (Sharkey 2014). However, the category-status model cannot account for the case of an intensely negative evaluation (what we refer to below as a “black sheep” effect) that follows a negative expectancy violation on the part of a producer/product that belongs to what was an initially identity relevant category (Adut 2005). We argue that the reason for this is that the identity relevance (and thus valence) of the attribute that produces the violation needs to be incorporated into a process model in order to distinguish between the two mechanisms sequences, and thus the two seemingly opposed outcomes.

Expectancy Violations for Identity Relevant Categories and Identity Irrelevant Attributes

Let us begin with the case of a negative expectancy violation in the case of a producer/product category with positive identity valence. How does the evaluation depend on the
identity relevance of the unexpected attribute? Consider for instance, the case of a coffee enthusiast who goes to a newly opened coffee shop that has received rave reviews for its quality brews (positive identity relevance), but finds that, while the coffee is indeed excellent, the service is slow and rather uninspired (expectancy violation with a negative affective tag). Because speed of service is not an identity relevant attribute for the “coffee enthusiast” self-category, we predict a mild negative evaluation. The case of a producer/product category that has negative identity relevance but generates a positive expectancy violation is precisely parallel to this case (predicted mild positive evaluations). Note that in the former case, this model does predict a generally “protective” effect of status to the social consequences of “deviance” (Sharkey 2014).

**The Black Sheep Effect**

In the case of a negative expectancy violation for a producer/product category with positive identity valence, where the expectancy violating category attribute has positive identity relevance, our model predicts that we should observe what in the social cognition literature is referred to as the *black sheep effect* (Biernat et al 1999; Marques & Paez 1994; Marques, Yzerbyt & Leyens 1988). This is a type of intense negative evaluation for what are perceived to be ingroup members with negative attributes (Marques et al. 1988). The extremity of the evaluation is the result of the threat that apparently deviant members of the ingroup present for the overall category valence (essentially an undesirable shift from a positively to a negative valued category) of the social identities (self-categories) with which the person identifies (with the motivational impetus provided by the self-esteem mechanism). The black sheep effect is in essence a fine-grained specification (in terms of socio-cognitive mechanisms) of a special case of the “unpleasant surprise” discussed above.
For instance, individuals high in group identification have been shown to provide lower evaluations of unfavorable ingroup targets than unfavorable outgroup targets (Biernat et al 1999). More importantly, it has been shown that the extremity of the negative evaluation depends on the attribute that violates the expectation: negative evaluations are more intense when ingroup deviants are believed to have violated norms more centrally associated with the ingroup category than when the violate norms that are not associated with the ingroup category (Marques et al 1988). In our formulation, expectancy violations based on identity irrelevant (or negatively relevant) attributes map onto deviance not associated with the ingroup category norms while expectancy violations based on identity relevant attributes map onto expectancy violations of deviance associated with ingroup category norms.

Our basic proposal is that we should observe black sheep effects in market categories, and they should be driven by (structurally) analogous mechanisms (Lizardo & Pirkey 2014; Vaughan 2014): ingroup identification corresponds to identity relevance of the overall category, and violation of an ingroup norm corresponds to an expectancy violation on an identity relevant producer/product category attribute. The key prediction is that upon encountering a producer product with positive identity relevance who produces a negative expectancy violation with respect to an identity relevant category attribute, audiences will engage in a recategorization process, whereby the expectancy violating producer/product will be recast as a deviant (atypical) instance of the overall (still positively relevant) category. Note that this in this mechanism sequence it is the negative expectancy violation (and related negative affect) that generates the typicality judgment (and the subsequent evaluation). This is distinct from the Standard Model, in which atypicality alone is seen as responsible for negative evaluations.
As an illustration, compare the coffee enthusiast example above to the case of an enthusiast of the microbrewery movement who, upon encountering producers that seem to meet the specifications of the microbrew category in terms of appearance, taste, and so on (positive identity relevance), is faced with an expectancy violation with a negative affective tag (e.g. formal ownership by a mass producer or identity as a contract-brewer). In this case the model predicts that this audience member will recategorize the producer/product as a deviant instance of its initial kind (‘fake microbrew”). The vehemence with which large-scale brewers who attempted to enter the microbrewery market segment were rejected by insiders is consistent with the black sheep dynamic in terms of both the valence and the intensity of the relevant evaluations (Carroll and Swaminathan 2000).

The shining exception effect. The shining exception effect, is the obverse of the black sheep effect. Here a category with negative identity relevance, generates a positive expectancy violation with respect to an identity relevant attribute. Recall that an identity relevant attribute is one that is shared with identity relevant product categories. This would entail for instance, encountering a producer/product with negative identity relevance (and thus baseline negative category valence) that displays a surprising attribute usually associated with identity relevant (and thus positively valenced) product categories (e.g. a fast food restaurant that serves fresh, tasty food from all natural ingredients). To make sense of this anomaly, the identity relevance model predicts that audience will engage in a recategorization process, that will render (as with the black sheep) that producer/product a “deviant” (atypical) example (usually a subtype) of its original (still negatively relevant) kind: a shining exception.

As noted above when discussing pleasant surprises, it is likely that the shining exception dynamic accounts for why there exist certain brands with exceptional customer loyalty in
otherwise low status (negative category valence) industries. A shining example (pun intended) is the In-N-Out brand, long considered anomalous for being a “chain restaurant with a cult following” (McNichol 2002). We would argue that this is a result of the fact that In-N-Out generates positive expectancy violations when it comes to identity relevant attributes for a specific audience segment. In particular, the fact that the In-N-Out brand reputation is built on their self-proclaimed belief to make food from scratch and entertain every whim of their prospective customers strikes us as particularly important. In-N-Out restaurants are not allowed to freeze their food; all meals are prepared upon order; and fries are cut by hand (a nod to “artisanal” food preparation). All of these expectancy violating attributes in relation to the “fast food restaurant” category schema (fresh food, manual preparation, and a customizable experience) are identity relevant attributes for potential customers that would typically stay away from burger chains (possibly the same set of customers that would stay away from mass produced beer). Even the existence of a “secret menu” available to customers with the proper know-how, is a positive expectation violation on an attribute that is likely to be identity relevant to high-cultural capital audiences, who value the possession of such insider knowledge (see e.g. Holt 2002).

**DISCUSSION**

**Incorporating an Affective Dimension the Standard Model of Category-Based Evaluations**

In this paper, we elaborate the standard model of the ways that categories link to producer/product evaluations in markets. According to this model, the valence of evaluation is a function of fit to category schemas, such that any mechanism that reduces fit (e.g. category-straddling) is also likely to result in negative evaluations. Under this model, category schemas
serve as a backdrop for the generation of “cold” category expectancies (matching observed attributes to category attributes), with the only possible outcomes being “fit” or “lack of fit.” This model has received an impressive amount of empirical support (see Hannan 2010 for a review) but the range of evaluative outcomes that it can account for is limited. For instance, the standard model is unable to account for instances in which poor fit due to code violations or category spanning does not lead to negative evaluations (Durand, Rao, & Monin 2007) or instances when good fit results in negative evaluation (Carroll & Swaminathan 2000), among other cases.

We argue that only a model that incorporates affective valences (and thus moves beyond the “cold cognition” approach implicit in the standard model) can account for the full range of evaluative outcomes of interest. To that end, we draw on recent research in social and cognitive psychology that shows that categories schemas (and the attributes that compose them) are affectively tagged, and the expectancies generated by schemas have a necessary affective valence, with “neutral” valence being an unusual special case. We introduce the concepts of category valence and valenced expectancy violation as a way to elaborate the basic mechanisms postulated in the standard model. In our model, all market categories: 1) possess a net affective valence (positive, negative, or neutral), 2) the affective valence of a category generates valenced expectations regarding the presence or absence of positive and negative attributes for any related producer/product, 3) the audience responses to both fit and lack of fit to category schemas are valenced, and 4) valenced responses can vary in intensity.

Drawing on expectancy violation theory, our basic proposal is that the ultimate valence of the evaluation of a producer/product that does not fit category schemas goes in the direction of the expectancy violation. And that the extremity of the evaluative judgment is a function of an
inconsistency between the default category valence and the valence of the attribute that produces the evaluation. Positive expectancy violations generate positive evaluations, and negative expectancy violations generate negative evaluations. Accordingly, even episodes of fit to category schemas are themselves valenced: fit to categories which are positively valenced result in positive evaluations; however, fit to categories that are negatively valenced result in routine negative evaluations.

Most importantly, “surprises” (both pleasant and unwelcome) generate intense judgments on the direction of the surprise. Producer/products who belong to negatively valenced categories (e.g. low status industry segments) but who display an unexpected negative attribute are likely to receive disproportionately positive evaluations, while producer/products with default positive valence who display a negative attribute are likely to be harshly devalued. Accordingly, we propose that, contrary to the standard model, expectancy violations produced by lack of fit to the category schema may increase audience appeal (if the direction of the violation is positive), and that producer/product from well-regarded categories have more to lose when they disappoint audience expectations than those who belong to neutral or negatively valenced product categories.

Our expanded model improves upon the standard model in important ways. Whereas the standard model allows only for positive or negative evaluations (and thus tacitly assumes that all evaluations are of the same strength or intensity), the expanded model can account for degrees of positivity or negativity in evaluations. In this sense, it appears that empirical tests of the standard model may have been hampered by restricting themselves to category domains with either neutral or mildly positive valence (Sharkey 2014), which preclude identification of pleasant surprises (which require relatively negatively valenced categories). This is not surprising given
that any study that restricts itself to mature industries or established market settings is unlikely to come across negatively valenced producer categories (as these would be selected out by the market process itself). Additionally, the standard model suggests that both positive and negative evaluations are each the result of a single mechanism sequence. In contrast, the expanded model allows for a variety of mechanism sequences that lead to the same outcome.

Implications for Research on Audience/Producer Interactions in Markets

Categories research in organization theory has relied upon indirect measures of audiences’ perceptions of boundaries (Negro, Kocak, & Hsu 2010), tacitly lumping together multiple audiences in the process. This assumption also justifies reliance upon expert audiences as broadly representative of trends in valuation across multiple audiences (Durand, Rao, & Monin 2007; Rao, Durand, & Monin 2003). Producers, however, frequently face multiple audiences, either through (1) segmentation by level of expertise or insider/outsider status (Zuckerman 1999), (2) membership in multiple categories (Hsu, Hannan, & Kocak 2009), or a combination of the two (Durand et al. 2007). This means that, as a rule, audiences cannot be “aggregated meaningfully” (Carroll & Hannan 2000:70), leaving them to face conflicting demands and expectations (Hsu & Hannan 2005; Negro et al. 2010). While more recent work in categories research has begun to account for audience segmentation (e.g. Pontikes 2012), these audience typologies are not theoretically motivated but are instead fitted to the specific substantive domain under consideration (e.g. “market takers” versus “market makers”).

Our valenced category model provides a more theoretically motivated elaboration of audience heterogeneity in the category-based evaluation of producer/products in markets, with an explicit link between the audience typology, category-based expectancy violations, and evaluation outcomes. We introduce the concept of category identity relevance to conceptualize
this heterogeneity. In general we propose that audiences partition themselves into clusters depending on whether they consider both a set of overall market categories (and the attributes associated with those categories) as relevant for processes of self-categorization into groups (e.g. “comic book fan”). Drawing on theory and research of social cognition, we propose that 1) categories and associated attributes may have identity relevance, 2) that identity relevance is valenced in that it entails a self-categorization process in which audience members determine if a given offering is positively or negatively related to a given social identity, 3) under the condition of identity relevance, actors’ evaluations may be the result of motivated cognitions.

**Implications for Research on the Reproduction of Category Boundaries in Markets**

Our proposed model sheds light on the mechanisms that reproduce boundaries in markets. In particular, our approach shows how market boundaries persist in spite of the existence of “chronically deviant” producer/products. We propose that via recategorization processes, audiences are able to maintain the overall category valence of identity relevant market categories (either positive or negative) while allowing for consistent exceptions. For instance, note that both the black sheep and shining exception effects, rather than serving to blur category boundaries, actually reinforce them. The reason for this is that both black sheep and shining exceptions are as a whole much more likely to be seen as atypical members of their respective categories (Kunda & Oleson 1997).

Even repeated episodes of positive expectancy violations in relations to producer/products with negative identity relevance are unlikely to change the overall relevance valence of the category to which they belong, as long as that producer/product is seen as an atypical instance of the overall category schema. This is consistent with social cognition research that shows that when persons who are exposed to individuals who are also judged to be atypical
members of an outgroup, they are less likely to modify their original evaluation of the category as a whole (Castano, Paladin & Yzerbyt 2002: 367). Our argument is that a similar mechanism operates in markets. This accounts for the stability of hierarchies of status worth in market categories even in the face of apparent violations by actors that seem to defy those boundaries or who display attributes whose valence is inconsistent with those of the category as a whole.

**Implications for research on marketing and brand loyalty**

Our model should be of interest to both scholars and practitioners in marketing because it is able to shed light on the mechanisms that result in both the ultimate valence and the *intensity* of evaluations. In particular, our model complements and extends existing research and theory regarding the affective outcomes of consumer-product/producer identity congruence. Furthermore, the notion of identity relevance that we introduce extends this reasoning from the realm of *personal* identity to that of *social* identity (Stets & Burke 2000). Within the marketing literature there is substantial support for the proposal that congruence between a consumer’s identity and that of a producer or product generates positive affect (Aaker 1999, Grohmann 2009; Malar 2011). That is, when a consumer perceives a producer or product to be “like me” or “for me” their relationship to that producer/product becomes positively valenced. At the most intense levels, consumers may develop a sense of ‘brand love’ for particular producers that are highly congruent with their identity, and this experience is related to higher evaluations of the brand and its products, as well as a resistance to negative information about the brand (Batra et al 2012). This research however, treats the link between person and the product as driven by exogenous preferences. Our model links the “for me/not for me” identity relevance process to the differentiation of social groups based on patterns of cultural choice (Bourdieu 1984; Peterson 1992). Our model opens up avenues of future research linking “for me” and “not for me”
judgments to higher level identification with members of consumption communities and identity movements.

---

1 A previous version of this paper was presented at the Seventeenth Annual Meeting of Organizational Ecologists, Barcelona, Spain (2014).

2 An alternative explanation for the negative evaluation of category spanners is that, by virtue of being spread across multiple categories, producers are spread too thin and thus produce products of objectively lower quality than their specialist peers. While there is evidence that such a producer-side effect occurs (Freeman and Hannan 1983), recent research by Negro and Leung (2013) and Leung and Sharkey (2014) demonstrates that even after accounting for differences in quality a “perceptually driven penalty” is applied to category spanning producers/products.

3 As in so-called trickle down fashion phenomena (Simmel 1957).


5 See http://www.businessinsider.com/m-night-shyamalan-movies-keep-getting-worse-2013-6. Note that while some sort of “regression to the mean” is a perfectly acceptable “technical” explanation for the subsequent decline in performance, this does not explain the disproportionately negative reaction that such regression engenders.
REFERENCES


Cambridge: Cambridge University Press.