Evaluative Schemas and the Mediating Role of Critics

Greta Hsu
Graduate School of Management, University of California, Davis, Davis, California 95616, grhsu@ucdavis.edu

Peter W. Roberts, Anand Swaminathan
Goizueta Business School, Emory University, Atlanta, Georgia, 30322
{peter_roberts@bus.emory.edu, anand_swaminathan@bus.emory.edu}

How do critics enable producers and consumers to come to mutually agreeable terms of trade? We propose that critics offer more guidance to those who set prices when their quality assessments are structured by clearer evaluative schemas. Schema clarity enables producers to accurately anticipate the quality assessments that critics will disseminate to the market. This allows their posted prices to center more faithfully on prevailing conceptions of quality. We then argue that the position of a producer within the market’s social structure—in terms of its prior coverage, reputation, and niche width—infuses the degree to which it is guided by clear evaluative schemas. We test these predictions in the market for U.S. wines. After elaborating a novel approach to inferring the clarity of evaluative schemas within different varietal categories, we demonstrate that list prices are less variable around expected levels when the schemas used to evaluate quality are clearer. Moreover, this effect is stronger among more relevant and more focused producers in each category.

Key words: mediated markets; critics; categories; sociology of markets

History: Published online in Articles in Advance March 23, 2011.

Introduction
Market categories help actors negotiate complex market realities by providing simplifying frameworks that structure otherwise problematic exchanges. Their specific implications are best understood in two-stage models of competition that center on the interface between “candidates” (i.e., producers) and “audiences” such as critics, analysts, and end consumers (Zuckerman 1999). In the first stage, producers vie for audience attention. Producers and products that locate clearly within established market categories are easier for audience members to identify and are therefore more likely to gain attention. In the second stage, producers try to differentiate themselves from others in their consideration sets in ways that audience members find more appealing.

Applying this model, several studies document the implications of categorical boundaries on market outcomes. Actors that that bridge or straddle established market categories tend to be ignored (Zuckerman 1999, 2000; Zuckerman and Kim 2003), deemed less appealing (Hsu 2006b, Rao et al. 2005), or otherwise devalued (Hsu et al. 2009). Zuckerman (2004) extends this basic orientation by showing how stocks that do not fit neatly into analysts’ categories are subjected to a greater diversity of interpretive models. As a result, they experience greater trading volume and price volatility in response to the arrival of new information about them.

In contrast, less attention has been paid to the role played by categorical understandings in the second stage of competition. This is somewhat surprising given that clear evaluative schemas are necessary precursors for agreements on terms of trade and thus for the persistence of markets themselves (Podolny 1993, 1994; White 1981). Yet we currently lack an understanding of how the evaluative schemas that form the basis for differentiation influence market outcomes.

We address this gap by building on recent sociological interest in the roles that mediators such as critics play in structuring the interface between consumers and producers (Hsu 2006a; Rao et al. 2005; Shrum 1991; Zuckerman 1999, 2000). On the consumer side, products that receive favorable critical evaluations are deemed better values and are therefore preferred by consumers. This leads to the often-found positive relationship between quality ratings and prices (Caves and Greene 1996). The influence of critics is particularly strong in markets for restaurant meals, movies, books, and wine, where one cannot know the quality of a good until it has been purchased and consumed. Therefore, consumers must rely on critics’ published reviews as proxies for hidden product quality. Critics also influence consumers in their role as gatekeepers, directing the attention of consumers toward certain offerings and away from others (Shrum 1991, 1996; Zuckerman 1999).

Critics also provide guidance to producers, who must adjust resource allocations, divest business units, and adopt specific missions in order to conform to the standards espoused by these mediators, precisely because of their subsequent influence over consumers (Epseland and Sauder 2007, Zuckerman 2000). Such research draws our attention to one of the most pervasive means
by which critics structure market outcomes: they establish and then apply the evaluative schemas that allow producers to effectively differentiate themselves from one another. This issue has been studied in various cultural domains (Greenfeld 1988, Shrum 1996, White and White 1993). In art worlds, for example, Becker (1982, p. 134) observes that critics often create “coherent and defensible aesthetic systems” that guide producers in their work. When their evaluative schemas are accepted by others, critics provide stability to art worlds by facilitating interactions among its many different participants.

In this paper, we propose that the language that critics use when communicating their quality assessments provide cues about the evaluative schemas that underpin them and therefore conditions their influence on producers’ pricing decisions. In categories where schemas are clearly communicated, there is a tangible basis for producers to anticipate critics’ reactions to their offerings. This allows their list prices to faithfully reflect the quality ratings that will be published after their products reach the market. We then examine differences among producers in the degree to which their pricing is guided by clearly communicated evaluative schemas. We accordingly elaborate our baseline account of the order-generating effects of clear evaluative schemas by predicting which producers are more likely to be influenced by them.

The setting for our study is the market for U.S. wines. This market has several features that are advantageous for our purposes. Wine critics are known to exert considerable influence over consumers’ purchase decisions (Landon and Smith 1997, Roberts and Reagans 2007). As a result, wine producers have an interest in setting prices that reflect the critics’ quality judgments. If producers set prices too low, they will not benefit from higher prices associated with superior quality. If they set prices too high, their products will suffer from lower or slower sales. However, wine prices are typically set before critics evaluate the products. This means that producers cannot simply align their prices with already-published ratings. Instead, they need to understand the critics’ underlying evaluative schemas. In light of this need for producers to anticipate critical evaluations, it is important that wine critics publish their quality ratings alongside written justifications in the corresponding tasting notes. This, we will argue, makes evaluative schemas accessible to producers seeking pricing guidance.

The overall market for wine comprises different varietal categories (e.g., merlot, pinot noir, chardonnay), and specific notions of what constitutes wine quality differ across these categories (Laube 1999). Varietal categories also differ in the clarity of critics’ discourse regarding what constitutes high quality. We take advantage of this variation and compare the effects of clarity in evaluative schemas on pricing variability across varietal categories, as well as across producer positions within categories.

In the following section, we outline our theory and the pricing order predictions that are implied. After describing an approach that we developed to infer the clarity of evaluative schemas, we demonstrate that clearer schemas are associated with significantly lower variance in list prices around those predicted by a robust model of wine price. This pricing order effect is weaker among producers that have not attracted prior critical coverage, have unfavorable reputations for quality, and participate across a wide range of varietal categories. Our results and their implications are discussed in the concluding section of this paper.

### Evaluative Schemas and Orderly Pricing Within Markets

In candidate-audience interface models, candidates (or producers) vie for the favor of the audience (or consumers), who selectively choose the producers they wish to engage with (Zuckerman 1999). As part of the competition to reach consumers, producers must also court the market’s influential mediators, who selectively grant their attention and occasionally their approval.

Numerous decisions and commitments are made in anticipation of these subsequent reactions. Publicly traded firms divest business units and change their corporate strategies in the hope of winning analysts’ subsequent coverage (Zuckerman 2000). Movie studios decide how much to invest in making and marketing a film and how widespread its distribution should be before film critics conduct their reviews (Ravid 1999). Restaurant owners make staffing decisions and set menus before they know how food critics will rate their offerings (Rao et al. 2003).

In all of these examples, producers require some understanding of the evaluative schemas held by mediators in order to effectively make their decisions. Sometimes, the criteria used to arrive at overall quality judgments are explicitly stated. In the market for diamonds, for example, the quality appraisals provided by experts (e.g., the Gemological Institute of America or the American Gem Society) come in the form of a certificate, which is “a ‘blueprint’ of a diamond [that] tells you the diamond’s exact measurements and weight, as well as the details of its cut and quality. It precisely points out all the individual characteristics of the stone” (see the Diamond Buying Guide). At other times, producers are hindered because they are only exposed to the final outcomes of mediators’ deliberations. Waguespack and Sorensen (2011), for example, describe how the Motion Picture Association of America deliberately avoids providing rationales for the parental advisory ratings they issue for films. This creates ambiguity for film producers who seek to balance the competing goals of providing edgy, graphic material and obtaining the least restrictive rating.
In intermediate cases, mediators issue detailed reviews that are full of description and analysis that lack a clear underlying structure. For example, from a New York Times review of the restaurant Babbo, one knows that it received three out of a possible four stars and that it was among the New York Times’ “Top Picks.” However, although there may be hints, it would be difficult for a restaurateur to discern from the following review exactly what criteria this judgment was based on:

Some restaurants revel in exquisite subtleties. Babbo, blessedly, goes straight for the gut, adding one big taste sensation atop another, gilding already delicious dishes with extra bits of texture and final flourishes of flavor. It’s doing this now as well as ever. Although its co-owners, Mario Batali and Joseph Bastianich, have branched out considerably since they opened Babbo six years ago, the restaurant remains their cherished center of gravity, and the proof is in the pasta, always perfectly cooked. Beef cheek ravioli comes with crushed squab liver and black truffles; goose liver ravioli is topped with browned butter and balsamic vinegar. There are adventurous selections like lamb’s tongue, but there are also lamb chops, a rib-eye steak and plenty of fish. Save room for dessert—in particular, the pine-nut crostata. But also know, going in, that Babbo can be frenetic and noisy, with Led Zeppelin on the soundtrack, and that the slightly jarring atmosphere traces one of the dividing lines between a restaurant with three stars, which it unequivocally deserves, and one with the highest rating of four. (Bruni 2004)

When the basis for quality judgments lacks structure, there is variability in how much guidance producers are able to glean. This is evidenced in a passage that describes the fear that the owner and employees at Babbo felt when they learned that their restaurant was about to be reappraised by the New York Times food critics: “The fear isn’t that a critic might have a personal agenda; it is merely that judgment is unpredictable and sometimes arbitrary, even if its consequences can be absolute: if your restaurant gets trashed, for whatever reason, your trade will suffer” (Buford 2006, p. 307, emphasis added).

This suggests that, across markets or across categories within a market, one may observe differences in the clarity of critics’ evaluative schemas. The absence of clear schemas has serious implications for market stability. As Velthuis (2005, p. 122) notes, producers in such contexts face considerable uncertainty because “shared standards of value are lacking, while the willingness of [consumers] to pay is almost impossible to estimate.” We focus on the specific issue highlighted by Velthuis and ask whether clarity in the evaluative schemas that underpin critics’ quality assessments has implications for the predictability of pricing behavior across market categories.

As noted earlier, products that are rated more favorably are generally perceived as higher quality and are therefore able to command higher prices. As a result, producers must look to critics when setting prices in order to align their decisions with the quality information that will later influence the consumer side of the market. Therefore, producers have an interest in monitoring the critics and their evaluative schemas for the guidance that generates orderly pricing. As these evaluative schemas become clearer, critics’ quality judgments go from being effectively random (i.e., “One does not know how good this is until after the critic tells us”) to being more predictable (i.e., “Any individual with product expertise can form a reasonable forecast of the critic’s judgment”). This latter scenario implies more guidance, which is reflected in a stronger correspondence between the prices that producers set and the quality scores that are subsequently published by the critics. Therefore, we predict the following.

**Hypothesis 1 (H1).** Producers’ list prices will be less variable around expected levels when critics’ schemas for evaluation are clearer.

This main prediction is prefaced on an important assumption: producers have an interest in setting prices that reflect the critics’ quality judgments because these judgments will influence the decisions of targeted consumers. Of course, there is likely to be variation in this respect. In particular, producers may differ in the degree to which they actively seek to be considered by consumers who are interested in quality and therefore influenced by critics’ ratings. Accordingly, a refinement of (as well as a test of the validity of) our main prediction considers which producers feel part of the quality-oriented segments of the market.

We first consider producers that have not yet been reviewed by the critics. These producers are less responsive to increases in the clarity of critics’ evaluative schemas for two reasons. First, by not providing coverage to a producer, critics have effectively placed that producer outside of the quality-oriented segments of the market (Roberts and Reagans 2007). There is less chance that its current products will be reviewed and less chance that consumers will pay attention to those reviews even if they are published. This lowered expectation of critical scrutiny lessens the producer’s incentive to monitor the critics’ ratings and reviews.

A secondary mechanism has to do with learning. Prior reviews provide a producer with direct exposure to how critics have perceived products and what dimensions have been emphasized. This provides covered producers with a more informed sense of how current products are likely to be evaluated. Together, both mechanisms lead to the following prediction.

**Hypothesis 2 (H2).** Relative to producers with prior critical coverage, those that have not yet received coverage will experience a smaller decrease in pricing variability around expected levels when critics’ schemas are clearer.
Critics also marginalize producers by providing consistent indication that their products are of very low quality. As negative ratings accumulate, a producer develops a reputation for poor product quality. This negative signal forces that producer to look for alternative ways to sell its products—ones that sidestep the stigma of future bad critical reviews, like advertising campaigns that emphasize history, availability, region, or other non-quality attributes. This logic resembles that which underpins middle-status conformity theory, wherein producers of low status have less incentive to conform to established market schemas (Phillips and Zuckerman 2001). As relegated producers seek alternative ways to effectively promote their products, they are less likely to monitor the critics and thus have a diminished desire to be influenced by them. Therefore, we predict the following.

HYPOTHESIS 3 (H3). Relative to producers with better reputations, low-reputation producers will experience a smaller decrease in pricing variability around expected levels when critics’ schemas are clearer.

A third moderating factor relates to having products reviewed across a range of product categories. The overall identities and reputations of these generalists are less tied to their quality ratings within any one category. This is likely to have an impact on the price that typical consumers are willing to pay for a product in any one category. Therefore, relative to their specialist counterparts, generalist producers will be less dependent on critics’ views about quality in any given category in their pricing decisions. This decreases their incentive to follow and adhere to critics’ schemas for evaluation in the focal category.

Further contributing to this effect is the fact that different product categories tend to reflect different combinations of product characteristics (Zuckerman et al. 2003). As a result, the schemas employed by critics when evaluating quality are also different. Monitoring and discerning the basis for critics’ ratings and reviews requires experience and takes time and effort. Because evaluative schemas vary across categories, effort spent monitoring critical reviews and ratings in one category does not necessarily translate into knowledge or skills that are portable across categories. The logic here comes from research on market categories. Producers that work across categorical boundaries run the risk of incurring penalties related to consideration, valuation, and experience accumulation. This last experience penalty (Hsu et al. 2009, Negro et al. 2010) is salient in the current context. As a result, we predict the following.

HYPOTHESIS 4 (H4). Relative to more specialized producers, producers whose reviews are spread across a range of product categories will experience a smaller decrease in pricing variability around expected levels when critics’ schemas are clearer.

Quality Judgments and Evaluative Schemas in Markets for Wines

The emergence of the modern wine critic followed the rise of the United States as a serious wine-consuming country in the early 1960s and then as a serious wine-producing country by the late 1970s. U.S. wine markets lacked the historical and institutional support to offer much guidance to the new consumers who were trying to address the problem of hidden quality information. They concomitantly lacked the ability to provide guidance to the many new producers vying to compete and set prices based on prevailing conceptions of quality. The Wine Spectator was launched into this void in 1976. In 1979, it was acquired by Marvin Shanken, who knew that “consumers wanted to know what to buy, and for the Wine Spectator to succeed it would have to tell them. That meant a tasting panel and blind tastings” (McCoy 2005, pp. 81–82). Since its reinauguration, the Wine Spectator has been widely regarded as one of the most influential critical voices in the U.S. wine industry.

Shrum (1991, p. 352) distinguishes the analytic elements of a review from its evaluative elements as follows: “Analytic elements provide an interpretive context for understanding,” whereas “evaluative elements are positive or negative judgments, often distinct from the descriptive passages of the review.” Wine reviews typically adopt a structure that contains both of these elements. The evaluative element is summarized in a numerical quality rating. According to the Wine Spectator (2004, p. ix), this rating reflects “how highly our tasting panel regards each wine relative to other wines.” The review’s tasting note houses its analytic elements. Although both the evaluative and analytic elements of a review speak to wine quality, the former offers a summary judgment of product quality, whereas the latter tries to articulate the basis for this overall impression.

The 100-point scoring system was introduced by Robert Parker, who saw these numbers as “convenient shorthand” (McCoy 2005, p. 133). His approach was quickly adopted by the Wine Spectator. The popularity of this system sparked a process of commensuration (Epseland and Stevens 1998) that has in a relatively short period of time condensed the multidimensionality of wine quality to the single quality score (Podolny and Hsu 2003). These ratings have come to exert a huge influence on consumer perceptions of wine quality (Swaminathan 1995), as evidenced in several empirical analyses of price–quality relationships (Benjamin and Podolny 1999, Landon and Smith 1997, Roberts and Reagans 2007). Given this impact, producers have an interest in understanding how the critics’ judgments are generated. However, in contemporary U.S. wine markets, the critics’ positions as objective mediators disallow a priori consultation. This makes the analytic element of the wine reviews extremely important for
producers. There, critics provide explanations of their judgments about quality (Becker 1982).

Observation suggests the producers do seek guidance from the critics’ written words. In fact, a cottage industry of wine consultants has arisen to meet demand among producers for guidance on how to translate the critics’ tasting notes into more concrete guidance. For example, the Wine Angels’ “Wine Marketer’s Companion” (2003, p. 71) stresses that “the presence of certain attributes in both the Wine Spectator and the Wine Advocate tasting notes correlate to one wine scoring higher than another. Knowing which attributes yield the most influence when making or marketing a wine could provide a competitive advantage in the marketplace.” However, there can be problems associated with trying to interpret tasting notes. It has been suggested that wine critics “report on several thousand wines a year. They try as best they can to describe the ones they like. It is a daunting task” (Prial 1994, p. C1). What makes the task daunting, according to Darlington (2005), is that wine is too complex to be completely summed up in a single number. In this respect, producers trying to forecast a critic’s quality judgments require relatively clear evaluative schemas that map the ratings onto the analytic elements that support them.

In the next section, we analyze the clarity of evaluative schemas within different varietal categories. We develop an indirect approach to inferring clarity in evaluative schemas by focusing on the language used to justify quality assessments. Our approach shows that varietal categories do indeed vary in the extent to which critics’ judgments are guided by clear schemas for evaluation. This allows us to test our main pricing order prediction and its three moderators.

### Inferring Evaluative Schemas Based on Descriptive Similarity

In markets where critics do not explicitly articulate their evaluative criteria, how are schemas for evaluation conveyed to others in the market? Research on cultural markets points us to the structure of the language that critics use in their evaluations (Podolny and Hsu 2003, White and White 1993).

As Podolny and Hill-Popper (2004, p. 98) observe, “[T]he language that one uses for talking about value and quality is a reflection of the underlying understanding of value and quality, and since it is a reflection, language is therefore a tangible guide to those perceptions.” Language is thus a fundamental tool by which other market actors can infer the schemas critics use in their assessments of quality. In this spirit, West et al. (1996, p. 120) propose that an accepted “consumption vocabulary” separates the relevant attributes from the myriad of details and thereby helps market actors diagnose the main drivers of consumer preferences. An accepted consumption vocabulary “provides consumers with a framework to correctly identify the relationship between a product’s features and their overall judgment” (West et al. 1996, pp. 121–122). By imposing a structure, consumption vocabularies allow producers to move from simply speculating about preferences to being able to predict them. Translating this to our current focus on wine critics, West et al. (1996, p. 134) suggest that “if [producers] have been able to learn from previous experiences what characteristics in wine [critics] like best, [they] will be able to make a better-educated guess about what [they] will like on the basis of the descriptions [that have been] offered.”

In the wine industry, notions of which characteristics constitute high quality vary from one varietal to the next. For example, whereas a critic may prefer a cabernet sauvignon wine that has intense flavors, a full body, and firm tannins, he or she would be unlikely to value those same characteristics in a wine made from merlot, a grape that tends to produce softer, suppler, and less tannic wines. And although characteristics like floral and citrus are rarely mentioned in descriptions of high-quality cabernets or merlots, they are often used for white grape varietals such as riesling and viognier. Consistent with these examples, Laube (1999, p. 16, emphasis added) observes that each wine varietal possesses distinct characteristics and thus follows a distinct schema for evaluation:

> Even when grown in different appellations and vinified using different techniques, a varietal wine always displays certain qualities, which are inherent in the grape’s personality. Understanding what a grape should be as a wine is fundamental, and knowing what a grape can achieve at its greatest is the essence of fine-wine appreciation.

Varietals also differ in the clarity of the schemas that map discourse about specific characteristics to overall assessments of wine quality. In our analyses, we infer the clarity of the evaluative schemas employed by Wine Spectator critics for different varietals based on the language used in their wines’ tasting notes. The Wine Spectator and Robert Parker’s Wine Advocate are widely regarded as the most influential critics in the wine industry. We focus on the Wine Spectator because its reviews are available from the early 1980s (when it began publishing its wine reviews) and because of the relative transparency and consistency of its review process. The vast majority of Wine Spectator reviews are conducted using a rigorous blind tasting methodology. Critics enter their tasting notes and ratings into a database before learning the identity of the wines, their producers, or their prices. As Wine Spectator states, “Many scientific studies have shown that judgment is strongly influenced by awareness of brand or price. Blind tasting helps to remove a critic’s bias. At Wine Spectator, all official reviews of newly released wines result from blind
tastings (any reviews not based on blind tastings are clearly specified).” Our data only include reviews from these blind tastings. As such, we can be sure that the critics do not know a wine’s producer or its list price while they perform their evaluations.

Our specific approach to assessing the clarity of evaluative schemas follows Dickie (1997, p. 139, emphasis added), who argues that “insofar as a critic uses principles, he or she will treat similar cases similarly . . . .” The test of whether or not a person uses principles [or clear evaluative schemas] in evaluating is whether or not the person evaluates cases similarly that are similar in the relevant respects.” The Wine Spectator publishes both the evaluation (i.e., a score on a 100-point scale) and the analysis (i.e., the open-ended tasting note) in each review. These scores are grouped into several bands: 90–100 = “Classic/Outstanding”: a great wine or a wine of superior character and style; 85–89 = “Very good”: a wine with special qualities; 80–84 = “Good”: a solid, well-made wine; 75–79 = “Mediocre”: a drinkable wine that may have minor flaws; and 50–74 = “Not recommended.” We propose that the similarity in the language used to describe wines within each quality band reflects the clarity of its critics’ understanding of what quality means for that varietal in that year.

When the descriptive words in the tasting notes for wines of similar quality are themselves more similar, one may infer that the Wine Spectator critics have a clearer schema of what quality is in that varietal category. This clarity provides guidance to producers that must anticipate how critics will react to a wine and then price their wines accordingly. On the other hand, when the text in the reviews of wines within quality bands displays little overlap, one may infer that the varietal category lacks clear schemas for evaluation. The lower descriptive similarity reflects ambiguity over what characteristics good wines should have and thus offers less help for producers looking for pricing guidance.

We use the following procedure to create a measure of the degree of similarity in the descriptions of wines of similar quality within each varietal category in each year. We first compile a list of over 1,400 wine words from several prominent wine glossaries. The aim here is to begin with a comprehensive list of words that are used by wine professionals and consumers alike. Many of the words in this list are different grammatical forms of the same root words. We combine these and then eliminate those words that are clearly nonevaluative (e.g., “wine,” “aroma wheel,” “mechanical harvester”) and those that are names of specific varietals or regions. This results in 445 words that populate our “long” list of wine terms.

Sensory research shows that only certain words from this list effectively facilitate the communication of sensory-related properties. Civille and Lawless (1986) specify three criteria for identifying these words. They must convey specific information relevant to the underlying natural, chemical, or physical structure of a product. They must be appropriate to the frames of reference of the intended perceivers (i.e., wine producers, critics, and consumers) so that the proper abstraction of concepts is possible. Finally, they should be elemental words and not combinations of other words. In the context of wine, the word “creamy” reflects a combination of “smoothness,” “viscosity,” and “fatty mouthfeel” and therefore conveys imprecise information about a wine’s attributes. To limit our list of words to those that are specific, appropriate, and elemental, we consulted a wine sensory expert. Applying these three criteria to the words in our long list, this expert extracted a “short” list of 320 words. The following example shows one of the sampled wine reviews and indicates the relevant descriptors according to the long and the short lists.

Long list: Rich yet elegant and polished, with supple currant, cedar, chocolate, berry and spice notes, turning complex and rich (relevant words italicized).

Short list: Rich yet elegant and polished, with supple currant, cedar, chocolate, berry and spice notes, turning complex and rich (relevant words boldfaced).

We rely on the short list of wine words to extract a set of descriptors from each wine review. We then measure the degree of similarity in the words extracted from the reviews of any two same-quality band wines taken from the same varietal category. Numerous measures exist to assess the similarity of a pair of objects along a set of binary variables that, in our case, reflects whether each review contains each descriptor in our short list. One of the basic distinctions between different binary measures of similarity is whether they treat the co-absence of attributes as relevant to the judgment of how similar objects are (Everitt et al. 2001). For example, if the reviews for two wines do not contain the descriptor “complex,” should the co-absence of “complex” be included in their calculation of similarity? For our purposes, counting co-absences in a calculation of similarity does not seem sensible. Given the brevity of wine reviews, most descriptors are absent from any one review. As such, we do not want to attribute a high degree of similarity to a pair of reviews simply because they both lack a large number of descriptors.

In our main analyses, we use the simplest measure that does not incorporate co-absences—the Jaccard coefficient, which takes the following form:

$$J = \frac{C_{\text{shared}}}{(C_{\text{shared}} + C_{\text{distinct}})},$$

where $C_{\text{shared}}$ counts the number of short-list words that are found in both reviews, and $C_{\text{distinct}}$ counts the short-list words that appear in one review but not the other. Figure 1 provides an example of how this measure is calculated for two different pairs of wines. Wines A and B are described in more similar terms—roughly half of the total number of relevant words that they contain are common to both reviews. On the other hand, less
Figure 1  Demonstration of Jaccard Scores

Wine A: A powerful, young wine, even at age 10. Dark, ripe, intense and deeply concentrated, this one is years from peaking. Packed with a rich, earthy core of currant, mineral, anise, sage, and cedar. Quite tannic in this tasting, the fruit poured through on the finish.

Wine B: Firm, dark, ripe, and intense. A young, vibrant wine with lots of earthy currant, spice, cedar, and sage notes, finishing with a tight band of tannin and good length.

Wine C: A beautiful wine, concentrated, velvety, and elegant, with delicious violet-tinged cherry and currant aromas and flavors, firm but not overwhelming tannins, and a crisp undertone of acidity to support the finish.

Shared: Dark, ripe, earthy, currant, cedar, sage, tannin, young, intense

Distinct: Mineral, anise, fruit, finish, firm, spice, length, rich, vibrant, tight, good

Shared: Firm, currant, tannin

Distinct: Dark, ripe, earthy, spice, cedar, sage, length, velvety, cherry, aromas, flavors, crisp, acidity, finish, intense, young, vibrant, tight, beautiful, elegant, delicious

Note. Words in boldface are in the short list; words in boldface or italics are in the long list.

than one-fifth of the wine words found in the reviews of wines B and C are common to both reviews.

Within each varietal category \((ij)\) and each year \((t)\), we calculate the average pairwise similarity between all same-quality band wines, and we then calculate the weighted-average similarity across the four positive (i.e., greater than 74) bands \((j)\):

\[
\text{Jaccard}_{ij} = \frac{\sum \left( w_{jit} \times J_{jit} \right)}{\sum w_{jit}},
\]

where \(w_{jit}\) is the count of the number of wines in quality band \(j\) in the focal varietal-year. A higher Jaccard measure implies greater overall similarity in how wines of similar quality are analyzed and described. It therefore captures the extent to which critics have clearer schemas for the characteristics that quality wines in that varietal category should possess.

Another consideration in the creation of this measure is whether consistency over time contributes to the guidance followed by producers. We explore this by creating a weighted-average similarity variable based on all reviews within a varietal category over moving three-year windows and comparing this to the one-year version of the variable. The three-year specification has much greater explanatory power and is thus the one that we use in our main models. The three-year Jaccard measure ranges from a low of 0.00 to a high of 0.251. This suggests that varietal categories ranged from having no correspondence at all in the descriptions of their similar-quality wines to having approximately one-quarter of their short-list words in common.

To check the validity of this measure, we returned to the sample of tasting notes and searched for cases where the critic documented that the list price was surprisingly low. According to the <b>Wine Spectator</b>, the only additions critics are permitted to make to their tasting notes after learning about the identity and price of a wine are expressions of the surprisingly low price of the wine in question. This is clearly something that should occur less often in varietal categories where producers are better able to anticipate the reactions of critics in their pricing decisions. Upon inspection of the data, this is always associated with finding the word “price” or “value” in the tasting note, something that occurred in 674 reviews. We counted the number of such occurrences within each varietal category and then computed the percentage of pricing surprises. These percentages ranged from zero (e.g., sauvignon blanc) to almost seven (e.g., gamay and chenin blanc). Across the 23 varietal categories that we study, there is a negative correlation \((\rho = -0.35)\) between this pricing surprise variable and the Jaccard score averaged across all available years. This indicates that critics were indeed less likely to be surprised by list prices in varietal categories where Jaccard scores tended to be higher.

A potential concern with our approach stems from its reliance on a single source of critical evaluations. This leaves open the possibility that our Jaccard measure reflects little more than the actual similarity in the objective features of wines in a varietal category. Critics from the <b>Wine Spectator</b> may simply be documenting all of the objective features of wines in their tasting notes. Thus, the varietal categories in which they appear to apply clearer schemas may be composed of objectively more similar wines. The previous remarks about the inherent complexity of wines and their evaluation suggest that the assessment of wines will never be as straightforward as the assessment of, for example, diamond quality. That said, we investigate this concern by comparing the language used by Robert Parker in his reviews with that used in our <b>Wine Spectator</b> reviews. This investigation is based on all reviews of U.S. wines published in the <b>Wine Advocate</b> between 1992 (the first year that his reviews are available online) and 2001. We first examine the most frequently used short-list words by each critic across various varietal categories. For example, Table 1 shows the eight most common
words used by Robert Parker and the *Wine Spectator* in their reviews of cabernet sauvignon and sauvignon blanc. Between 1994 and 1996, only two of the eight most frequently used descriptors for cabernet sauvignon were common to both critics (in italics). For sauvignon blanc, just three words were common to both lists (in italics).

Although cursory, these comparisons suggest that the language and therefore evaluative schemas that underpin different critic’s assessments of wine quality can diverge in meaningful ways. In a more systematic examination, we calculate the semantic similarity between the descriptions of specific wines that were reviewed by both critics. After identifying almost 6,000 wines that were reviewed by both Robert Parker and the *Wine Spectator* between 1992 and 2001, we use our Jaccard approach to assess their semantic similarity. Within each varietal category and each year, we calculate the average pairwise similarity in short-list words for the wines reviewed by both critics. This ranged from 0 (no words in common) to 0.286, with a mean of 0.082. On average, less than one-tenth of the descriptors were in both Robert Parker’s and *Wine Spectator’s* reviews of the same wines. The fact that two experienced critics can apply such different schemas clearly goes against the notion that our measure of schema clarity is driven by objective wine similarity.

### Table 1 Most Frequent Descriptors in Two Varietal Categories

<table>
<thead>
<tr>
<th>Wine Spectator</th>
<th>Robert Parker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cabernet sauvignon</td>
<td>Cabernet sauvignon</td>
</tr>
<tr>
<td>tannin</td>
<td>tannin</td>
</tr>
<tr>
<td>finish</td>
<td>finish</td>
</tr>
<tr>
<td>cherry</td>
<td>cassis</td>
</tr>
<tr>
<td>currant</td>
<td>nose</td>
</tr>
<tr>
<td>spice</td>
<td>taste</td>
</tr>
<tr>
<td>ripe</td>
<td>color</td>
</tr>
<tr>
<td>plum</td>
<td>fruit</td>
</tr>
<tr>
<td>oak</td>
<td>sweet</td>
</tr>
<tr>
<td>Sauvignon blanc</td>
<td>Sauvignon blanc</td>
</tr>
<tr>
<td>finish</td>
<td>finish</td>
</tr>
<tr>
<td>fruit</td>
<td>fruit</td>
</tr>
<tr>
<td>crisp</td>
<td>crisp</td>
</tr>
<tr>
<td>spice</td>
<td>dry</td>
</tr>
<tr>
<td>pear</td>
<td>fresh</td>
</tr>
<tr>
<td>floral</td>
<td>honey</td>
</tr>
<tr>
<td>bright</td>
<td>body</td>
</tr>
<tr>
<td>apple</td>
<td>melon</td>
</tr>
</tbody>
</table>

**Analysis and Results**

In auction-type markets, predictability is reflected in how prices respond to new quality information that leads to revisions on the supply and demand sides of the market. Thus, when Baker (1984) and Zuckerman (2004) focus on options and stock markets, they rightly examine the volatility of price reactions to new information. In markets with administrated prices, however, there is little opportunity to adjust prices after products are released (Blinder et al. 1998, Velthuis 2005). Here, predictability is reflected in list prices that more accurately and systematically reflect the information that will ultimately shape consumer valuations.

Following Sorensen and Sørensen (2001) and Sørensen (2002), we employ a multiplicative heteroscedasticity, or variance function model (Davidian and Carroll 1987), to assess the impact of quality schemas on list price variability around expected levels. Here, real list prices \( p_{ijt} \) are broken into two components:

\[
p_{ijt} = \mu_{ijt} + \sigma_{ijt} + \epsilon_{ijt},
\]

where \( \mu_{ijt} \) is a function that describes the expected price of product \( i \) in varietal category \( j \) at time \( t \):

\[
\mu_{ijt} = E(p_{ijt}) = \beta'X_{ijt-1}.
\]

Pricing variance is reflected in \( \sigma_{ijt} \), and \( \epsilon_{ijt} \) is a standard error term. More specifically, \( \sigma_{ijt} \) is modeled as

\[
\mu_{ijt} = \text{Var}(p_{ijt}) = \exp(\gamma'Z_{ijt-1}),
\]

where \( \gamma \) is a vector of coefficients reflecting the influence of the \( Z_{ijt-1} \) variables on the variance of product prices around expected levels. Because the first (mean) and second (variance) moments of a normal distribution are independent of each other, we maximize the likelihood function to obtain the estimates of the \( \beta \) and \( \gamma \) parameters (Greene 2000). This latter variance function allows us to test whether the variance around expected list prices is lower in varietal categories that evidence clearer schemas for quality. More specifically, we predict that as quality schemas firm up, prices should become less variable (\( \gamma_{\text{T}} < 0 \)).

We begin with a baseline model of a wine’s real list price (Combris et al. 1997, Landon and Smith 1997), which is the suggested retail price at the time of release divided by the consumer produce index normalized to unity in 1983. Because real wine prices are highly skewed, we use the natural log of this real price variable as our dependent variable. The mean function includes the quality score reported in the *Wine Spectator*. The natural log of each wine’s age at release is included as a control variable because older wines tend to command higher prices. Quality, quantities produced, and therefore wine prices tend to vary systematically across vintages (Ashenfelter 2008), and so we include a set of indicator variables for each of the post-1980 vintages in our data. We include another set of indicator variables that specify the different regions in which the wines were grown, as this has been shown to influence the prices of U.S. wines (Benjamin and Podolny 1999). The regions...
are the Bay Area, Carneros, Mendocino, Napa, Sierra, Sonoma, South Coast, other California, Finger Lakes, Long Island, other New York, Oregon, and Washington. To account for any unobserved varietal category effects on wine price, we also include the set of 23 varietal category indicator variables (including, for example, cabernet sauvignon (and blends), merlot, pinot noir, chardonnay, sauvignon blanc, and sparkling wines).

To account for any trend in the prices for U.S. wines over time, the model includes a variable that counts the number of years since the first time any wine in the focal varietal category was reviewed. Because of competitive effects, the number of wines rated in the same varietal category should depress prices, and so we include the natural log of rated-product density in the focal varietal category. When wines in a category are sold by a more concentrated set of producers, the prospect for coordination might have an offsetting effect on prices. We therefore include a Herfindahl measure of the degree of concentration of rated wines across producers.

Because producers with superior reputations for product quality tend to charge higher prices for their wines (Benjamin and Podolny 1999, Landon and Smith 1997, Roberts and Reagans 2007), we include the average of all prior quality scores received by the focal producer. We impute the reputation score with the overall sample average for producers that never before received a published review from the Wine Spectator. We also include an indicator variable set to 1 for observations that represent a producer’s first published review. These producers have no reputations and no experience anticipating the reactions of critics to their wines. We therefore expect the coefficient on this variable to be negative. We also include the natural log of producer age—indicated by the focal year minus the vintage year of the first wine reviewed—as a control variable but are agnostic about the expected direction of its effect on prices. We do not have a direct measure of the cost structures of different producers but control for variation in costs using a proxy for the economies of scope that are available to producers with broader niche widths. We expect these producers to have lower costs and thus an ability to sell at lower prices. To calculate each producer’s niche width, we use Simpson’s (1949) index of diversity, which measures the diversity of a distribution over a set of discrete categories:

$$W(\mu(x, t)) = 1 - \sum_{l \in l(v)} \mu^2(l, x, t),$$

where \(l(v)\) denotes the set of varietal categories in which the producer has had wines reviewed over past three years, and \(\mu(l, x, t)\) refers to proportion of the producer’s total reviewed wines that came from varietal \(l\) over that same time period.

The variance function includes the three-year Jaccard variable, which tests our main prediction. It also includes as controls the variables that measure the number of years since the first review in the focal varietal category was published, the natural log of the count of reviewed wines, and the concentration of published reviews across producers. Finally, varietal category fixed effects are included to net out the influence of systematic unobservable factors that are specific to each varietal market.

Starting with all wines produced after the 1980 vintage and reviewed in the Wine Spectator after 1985, we exclude reviews without list prices, quality ratings, tasting notes, or known grape varietals. We also set aside the reviews that make explicit reference to price or value in the tasting notes. Because of the demonstrated lack of correspondence between the schemas invoked by the Wine Spectator critics and those applied by Robert Parker, we are wary of the implications of having non-conforming guidance from two important critics. To focus on reviews less likely to be influenced by this potentially confounding issue, we set aside the observations for which the focal producer had received at least one review from Robert Parker in the focal varietal category during the previous three years. This leaves a sample of 20,019 reviews spread across 23 varietal categories. Descriptive statistics and correlations for these variables are reported in Table 2.

Our main model, Model 1, is presented in the first column of Table 3. Each of the expected effects in the mean function of the model is significant with the exception of the concentration variable. The coefficient estimates suggest that real wine prices trend upward over time. Producer age exerts a positive effect on prices, whereas the producer’s niche width has the expected negative impact. In the variance function, the varietal category fixed effects are jointly significant. With the passing of time, the elaboration of a varietal category has a small positive impact on pricing variance around expected levels. On the other hand, increasing the number of prior-year reviews available as examples decreases ambiguity for those setting prices. The effect of the three-year Jaccard measure is negative and significant. This result provides support for our main pricing order prediction. When evaluative schemas are clearer, list prices tend to be less variable around expected levels.

We obtain virtually identical results in an unreported model that includes the Jaccard measure in the mean function of the model. There, the effect of the Jaccard measure on price levels is negative and significant, but its effect in the variance function remains the same. Another concern is that the Jaccard effect might be an artifact of the average number of short-listed descriptors contained in the reviews, which upon inspection is found to be correlated with the three-year Jaccard variable. However, another unreported variant of Model 1 finds that the average number of words exerts no significant effect on pricing errors while the estimated Jaccard effect is virtually unchanged.
To further investigate the robustness of our main finding, we conducted several supplementary analyses. First, we replicated the results in Model 1 using several alternative similarity measures (see Appendix A). We also replicated our main finding using a series of quality band-specific Jaccard measures. Recall that the Jaccard measure in Model 1 is based on the average pairwise similarity across the four positive quality bands. To test whether producers take guidance from wines in the different quality bands, we sequentially replace this averaged similarity measure with a series of measures based on the reviews of wines from specific quality bands. The coefficient estimates for the quality band-specific Jaccard measures demonstrate that producers take guidance from increased semantic consistency within all positive quality bands. The guidance effect is particularly strong from previously reviewed wines that received scores in the 80–84 and 85–89 ranges. There is a smaller effect associated with previously reviewed wines receiving scores of less than 75; this, along with their characterization as being flawed (i.e., classified as “Not recommended”), affirms our decision to employ a Jaccard measure that incorporates wines from all positive quality bands reported in the Wine Spectator.

Model 2 adds several variables to the variance function that test H2 through H4. We isolate cases in which a focal producer had never before received a published Wine Spectator review—i.e., those for which Producer’s first review equals 1. For producers that received reviews before the focal year, we isolate those whose reputations fall inside the lowest two quality bands (i.e., “Not recommended” and “Mediocre”). Finally, we isolate a group of generalist producers that fall in the highest 25% of the distribution of the niche width variable. The coefficient estimates on the interactions between these three variables and the three-year Jaccard variable test our moderating predictions. Although the main effect of having no prior Wine Spectator reviews on pricing variance is not significant, the effect of its interaction with the Jaccard measure is positive and significant. In fact, its effect size is such that it almost completely offsets the negative main effect of the Jaccard measure on pricing variance. The interaction between the low-reputation indicator variable and the Jaccard measure is also positive and significant, as predicted in H3. Finally, generalist producers tend to price with lower variance around expected levels. Moreover, consistent with H4, the effect of the Jaccard measure is significantly less negative among these generalist producers. These results offer support for all three moderating predictions.

We close our analysis by addressing potential concerns about a mismatch in level of analysis between our theorizing and the analysis summarized in Table 3. Our theory is at the category level, whereas the results from our main analyses are at the product level. To determine the effects of schema clarity on pricing volatility at the category level, we first estimate (using ordinary least squares regression) the baseline pricing model to obtain an estimate of what the price of each wine should be. Our predictions are about the extent to which producers tend to deviate from these estimates when setting their prices in each varietal-year. In a second stage, we estimate the effects of different variables—including the three-year Jaccard variable—on the average of the errors around predicted prices taken at the varietal-year level. The results, which are described and reported in Appendix B, are consistent with our main findings.

### Discussion and Conclusion

Much like other social structures such as certain networks (Baker 1984), status hierarchies (Podolny 1993, 1994), and well-defined market roles (White 1981), market categories with clear boundaries and clear evaluative schema are theorized to facilitate and sustain orderly patterns of exchange. However, although there is accumulating evidence of the implications of categorical boundaries on market outcomes (i.e., the first stage of competition in candidate–audience models), less attention is being paid to documenting the influences of evaluative schema on the ability of producers to effectively differentiate their offerings in the second stage of competition. Our analysis establishes a link between the presence of clear evaluative schemas within market categories and a more robust correspondence between list

### Table 2 Descriptive Statistics and Correlations—Bottle Level (N = 20,019)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. dev.</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
<th>(7)</th>
<th>(8)</th>
<th>(9)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) In(Real price)</td>
<td>2.343</td>
<td>0.483</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) Years since first rating</td>
<td>9.748</td>
<td>4.308</td>
<td>0.205</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) In(Reviewed wines)</td>
<td>5.277</td>
<td>1.151</td>
<td>0.176</td>
<td>0.370</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) Concentration</td>
<td>0.021</td>
<td>0.043</td>
<td>−0.078</td>
<td>−0.209</td>
<td>−0.707</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5) In(Producer age)</td>
<td>1.913</td>
<td>0.871</td>
<td>0.048</td>
<td>0.079</td>
<td>0.033</td>
<td>−0.034</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6) Producer reputation</td>
<td>83.896</td>
<td>3.172</td>
<td>0.383</td>
<td>0.137</td>
<td>0.056</td>
<td>−0.023</td>
<td>−0.095</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7) Producer’s first review</td>
<td>0.125</td>
<td>0.331</td>
<td>0.008</td>
<td>−0.040</td>
<td>−0.015</td>
<td>0.021</td>
<td>0.606</td>
<td>0.128</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8) Niche width</td>
<td>0.624</td>
<td>0.224</td>
<td>−0.273</td>
<td>0.102</td>
<td>−0.027</td>
<td>0.010</td>
<td>0.415</td>
<td>−0.170</td>
<td>−0.345</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(9) In(Bottle age)</td>
<td>0.921</td>
<td>0.404</td>
<td>0.354</td>
<td>0.126</td>
<td>0.109</td>
<td>−0.039</td>
<td>−0.133</td>
<td>0.023</td>
<td>0.012</td>
<td>−0.171</td>
<td></td>
</tr>
<tr>
<td>(10) Quality score</td>
<td>84.205</td>
<td>4.828</td>
<td>0.411</td>
<td>0.155</td>
<td>0.087</td>
<td>−0.041</td>
<td>0.047</td>
<td>0.332</td>
<td>−0.054</td>
<td>−0.053</td>
<td>0.013</td>
</tr>
</tbody>
</table>

Hsu, Roberts, and Swaminathan: Evaluative Schemas and the Mediating Role of Critics
Currently, we know little about exactly how critics fulfill forms of uncertainty faced by consumers and producers. The many roles played by critics in reducing the various uncertainties engendered by the market are also hard to quantify. We suspect that something of this nature may be more likely if critics are able to carve out distinct niches of this kind of competition for influence on the stabil- ization and differentiation that has already been docu- mented among market producers.

Zuckerman et al. (2003) typecasting, she may be more likely to gain legitimacy and therefore consideration by producers making their pricing decisions. Here, there may be trade-offs inherent in the struggle for critics to maintain their relative influence over time. As discussed in the preceding paragraph, one trade-off is between adhering to clear and stable evaluative schema versus leading the market by continually updating or refreshing schema. In the multicritic world, there may be another trade-off between establishing a truly distinctive schema versus one that corresponds more closely to those held by other influential critics. A critic may be better able to build a distinct reputation by establishing her own evaluative schema rather than adopting those advocated by others. On the other hand, in a pattern reminiscent of the Zuckerman et al. (2003) typecasting, she may be more likely to gain legitimacy and therefore consideration by conforming to the schema espoused by others. In the end, one might expect the same tension between consider- ation and differentiation that has already been docu- mented among market producers.

More broadly, one wonders about the overall impact of this kind of competition for influence on the stability and predictability of market exchanges. Order seems more likely if critics are able to carve out distinct niches of producers and/or audience types in which to specialize. We suspect that something of this nature may be going on in the U.S. wine industry, as the majority of the wines in our data were reviewed by Robert Parker or the Wine Spectator, but not both. On the other hand,
when the different critics elect to, or are forced to, vie for the same niche and then promote divergent conceptions of quality, producers and consumers will find it more difficult to achieve agreement on appropriate terms of trade. In the extreme, clashing evaluative schema may throw the legitimacy of critics themselves into question. If every critic proposes a radically different ranking of products, others in the market will surely question the validity of their judgments and may choose to rely on other indirect indicators of quality, such as producer status (Podolny 1993). These kinds of questions and concerns clearly suggest the value of future research.

In this way, our study paves the way for future exploration of the implications of both producers’ and critics’ interests and agency on the evolution of the evaluative schemas that provide order in a market. We have demonstrated the impact that these schemas have on the degree of order observed across different market categories. This provokes one to ask about the factors that shape the paths through which these evaluative schemas evolve. The answer to this question is likely tied both to the complexity of features inherent in products and to the diversity of producers and critics that populate the different markets. Heterogeneity in experience levels, backgrounds, geographical locations, and strategies (Klepper and Simons 2000, McKendrick and Carroll 2001, McKendrick et al. 2003, Simons and Roberts 2008) increases the diversity among products that are grouped together in the same market categories. As they all seek to establish advantageous positions, they may be more likely to endorse and promote diverging schemas for evaluation. This might hinder or upset the development of clear and agreed-upon evaluative schemas. By contrast, markets in which a stable and homogeneous set of incumbents dominate may tend to see increasing clarity regarding the basis for evaluation with the passage of time. This type of focus on evolution in evaluative schemas will likely reveal additional factors that shape the extent to which mediators effectively structure market outcomes.

Our study provides an intriguing first glimpse into how the structure of evaluative schemas condition the mediating role performed by critics within markets. Further exploration of the issues we raise above will allow organization theorists and economic sociologists to move past repeated demonstrations that critics have important effects on markets and market order and begin to address important questions about how these effects are realized. Accumulating theory and evidence of this type will help to explicate the still somewhat mysterious link between social structuration and market order.

Acknowledgments
The authors thank seminar participants at the University of Chicago, London Business School, Massachusetts Institute of Technology, Stanford University, and University of Michigan for valuable comments that led to improvements in this paper.

Appendix A. Assessing Robustness Using Different Indicators of Clear Evaluative Schemas
When we use a Jaccard measure based on our long list of descriptors, the estimated variance effect remains negative and significant. We also created a pair of measures that are similar in structure to the Jaccard but differ slightly in their construction. The Jaccard measure is the most basic indicator of similarity: the proportion of matches out of all the descriptors found in the two compared reviews. A Dice measure places more emphasis on commonality by giving twice the weight to matched descriptors, whereas a Kulczynski II measure adjusts for differences in the number of descriptors found in the two reviews. The latter measure addresses the potential concern that the Jaccard measure misstates similarity when one of the reviews has relatively few descriptors. If review A houses 2 descriptors and review B has 10, then the similarity to B from A’s perspective is stronger if there is one matching descriptor (1 out of 2) relative to the similarity to A from B’s point of view (1 out of 10 matches). As seen in Table A.1, the correlations among these three similarity measures are very high. It is therefore not surprising that we get virtually identical results across the three variants of Model 2 (see Table A.2).

Another potential indicator of evaluative schemas is the degree of repetition of the specific terms used to analyze wines, which reflects the consistency with which schemas for evaluation are employed. Greater repetition is reflected in a higher concentration of words in the reviews of all wines within the varietal category (Podolny and Hill-Popper 2004, p. 101, footnote). To examine this, we replaced the Jaccard measure with a Herfindahl measure that calculates the percentage of total wine review words from the short list accounted for by the eight most frequently used terms in each varietal category. Note first that the Jaccard and Herfindahl measures are correlated at 0.69. Moreover, the negative coefficient on this Herfindahl measure again corroborates our results.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coef.</th>
<th>Robust std. error</th>
<th>Log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jaccard (3-year), short</td>
<td>−2.606**</td>
<td>0.587</td>
<td>−5,822.7</td>
</tr>
<tr>
<td>Jaccard (3-year), long</td>
<td>−2.925**</td>
<td>0.713</td>
<td>−5,826.2</td>
</tr>
<tr>
<td>Dice, short</td>
<td>−1.579*</td>
<td>0.355</td>
<td>−5,822.4</td>
</tr>
<tr>
<td>Kulczynski II, short</td>
<td>−1.546**</td>
<td>0.344</td>
<td>−5,821.8</td>
</tr>
<tr>
<td>Herfindahl, short</td>
<td>−0.179*</td>
<td>0.055</td>
<td>−5,834.4</td>
</tr>
</tbody>
</table>

**p < 0.01.
Appendix B. Aggregating to the Varietal Category Level of Analysis

In the first step of this supplementary analysis, we estimate a baseline hedonic pricing model to obtain an estimate of each wine’s price. The measures in this model are the same as those in the mean function of Model 1 in Table 3 (and their estimates are consistent). To test our main prediction, we then extract the residuals from this baseline model and compute the average of their squares at the varietal-year level, thereby creating a variable that indicates the amount of error around the prevailing pricing equation. This produces pricing error observations for 270 varietal-year observations. We then regress this variable on the three-year Jaccard variable while controlling for the number of years since the first review in the focal varietal category was published, the natural log of the count of reviewed wines, and the concentration of these reviews across producers. Model B1 shows that varietal categories with higher Jaccard scores experience lower pricing variance.

To test our moderating predictions, we compare the degree to which producers in different market positions are influenced by the clearer evaluative schemas implied by higher Jaccard scores. More specifically, we successively average the squared errors from Model 1 in Table 3 to different groups of producers within each varietal-year and then estimate the effect of the Jaccard measure within each group. Model B1a in Table B.1 is based on the 202 varietal-year observations that include first-review producers. The effect of the Jaccard measure on pricing variance is lower and no longer significant at conventional levels. Model B1b shows that the small effect of the Jaccard measure among the low-reputation producers is not statistically significant at conventional levels. Model B1c shows that the smaller estimated effect among the generalists is not significant. This pattern is consistent with our prediction (H4) that the guidance provided by critics’ evaluative schemas is not followed by producers that participate in a large number of varietal categories.

Table B.1  The Effect of Evaluative Schemas on Pricing Errors—Category-Level Analysis

<table>
<thead>
<tr>
<th></th>
<th>Model B1 (main)</th>
<th>Model B1a (producer’s first review)</th>
<th>Model B1b (low reputation)</th>
<th>Model B1c (generalist producers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.089**</td>
<td>0.053</td>
<td>0.074</td>
<td>0.025</td>
</tr>
<tr>
<td>(0.020)</td>
<td>(0.045)</td>
<td>(0.047)</td>
<td>(0.027)</td>
<td></td>
</tr>
<tr>
<td>Jaccard (three-year)</td>
<td>−0.435**</td>
<td>−0.382</td>
<td>−0.323</td>
<td>−0.170</td>
</tr>
<tr>
<td>(0.125)</td>
<td>(0.210)</td>
<td>(0.242)</td>
<td>(0.154)</td>
<td></td>
</tr>
<tr>
<td>Years since first rating</td>
<td>0.003*</td>
<td>0.008*</td>
<td>0.008</td>
<td>0.002</td>
</tr>
<tr>
<td>(0.001)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.001)</td>
<td></td>
</tr>
<tr>
<td>ln(Reviewed wines)</td>
<td>0.008*</td>
<td>0.009</td>
<td>0.001</td>
<td>0.012*</td>
</tr>
<tr>
<td>(0.003)</td>
<td>(0.010)</td>
<td>(0.011)</td>
<td>(0.005)</td>
<td></td>
</tr>
<tr>
<td>Concentration</td>
<td>0.103</td>
<td>0.115</td>
<td>0.057</td>
<td>0.233*</td>
</tr>
<tr>
<td>(0.045)</td>
<td>(0.147)</td>
<td>(0.183)</td>
<td>(0.105)</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>269</td>
<td>202</td>
<td>177</td>
<td>247</td>
</tr>
<tr>
<td>R²</td>
<td>0.087</td>
<td>0.036</td>
<td>0.047</td>
<td>0.058</td>
</tr>
</tbody>
</table>

Note. Robust standard errors.

*p < 0.05; **p < 0.01.

Endnotes


2Even producers that do not fully accept critics’ views about what constitutes quality know that they must abide by prevailing opinions to continue their roles as commercially successful market producers (Fine 1992).

3The other is Robert Parker of the Wine Advocate, who made his reputation predicting the quality of the 1982 Bordeaux vintage (Brook 2001).


5Thanks to Hildegard Heymann, professor of viticulture and enology at the University of California, Davis.

6As we elaborate in Appendix A, a number of alternative similarity measures yield results that are similar to the ones we report in our main analyses.

7Because only a very small percentage of products are traded in auction-like settings, there are few opportunities to update or modify list prices once products are in the market (Blinder et al. 1998). Moreover, in the wine industry, price adjustments within retail channels are further constrained by the fact that quality ratings are typically published alongside producers’ suggested retail prices, which serve as visible commitments to those list prices.

8All of the results that we report are replicated in the larger sample of 28,511 observations.

9Note that wine critics often achieve central positions and even celebrity status, whereas diamond evaluators do not.

References


**Greta Hsu** is an associate professor of management in the Graduate School of Management at University of California, Davis. Her current research investigates evolution in the production and performance of genre spanning films (with Giacomo Negro and Fabrizio Perretti), the manipulation of category claims in the tobacco industry (with Stine Grodal), and divergence across constituent groups in their identity categorizations for an organization (with Kimberly D. Elsbach).

**Peter W. Roberts** is an associate professor of organization and management in the Goizueta Business School at Emory University. His Ph.D. is from the University of Alberta. Recently, he has directed his interests in entrepreneurship, evolution, and organizational performance toward topics in the field of social enterprise, with projects focusing on social entrepreneurs, for-profit microfinance institutions, and company-sponsored philanthropic foundations.

**Anand Swaminathan** is the Goizueta Professor of Management at Emory University. His research examines the effect of founder roles on audience appeal in the Toronto restaurant industry (with Peter Roberts and Giacomo Negro), job structures in the California wine industry (with Heather Haveman and Eric Johnson), career mobility among NFL coaches (with Jim Wade and Andreas Schwab), and resource-partitioning processes among Wisconsin banks (with Giacomo Negro and Fabiana Visentin).