Typecasting, Legitimation, and Form Emergence:

A Formal Theory*

Greta Hsu
Univ. of California at Davis

Michael T. Hannan
Stanford University

László Pólos
Durham University

Running head: Typecasting, Legitimation, and Form Emergence

June 8, 2010

*We thank Glenn Carroll, Özgecan Koçak, Giacomo Negro, Filippo Wezel, and the referees for comments, and the Stanford Graduate School of Business, the Stanford GSB Faculty Trust, and Durham Business School for generous financial support. Please direct correspondence to: Greta Hsu, Graduate School of Management, Univ. of California at Davis, One Shields Ave., Davis, CA 95616
Abstract

We propose a formal theory of multiple category memberships. This theory has the potential to unify two seemingly unconnected theories: typecasting and identity-based form emergence. Typecasting, a producer-level theory, considers the consequences of specializing versus spanning across category boundaries. Identity-based form emergence considers the evolution of categories and how the attributes of producers entering a category shape its likelihood of gaining legitimacy among relevant audiences. Both theory fragments treat the processes by which audience members assign category memberships to producers. This paper develops this common foundation and clearly outlines the arguments that lead to central implications of each theory. We formalize these arguments using modal expressions to represent key categorization processes and the theory-building framework developed by Hannan, Pólos, and Carroll (2007).
Introduction

Categorization has attracted considerable interest from economic, cultural, and organizational sociologists in recent years. Empirical work on this subject covers a range of issues. One substantial branch treats category emergence, considering issues such as how the existing structure of a market shapes new category formation (Ruef 2000; Pontikes 2008; Simons and Roberts 2008) and how new categories arise through the recombination of existing ones (Jensen, forthcoming; King, Clemens, and Fry, forthcoming). It also analyses how key actors, such as producers and consumers (Rosa, Porac, Runser-Spanjol, and Saxon 1999), media outlets (Kennedy 2005; 2008), industry associations (Waguespack and Sorenson forthcoming) and social movements (Carroll and Swaminathan 2000; Schneiberg, King, and Smith 2008), affect the creation of new organizational categories.

Research has examined other processes of category evolution including the convergence of category understandings (Koçak 2008; Cattani, Ferriani, Negro, and Perretti 2008; Koçak, Hannan, and Hsu 2009) and divergence (McKendrick and Carroll 2001; McKendrick, Carroll, Jaffee, and Khessina 2003; Bogaert, Boone, and Carroll forthcoming), as well as the reconstitution (Lounsbury and Rao 2004) and erosion (Rao, Monin, and Durand 2005) of categories.


This proliferation of research indicates a growing awareness of the importance of categorical issues to social and economic dynamics, and it demonstrates the vitality of this broad topic area. Yet, as this research area continues to progress, researchers should keep in mind the challenge of building cumulative knowledge. Little attention has been directed to connecting the ideas and
intuitions that animate the different theory fragments on categorization. Learning how these separate fragments fit together—or what changes must be made to make them fit—is vital to continued advancement of this area of research. Theoretical integration allows for checking the consistency of different lines of reasoning as well as identifying core ideas that can be transferred over contexts. It also points to areas of conceptual intersection that remain ambiguous and require future exploration.

We begin to address this challenge by developing a formal theory of multiple category memberships. A desired by-product of a formalization of a theory is the discovery of unsuspected connections with other theories. We believe that the work presented here exemplifies this advantage. Originally, we set out to formalize typecasting theory. This theory focuses on well-established categories; it considers the implications for individual producers of specializing in one category versus generalizing across categorical boundaries (Faulkner 1983; Zuckerman, Kim, Ukanwa, and von Rittman 2003; O’Mahony and Bechky 2006; Leahey 2007; Hsu, Hannan, and Koçak 2009). Research suggests that audiences can better make sense of specialists, but association with a single category restricts the range of future opportunities. We sought to clarify the mechanisms at work in this kind of process by applying some nonstandard logical tools.

Once we had this formalization in hand, we realized that with some minor extensions it could explain a seemingly unrelated phenomenon: the differing consequences for category evolution of the founding of producers devoted to the activities associated with a new label versus the arrival of de-alio entrants (producers who are already associated with some other category labels). More generally, identity-based theory of form emergence considers how the variability of the attributes of the producers associated with an emerging category shapes the likelihood that it will gain constitutive legitimacy (become taken for granted) for relevant audiences (McKendrick and Carroll 2001; McKendrick, Jaffee, Carroll, and Khessina 2003). Work in this area finds that a category is more likely to become a taken-for-granted form when new entrants lack an affiliation with other categories and focus solely on the activities associated with the category.
An important contribution of our paper is to develop the logic behind these implications, which centers on processes core to the typecasting dynamic.

The end result is a formal structure that unifies two theory fragments that appear dissimilar on the surface. These theory fragments have progressed largely independently, not surprisingly given the differences in levels of analysis and key outcomes. Such differences suggest that integrating the two theories posed a nontrivial challenge. Yet, we show a strong conceptual connection. Both address the effects of the positioning of producers in a space of categories on an audience’s understandings. In this paper, we flesh out these connections to clarify the processes that lie at heart of theories of categorization. In particular, we build on a common foundation—a theory of partiality in memberships expressed in a modal language. We demonstrate that this foundation can be used to unify the key claims of both theory fragments.

Behind the scenes we use the formal theory-building tools and framework developed by Hannan, Pólos, and Carroll (2007) and extended by Pólos, Hannan, and Hsu (2010). The key constructions are modal logical models that allow for subtle formalization of key sociological concepts such as legitimation, identity, and social form as consequences of the beliefs of the members of relevant audiences. As we aim to illustrate, this approach to theory building has value for producing coherent, integrative sociological theories.

We begin with an overview of our conceptual approach to categorization. We then highlight key concepts from recent theoretical work by Hannan et al. (2007) on the emergence of categories and forms that figure prominently in the theory we build. We extend this theory to develop a theorem that fits the typecasting imagery developed by Zuckerman and colleagues. Then a few additional considerations, we establish that the McKendrick–Carroll argument also follows from the new framework. We conclude with discussion of additional implications of the general theory.
Preliminaries: Categories and Forms

Much of the recent interest in categorization stems from dissatisfaction with traditional sociological approaches to conceptualizing the notion of form (as applied to organizations and other kinds of producers). In the case of organizations, scholars generally agree that form refers to “those characteristics of an organization that identify it as a distinct entity and, at the same time, classify it as a member of a group of similar organizations” (Romanelli 1991: 81–82). Yet, there has been considerable disagreement about how to approach the task of identifying the presence and boundaries of a form (Hannan and Freeman 1986; Romanelli 1991; Pólos, Hannan, and Carroll 2002). One common approach identifies forms by looking for common patterns of features on the belief that forms can be assessed in purely objective terms. Other research looks to boundary-creating processes such as social network ties and personnel flows to understand form distinctions (DiMaggio 1986; Hannan and Freeman 1986).

Yet, researchers increasingly argue that such approaches lose sight of the importance of the social meanings and interpretations of contemporary audiences (McKendrick and Carroll 2001; Pólos et al. 2002; Baron 2004; Hsu and Hannan 2005; Ruef and Patterson 2009). How can researchers ensure that the forms they study empirically actually represent instances of meaningful social units? An audience-based theory of categories (an antecedent to forms) seeks to provide a resolution to such concerns.

This new approach concentrates on the interface between producers and audiences, and it tries to model what audience members perceive when they “see” a producer and what they expect of the bearers of a category label (Pólos et al. 2002; Hannan et al. 2007). Producer and audience member are conceptualized as roles. Examples of the kinds of actors who commonly play the producer role, of course, includes firms that sell goods and services in markets. But it applies as well to individuals who seek employment, to political parties seeking the favor of donors and voters, and social movements trying to mobilize members and other resources. Similarly, the audience role involves evaluat-
ing and potentially rewarding producers with essential material or symbolic resources. Those who play the audience role generally includes customers, investors, donors, third-party intermediaries such as critics and analysts, and even other actors on the producer side of the interface.

In this approach, a category is a type of collective identity; it involves a typification of commonality. In the simplest case, audience members recognize similarities among a cluster of producers and come to regard these similar entities as members, to varying degrees, of a common (fuzzy) set and agree on a label for the set. In some cases, these typifications become associated with an agreed-upon meaning, in which case we refer to categories. Finally, categories sometimes become forms—that is, they become highly legitimated or taken-for-granted by members of an audience.

This formulation explicitly addresses the messy nature of socially constructed categories. It also, in emphasizing what audience members see, requires the use of new conceptual tools. In particular, we need a formal representation of the degree to which an agent views a producer as fitting a label and its meaning. Such a conceptualization allows modeling of partiality in category memberships and the impact of partiality on category dynamics (Hannan 2010).

A broad range of recent empirical work has adopted important elements of this new approach to conceptualizing categories and forms (but without necessarily casting the argument in a model of partial memberships). We provide a brief (and by no means comprehensive) overview of key lines of research here.¹ One line uses ideas about identity to explain form emergence. For example, Ruef (2000) finds that the distribution of existing organizations in the identity space of the healthcare domain affects the likelihood that new forms with similar identities will emerge. A key implication is that the fit of a novel category within preexisting categorical understandings shapes the likelihood of its acceptance and legitimation. King et al. (forthcoming) study this issue from a

¹Hannan (2010) provides more systematic coverage.
different angle, by examining how the local institutional context of an emerging form shapes which combinations of identity-defining elements get adopted by producers, and thus how the collective identity of the form develops.

Another line of research draws attention to the effect of the activities of external audiences on category evolution. For example, Rao, Monin, and Durand’s (2003) study of the nouvelle-cuisine identity movement in French gastronomy finds that greater sociopolitical legitimacy of movement activities and more widespread theorization of the new cuisine by culinary journalists and critics influenced the propensity of French chefs to “defect” from classical to nouvelle cuisine. And in a study of the market for computer workstations, Kennedy (2005) finds that media outlets, through their coverage decisions, affect how firms in emerging categories identify their rivals and thus how they perceive the categorical structure of their market.

Research has also investigated how external audiences’ perceptions shape critical reception and commercial success. Zuckerman (1999) find that firms that fail to establish themselves clearly as members of financial analysts’ categories (by diversifying across the categories that the analysts use in covering the market) are less likely to receive coverage by a buy-side analyst. Such inattention reduces attractiveness to investors and impairs stock market returns accordingly. Likewise, Hsu (2006) finds that audiences have trouble making sense of Hollywood films that span established genres. Genre-spanners fit poorly with audience tastes and therefore have low appeal. This line of research suggests that categories exert strong constraints on individual producers to conform to the expectations inherent in the meanings of categories.

The strength of category-related expectations (and thus the penalties associated with violating them) can evolve over time with changes in the composition of a category. For example, Rao et al. (2005) find that penalties imposed on chefs who borrow elements from both classical and nouvelle cuisine weaken as many straddle these category boundaries, which suggests that widespread spanning causes shifts in categorical understandings. High levels of category spanning appear to cloud audience members’ beliefs about what it means to a member
of a category. Negro et al. (2010) find evidence of this kind of pattern in their study of critical reactions to elite Italian wines.

The repercussions of category spanning change as the identities of individual producers change. For example, Zuckerman and colleagues (2003) find that conforming to the expectations of a single category increases the likelihood that an inexperienced producer will gain attention from relevant audiences. However, such a simple and clear identity restricts future opportunities outside of a producer’s initial category. So an identity that spans multiple categories can prove more beneficial for experienced producers in allowing a wider range of activities to be acceptable to relevant audiences. In other words, participating in a diverse array of categories incurs weaker penalties for more experienced producers. However, Negro et al. (forthcoming) find the opposite pattern for Italian winemakers. Clearly more needs to be learned about this aspect of typecasting.

These various lines of research share a common conceptualization of categories as collective identities constructed over time through the actions of interested social agents. As such, they can be viewed as fragments of a larger research program. As noted earlier, we believe that learning how these separate fragments fit (or do not fit) together can lead the way for continued development in this area of research. Such work demands careful elucidation of the mechanisms claimed to be operative in the various theory fragments.

Indeed the language used by sociologists to describe processes of identity construction, categorization, and legitimation is notoriously vague. It generally leaves open what it means to have a categorical identity and what it means for a category to become taken for granted. A lack of clarity on such central questions makes it hard to develop cumulative knowledge about these processes. We think much can be gained from hazarding a formal treatment. Accordingly, paper tries to offer a formal and precise statement of mechanisms that can yield the desired implications (theorems) of two lines of work. Of course, whether the formal statement we offer is the only possibility or even the best of several possible formal renderings is a subject for further theoretical research.

To be sure, formalization exacts a price. This approach requires reliance
on explicit simplifying assumptions that might not fit real situations. Instead of seeking (and claiming) realism, we strive to represent the core insights of the theory fragments. If successful, this strategy allows us to learn more how the arguments work and how they can be modified to capture substantively important complications.

**Fuzzy Category Theory**

In their theory of the emergence of categories and forms, Hannan et al. (2007) establish a formal foundation that proves vital to our integration effort. The theory considers a domain as consisting of a dual role structure (producer and audience) and a language that spells out the meanings of these roles. As we noted above, the term producer refers to an agent who produces products or provides services in the domain. Audience members are agents who evaluate these offerings and potentially reward producers of offerings that they find appealing with essential material or symbolic resources. Relevant audience members include organizational insiders (e.g., employees, managers of producer organizations) as well as various kinds of outsiders (e.g., buyers and suppliers, investors, critics, regulators, and other producers).

The basic linguistic objects are labels. Applying labels (such as “university,” “labor union,” or “comedian”) facilitates cognitive processing of and communication regarding producers. Labeling thus appears to be an important step in the social construction of categories. We begin our formal story with a labeling function, defined as a mapping of triplets of audience members, producers, and time points to (the powerset of) the set of available labels.

In the cases of most sociological interest, labels get paired with schemas that articulate what determines a label’s applicability—what it means. Schemas

---

2Hsu and Hannan (2005) and Hannan et al. (2007) review relevant research in cognitive science and discuss implications for sociological theories of categories.

3A powerset of a set is the set of all of its subsets.
provide abstract models or representations of the feature values and/or patterns of social relations that agents regard as consistent with a given label. In other words, schemas establish the meanings (or intensions) of labels. For example, the label “microbrewer” might be paired with a schema including features such as a small-scale operation, use traditional handcrafted methods of production and of traditional ingredients, and so forth.

We refer to a pair consisting of a label and a schema as a type. Following several major lines of work in cognitive psychology and cognitive science, we assume that assessments of producers’ membership in a type can be partial, a matter of degree (Hannan 2010). Based on fit to the configurations of features/relations in a schema, an observer regards producers as full-fledged members of a type, others as having a moderate or low standing as a member, and still others as completely outside the type boundary. In the construction on which we build, the degree of fit of a producer’s characteristics to a schema is reflected in a grade of membership (GoM) function.

Types can have positive, neutral, or negative valuation. The interesting case for issues related to typecasting concerns positive valuation. In such cases, greater fit of a producer’s features with an audience member’s schema yields greater intrinsic appeal (fit with tastes for offerings of that type). In other words, audience members generally prefer offerings that meet their expectations for a category. We represent this idea with the function \( \tilde{\alpha}(l, x, y, t) \) that tells the intrinsic appeal of the offering of producer \( x \) in type \( l \) to audience member \( y \) at time point \( t \). A type has positive value when the expected intrinsic appeal of a producer’s offering increases with the producer’s GoM in the audience member’s meaning of the label.

Some Formal Preliminaries

Taken for grantedness (or constitutive legitimation) is the central process in the theory we advance. A category is legitimated in this sense if agents in the system being analyzed take it for granted as an element in the structure
Although this idea has been very important in theory and research, it has until recently remained a primitive, intuitive idea—it remained unclear exactly what it means to take something for granted. Hannan et al. (2007) and Pólos et al. (2010) analyzed this concept using premises stated in a modal language. They built on the idea that the legitimation of a category means a heavy reliance on defaults. Before we can explain how we do this, we need to establish a few formal notations and introduce some notation.

Modal Operators for Perception, Default, and Belief

We define modal operators for perception, default, and belief. In logic, the term modality originally referred to qualities of the truth of a proposition, especially the possibility and necessity of a statement. The technical apparatus for analyzing logics with operators for possibility and necessity was eventually generalized to treat statements about an agent’s “attitude” toward an object or relation; and the term modality now generally extends to include expressions of perceptions, beliefs, and valuations. We use this term modality in this extended sense.

We refer to an agent’s information state about a factual situation as a set of beliefs. Perceptions contribute to beliefs in an immediate way. What agents directly perceive updates their beliefs. Therefore the temporal order of perceptions matters: more recent perceptions replace older ones if they disagree. Yet, perceptions provide limited descriptions of the actual world in two senses.

First, perceptions can be inaccurate. A wide body of cognitive research suggests that people’s existing schemas shape how they perceive a given situation (DiMaggio 1997). This often works to minimize conflict in more recent and older perceptions, as agents unconsciously overlook or distort perceptions to fit their pre-existing beliefs.

Second, agents generally have partial perceptions of relevant conditions. Partiality can result in some propositions being perceived as true and others false,
while leaving open the truth/falsity of others. Because such partiality generates uncertainty, mechanisms sometimes emerge to eliminate gaps. We propose that agents rely on schematic defaults to “fill in” missing facts when they lack a relevant perception and an applicable default is available.\textsuperscript{4} That is, defaults shape beliefs only in the absence of current perception of the facts in question. And although beliefs based on taken-for-granted assumptions shape information states (and thus behavior), defaults get exposed to revision due to direct perceptions that conflict with the assumed facts.

Pólos et al. (2010) defined a model for the language containing these operators and provided its formal semantics. This model was designed to satisfy the following constraints:

1. perception is partial at all time points;
2. beliefs must be grounded in either perception or taken-for-granted assumptions;
3. as seeing is believing, perception (at least temporarily) overrides earlier beliefs;\textsuperscript{5}
4. defaults shape beliefs (unless there is perceptual evidence to the contrary);
5. lasting beliefs develop if lasting taken-for-granted assumptions are not contradicted by perceptual evidence.

We introduce three new logical constants—modal operators—that are defined for an (arbitrary) audience member $y$ and a sentence (formula) $\varphi$.\textsuperscript{6} We use the following notation for these new logical constants:

\textsuperscript{4}DiMaggio (1997) discusses the sociological implications of research in cognitive science on a key related mechanism: automatic cognition—the implicit, automatic reliance on default assumptions about features embedded in schemas.
\textsuperscript{5}We do not mean that perception cannot be shaped by beliefs, rather that whatever the agent perceives, correctly or incorrectly, overrides beliefs.
\textsuperscript{6}We use the classical logical constants, such as $\land$, $\lor$, $\rightarrow$, $\leftrightarrow$, $\exists$, $\forall$, and $=$ in the usual manner.
P \_y \varphi(t) stands for “The focal agent \_y perceives that \varphi(t) is the case.”

D \_y \varphi(t) stands for “The focal agent \_y takes for granted that \varphi(t) is the case.”

B \_y \varphi(t) stands for “The focal agent \_y believes that \varphi(t) is the case.”

\_y \rightarrow \varphi(t) stands for “The focal agent \_y has taken for granted that \varphi is the case for some period up to and including time t.”

Nonmonotonic Reasoning

The theory on which we build states (some) definitions, postulates, auxiliary assumptions, lemmas, and theorems in a nonmonotonic logic (Pólos and Hannan 2002, 2004). In formal terms, models of arguments are given in terms of sequences of intensions of open formulas. It contains a formal language to represent causal stories and defines a new kind of quantifier, denoted by \( \mathfrak{N} \). Formulas quantified by \( \mathfrak{N} \) state what is expected to “normally” be the case (by default) according to a causal story. The normal case is what we assume to be the case if we lack more specific information that overrules the default. The implications of a set of rules with exceptions, which we call provisional theorems, are the logical consequences of a stage of a theory. Provisional theorems have a haphazard existence: what can be derived at one stage of theory development might not be derivable in a later stage if more specific considerations are brought into the picture. So the status of a provisional theorem differs from that of a causal story. The syntax of the second language codes this difference. It introduces a “presumably” quantifier, denoted by \( \mathfrak{P} \). Sentences (formulas) quantified by \( \mathfrak{P} \) are provisional theorems at a stage of a theory if they follow from the premises at that stage.

Our arguments rely partly on auxiliary assumptions, which make certain analyses tractable. Because auxiliary assumptions have a different status from causal claims that are believed to be true in the world, we mark them with a different quantifier, \( \mathfrak{A} \) (for “assumedly”). They play the same role in inference as does a formula quantified with \( \mathfrak{N} \). We use a non-standard convention to
simplify the expressions for quantification.\(^7\)

**A Modal Model of Taken for Grantedness**

**Defaults and Induction**

Audience members often have partial perceptions of the fit of producers to their schemas for labels. In some cases, they see (or treat as a default) only that a producer claims a label or that some other agent (perhaps a critic or another kind of gatekeeper) applies the label to the producer. Hannan et al. (2007) claim that such situations offer the analytic leverage needed to define legitimation. The key issue is how many schema-consistent features must be checked (in terms of beliefs) before an agent assumes that the remaining unchecked features conform to his/her schema.

This idea can be represented in terms of a test code, a partial segment of a schema that an audience member uses to make inferences about fit to the rest of the schema on which she has no beliefs (perceptions or defaults). If the agent believes that a producer “passes” the test, then she induces that the unperceived/non-default values of schema-relevant features also fit the schema. For example, an agent’s test code for the type “classical French restaurant” might consist of specific ingredients used in dishes and the naming and presentation of dishes (Rao et al. 2003). If a restaurant displays feature values on the test items that fit the agent’s schema, then he will conclude that other

---

\(^7\)We suppress most of the details about the variables of quantification by relying on the following rules.

1. The out-most quantifier of a formula (the one whose scope is the whole formula) binds all the free variables. This assumption allows us to omit the (sometimes long) lists of variables that follow these quantifiers in the standard notation.

2. If the quantifier whose scope is the whole formula is universal, then we omit the quantifier as well.
unobserved schema-relevant features, such as the position of the chef in the restaurant’s power structure and the organization of the kitchen, fit as well.

In technical terms, (believed) satisfaction of a test triggers the default that the unperceived/non-default feature values also satisfy the schema. The definition proposed here modifies the one offered by Hannan et al. (2007, Def. 4.1), which holds that induction “fills in” all non-perceived feature values when an agent perceives that a producer satisfies a test. Focusing on perception alone overlooks the power of existing defaults. Recall that defaults become beliefs when there is no contrary perception. There does not appear to be any reason to think that people will override existing defaults based only the passing of a test on other features. So we refine the earlier conception in line with this intuition. That is, we propose that induction works on features about which an agent has no belief (based either on a perception or a default). For most cases of interest these beliefs will be based on current or recent perception. Nonetheless, it makes the argument more general to allow for situations in which agent’s have defaults based on a prior history of experience with the producer but lacks any current update on the producer’s characteristics.

**Definition 1.** An induction from a test is a situation in which the belief that a producer’s patterns of feature values and/or relations satisfy a test is enough to trigger the default that the values of unchecked features (absent a prior belief to the contrary) also satisfy the schema. (LOT D.4.1\(^8\) slightly modified)

Let \( f_I \) denote an indexed set of \( I \) features and \( f_J \) denote an indexed subset of the relevant features: \( 0 \leq J < I \). We denote the corresponding sets of values of the features as \( S_I \) and \( S_J \), respectively. We use the expression \( f_{i,x,t} \in s_i \) to mean that the \( i \)th feature of the object \( x \) has a value that complies with the

---

\(^8\) This notation indicates that the definition comes from Hannan et al. (2007); it is their Definition 4.1. Postulates and Theorems are denoted by P and T, respectively.
schema $\sigma(l, x, t)$ at the time point $t$; and let $\sigma(l, y, t) = S_I$.

$$\text{induc}(\sigma(l, y, t), t_J) \leftrightarrow \forall i, j, x \left[ (l \in I(x, y, t)) \land (j \in J) \land (i \in I \setminus J) \land \left( \square_y(f_{j,x,t} \in t_j) \rightarrow \neg \square_y(f_{i,x,t} \not\in s_i) \right) \leftrightarrow \Diamond_y(f_{i,x,t} \in s_i) \right].$$

In that case, we refer to $t_J = \{t_j \mid j \in J\}$ as $y$’s test for judging conformity to the schema $\sigma(l, y, t)$, in notation, $\text{TST}(\sigma(l, y, t), t_J)$, and we say that the test has $J$ items.

If every relevant feature must be checked before applying a default, then the agent does not take anything for granted. If only a small fraction of the relevant features must be checked (perhaps only a claim to the label), then the agent uses defaults in a powerful way. These comparisons make the most sense when we consider the minimal test, one that involves the fewest features.

**Definition 2** (Minimal test for induction). The set of values of the $J$ feature values, $t_J$, is $y$’s minimal test for induction for the schema for $l$ at time $t$, in notation $\text{MT}(\sigma(l, y, t), t_J)$, if and only if (1) it is one of $y$’s tests for conformity with the schema; (2) it contains no more test features for the schema than any other of $y$’s tests; and (3) $y$ induces satisfaction of the schema $\sigma(l, y, t)$ on the untested features from this test. (LOT D.4.2)

The relative size of the minimal test for induction for fit to a schema relates directly to the degree of taken-for-grantedness of the label. By size we mean the number of features that belong to the schema or the test.

**Definition 3.** Taken for grantedness.

A. The degree to which an audience member takes for granted that the untested feature values and social relations of a labeled producer conform to a schema for the label is the ratio of the size of the untested part of the schema to size of the whole schema. (LOT D.4.3)

Let $(l, S_I)$ be a type for a focal agent $y$ at time $t$ and let the agent’s minimal
test code for this type have \( J \) elements.

\[
g(l, x, y, t) \equiv \begin{cases} 
(I - J)/I & \text{if } l \in I(x, y, t); \\
0 & \text{otherwise.}
\end{cases}
\]

**B.** The degree of taken for grantedness of a label to an audience member equals the average level of taken for grantedness assigned by the agent to the producers to which the agent applies the label. \((LOT \; D.4.4)\)

\[
G(l, y, t) \equiv \sum_{x|l \in l(x, y, t)} \frac{g(l, x, y, t)}{|\{x|l \in I(x, y, t)\}|},
\]

where \(|\cdot|\) denotes the crisp cardinality of a set, the number of distinct elements that it contains.

This definition sets taken for grantedness to zero if the audience member does not apply the label to the object or needs to see every (nonlabel) feature before making an induction, which is no induction at all—nothing is taken as satisfied by default. It sets taken for grantedness to one if merely applying the label shifts to reliance on defaults about schema-conformity on all other relevant features. In this case, the test on feature values is empty, \( J = 0; \) and the test is passed automatically whenever the agent applies the label.

Some language is needed to express the difference between an agent’s applying a schema to a label and taking conformity with the schema for granted. Hannan et al. use the terms type and concept. When an agent takes-for-granted that members of type will conform to her schema for it, the type is a concept.\(^9\)

**Definition 4.** A concept is a type for which the focal audience member treats conformity to the schema for the type label as taken-for-granted for (nearly) all those producers/products to which he/she assigns the type label. \((LOT \; D.4.5)\)

\[
\text{CONCEPT}(l, y, t) \leftrightarrow ((l, \sigma(l, y, t)) \in \text{ty}(y, t)) \land (G(l, y, t) \geq g \approx 1).
\]

\(^9\)The term “concept” can be regarded as the individual-level analogue of a highly taken-for-granted category, or form (which are aggregated concepts that refer to the perceptions among members of an overall audience).
Incomplete Beliefs and Defaults in Typecasting

Reliance on defaults about concept membership shapes how audience members regard the producers to whom they apply a label. We claim that defaults also underlie the typecasting dynamic that Zuckerman et al. (2003) highlight in their study of the careers of Hollywood film actors. As noted earlier, this research finds that actors who are strongly identified with a single type (genre) of work often find it difficult to obtain future work in other types difficult. Presumably, audience members assume that each type of work requires a distinct set of skills, so clear identification with one type of work implies that an actor lacks the skills necessary for others.

In our understanding, a typecasting process depends crucially on partiality of available information and clashes between applicable schemas. Sometimes agents have full information about the properties of a producer and can tell whether it fits one or another schema. In such a situation, there is no reliance on typecasting—instead the agent relies on direct perception. But, when perception is incomplete, knowledge that a producer fits one type generally gets treated as evidence that it likely does not fit other types whenever the schemas for the applicable categories clash in the sense that high fit to one implies a low fit to the other.

More generally, the typecasting dynamic suggests that the belief that a producer is a member of one type will (1) increase the producer’s appeal in exchanges of that type and (2) prevent acceptance of its membership in others. To build to that multiple-type case, we first need to consider how taken for grantedness affects the assignment of GoMs in cases in which test codes are satisfied but the audience member does not have a belief about schema satisfaction for some relevant features. In doing so, we contrast concepts and mere (i.e., less taken-for-granted) types.

We construct our arguments at the audience-member level. This simplifies our formal model and accords well with Zuckerman’s focus on the reactions of individual agents (investment analysts and casting directors) to category span-
ning. We believe that our formal results can be aggregated audience member by audience member to derive implications for a collective audience.\textsuperscript{10} In the interest of brevity, we do not develop these aggregate implications formally.

In more general, schemas might assign variable weights to features in assessments of fit to labels. Such weights would assign greater or lesser penalties to mismatches on certain features. Likewise, there might be a structure of conditionality among the elements of a schema, meaning that the value of one feature affects what is schema-conforming on another. Allowing this kind of generality makes it impossible to generate simple models that apply across situations. (Empirical work that focuses on a particular domain can, of course, investigate such subtleties.) Therefore we make a simplification in the interest of theoretical analysis. We analyze situations in which fit to a schema can be assessed simply by counting matches and mismatches of the values of relevant features to the schema. We call this a flat schema.

**Definition 5.** An agent’s schemas (for a pair of labels) are flat if higher grades of membership are assigned to objects with more matches to the schema and fewer mismatches.

Let $\sigma(l, y, t) = S_I$, $\sigma(l', y, t') = S_{I'}$; and let $p^+(l, x, y, t)$ and $p^+(l', x', y, t')$ denote the proportion of features values for which $y$ believes schema conformity for the triplets of labels, producers, and time points.

$$\text{FLAT}(l, l', y, t, t') \leftrightarrow \forall x, x' [(p^+(l, x, y, t) > p^+(l', x', y, t'))$$
$$\implies \text{E}\{\mu_{i(t)}(x, y, t)\} > \text{E}\{\mu_{i(t')}(x', y, t')\}].$$

How can we represent the idea that audience members often have incomplete beliefs about schema satisfaction? Because we want to make the argument general and we do not have any prior expectations about patterns, we develop a

\textsuperscript{10}This is not to claim that there are no interesting dynamics at the audience level, just that we do not focus on them. See Hannan et al. (2007, Chapter 5) and Koçak et al. (2009) for treatments of such dynamics.
simple baseline probability model that allows us to compare situations that are alike on average. (We state the elements of the probability model as auxiliary postulates, which means that they serve as analytical conveniences not as claims about the world.) We want this part of the model to be simple, because we do not want it to shape the main qualitative implications of the argument.

The first step defines a common probability over features that no belief is available. Our baseline model holds that the schema-relevant features do not differ in the probability that a belief about conformity to the schema is lacking. In other words, each audience member has available beliefs on a random sample of schema-relevant features.

**Auxiliary assumption 1** (Beliefs available for random samples of features).

*Beliefs about fits to the applicable schemas for the focal labels are available at random in the sense that the probability that the audience member does not have a belief about the value of a feature is the same for all relevant features in the schemas.*

Let \( \sigma(l, y, t) = S_I \).

\[
\forall x, t \left[ \exists \pi \forall i (i \in I) \rightarrow \Pr \{ \neg \exists v \left[ B_y (f_{i,x,t} = v) \right] = \pi_x \} \right].
\]

The key intuition behind typecasting relies on a counterfactual: had the agent had full information about two producers (who differ in their histories of prior labels and memberships), she would have no reason to prefer one to the other. According to the counterfactual, the audience member would generally regard the producers as having equal grade of membership in terms of schema satisfaction. We think that the central challenge formalizing the typecasting argument is building a coherent representation of the counterfactual.

To represent this counterfactual, we assume (as the second element in the baseline probability model) that the two producers being compared would be equally likely to satisfy the relevant schema if the audience member had a positive belief about their feature values. Otherwise, the judgment of fit to schemas would reflect relevant differences between the producers—no typecasting need
occur. We implement this imagery by assuming that the probability that an agent believes that the value of an arbitrary feature conforms to the schema is the same for any two producers being compared. In other words, the producers are equivalent in expected-value terms.

**Auxiliary assumption 2.** The probability that an audience member believes that one of a producer’s feature values satisfies her schema for a label (conditional on having a belief) is the same for all producers.

Let $\sigma(l, y, t) = S_I$.

$$\forall x, t \exists p \forall i \in I \rightarrow \Pr\{\exists y(f_{i,x,t} \in s_i) \mid \exists v(\exists y(f_{i,x,t} = v)) = \rho_x] \}.$$

This probability model, when applied to flat schemas, implies a pattern that agrees with the core intuition about the constraints imposed by typecasting. We develop this implication for a simplified situation that makes the analysis tractable. The simplification considers situations in which each audience member has flat schemas for two labels, $l$ and $l'$, has minimal test codes for each schema, and the probabilities that the audience member has a belief about a $l$-schema-relevant feature and that beliefs indicate schema conformity are equal for the two triplets of producers, audience members, and time points. In other words, we construct a set of worlds that satisfy the theoretically relevant conditions. We define these worlds as follows.

**Definition 6.** Nearest-possible-world condition 1.

$\Phi(t, t')$ indicates that the following nearest-possible-world conditions hold over the time interval beginning at $t$ and ending at $t'$. For the focal audience member:

1. $l$ is a concept over the interval $[t, t']$;
2. the schemas for it do not change over the interval;
3. $l'$ is a type at (at least) the end point of the interval;
4. the schemas for $l$ and $l'$ are flat over the interval;
5. on average, beliefs about the $l$-relevant and $l'$-relevant feature values of all the producers in the domain are incomplete to the same degree for the producers being compared:

$$\exists \pi [\forall x, x' [\pi_x = \pi = \pi_{x'}]];$$

6. the producers fit the schema for $l$ to the same degree within the interval on average:

$$\exists \rho [\forall x, x' [\rho_x = \rho = \rho_{x'}]].$$

With these assumptions in place, we can now begin to derive implications about the consequences of defaults based on inductions from minimal tests. We begin with the case of one label at two time points with a constant schema.

**Theorem 1.** With random availability of beliefs for a single label with a flat schema and with a common probability of forming a schema-conforming belief, agents presumably assign higher grades of membership to producers in the meaning of a label when conformity with the label is more taken for granted.

$$\mathbb{P} [\Phi(t, t') \land (l = l') \land (g(l, x, y, t) > g(l, x', y, t')) \Rightarrow E\{\mu_i(t)(x, y, t)\} > E\{\mu_i(t')(x', y, t')\}].$$

(Although considering the details of the proofs of this and subsequent theorems is not essential for understanding the flow of argument, the proofs do show vividly how the assumptions, definitions, and modal operators interact in shaping typecasting processes. Nonetheless, we place a proofs in an appendix to minimize the amount of technical material in the main text.)

The foregoing theorem shows clearly the impact of growth taken for grantedness for a given label (which we can see by setting $l' = l$). When agents have only partial information about the values of relevant features, growing taken for grantedness causes them to see the producers as more closely fitting these schemas.
Next, we remove the restriction to one label at different time points by comparing membership assignment for types versus concepts (highly taken for granted types).

**Theorem 2.** *When audience members have partial observations on some type-relevant producer characteristics, they presumably assign higher grades of membership in the type to objects when the type is a concept.*

\[
\mathbb{P} [\Phi(t, t') \land \text{CONCEPT}(l, y, t) \land \neg \text{CONCEPT}(l', y, t')
\Rightarrow E\{\mu_i(l)(x, y, t)\} > E\{\mu_i(l')(x', y, t')\}].
\]

The argument behind Theorem 1 when combined with Definition 5 (positively valued type) also implies a corresponding difference in the intrinsic appeal of the offerings forbearers of concept labels than forbearers of (mere) type labels in situations of partial knowledge.

**Corollary 1.** *When audience members have partial observations on some producer characteristics relevant to a positively valued type, they presumably find the offerings of members of the type more appealing when the type is a concept.*

\[
\mathbb{P} [\Phi(t, t') \land \text{CONCEPT}(l, y, t) \land \neg \text{CONCEPT}(l', y, t') \land \text{PVT}(l, y, t) \land \text{PVT}(l', y, t')
\Rightarrow E\{\tilde{\alpha}(l, x, y, t)\} > E\{\tilde{\alpha}(l', x', y, t')\}].
\]

As noted earlier, we focus our arguments on positively valued types, which have the most relevance for typecasting concerns. Of course, if the valuation of the type were negative (e.g., as for “sweatshop,” “ambulance chaser,” and “slave ring,”) then the theorems and corollary in this section would run in the opposite direction.

**Typecasting**

The argument made to this point, together with the dynamic interactions defined for the modalities, yields what struck us as a surprising implication. Sup-
pose that an audience member initially believes that a producer passes the minimal test code for a concept and does not perceive any violations of her schema on the non-test items. Suppose further that this agent later develops the belief that the same producer also passes the minimal test code for a clashing concept and does not display any observable violations of the schemata for that concept. What happens?

To provide a formal answer to this question, we first define clashes between schemas. To simplify what follows, we consider pairs of labels whose schemas clash (for an audience member) but only outside of their minimal test codes.

**Definition 7.** Schemas for a pair of labels clash outside an agent’s minimal test codes for them, in notation \( \text{CLASH}(l, l', y, t) \), if (1) the pair of minimal test codes can be jointly satisfied and (2) the pair of full codes cannot be jointly satisfied.

As we thought about these issues, we first reasoned that a schema clash might block the application of defaults with the consequence of lowered GoMs in both concepts. However, the default modality does not work this way, as we have modeled it (based on general considerations, not the scenario under discussion). Once a default has been set, it has the status of a fact unless and unless and until it gets overridden by a new perception. So, in the scenarios we consider, the audience member treats all of the schema-relevant items as satisfying the schema and also treats the defaults as facts when considering membership in the clashing concept. The result is that she decides that the producer does not fit well the newly relevant (clashing) concept, and she does not alter her judgment of the producer’s typicality in the original concept. This conclusion fits our reading of the typecasting imagery.

This argument can be built on the reasoning behind Theorem 1. The key step in linking this argument to typecasting is constructing meaningful simplifying

---

11 Hannan et al. (2007) define schema clash indirectly with a meaning postulate (MP5.1): “Normally, the higher a producer’s grade of membership in a type whose schema clashes with that of a focal category, the lower the producer’s grade of membership in the focal type.” Unfortunately this construction presumes part of what we want to derive.
assumptions that allow us to capture the key insights. We try to accomplish this task by defining a set of simple worlds that allow the typecasting dynamic to operate without interference from other relevant processes. We define the properties of this set of worlds as follows.

**Definition 8. Nearest-possible-world condition 2.**

$\Psi(t, t')$ indicates that the following closest-possible-world conditions hold over the time interval beginning at $t$ and ending at $t'$. The focal audience member

1. has schemas and minimal test codes for $l$ and $l'$ that do not differ for the time points $t$ and $t'$ and the property that
   (a) the parts of the schemas involved in the minimal test codes for $l$ and $l'$ do not clash;
   (b) the schemas for $l$ and $l'$ that clash beyond their minimal test codes and clashes outnumber non-clashes;

2. applies the label $l$ to the producer $x$ and believes that the producer $x$ passes her minimal test code for $l$ over the relevant time interval;

3. applies the label $l'$ to both producers at the later time point $t'$;

4. either does not apply the label $l$ to the producer $x'$ earlier in the time interval: or does not believe that $x'$ passes her minimal test code for $l$ during the interval.

**Theorem 3.** When two concepts with schemas that clash only outside audience members’ minimal tests for them, membership in one concept at an earlier point in time (1) yields a higher fit to that schema at subsequent times but (2) reduces the fit to the other schema at a later point in time when the audience members do not generally have beliefs about a producer’s conformity to schema on all relevant features.

\[
\mathbb{P} \left[ \Phi(t, t') \land \Psi(t, t') \land (t' > t) \rightarrow E\{\mu_{i(t)}(x, y, t')\} > E\{\mu_{i(t)}(x', y, t')\} \right. \\
\left. \quad \land E\{\mu_{i(t')}^{(l)}(x, y, t')\} < E\{\mu_{i(t')}^{(l')}^{(l)}(x', y, t')\} \right].
\]
As the proof (in the Appendix) makes clear, this result depends upon partiality of beliefs. If the focal agent has complete beliefs about the producer's values on the schema-relevant features, then the process modeled in this implication would not follow.

If the types have positive value, then the pattern claimed by Zuckerman and collaborators immediately follows.

**Corollary 2** (Zuckerman et al. 2003). In the case of two concepts with schemas that clash only outside audience members’ tests for them, membership in one at an earlier point in time presumably (1) enhances the intrinsic appeal of the producer’s offering in the first concept but (2) reduces the appeal of its offering in the other at a later point in time when the audience members do not generally have beliefs about a producer’s conformity to schema on all relevant features.

\[
\mathbb{P} [\Phi(t, t') \land \Psi(t, t') \land (t' > t) \land \text{pvt}(l, y, t) \land \text{pvt}(l', y, t')]
\rightarrow E\{\tilde{\alpha}(l, x, y, t')\} > E\{\tilde{\alpha}(l', x', y, t')\} \land E\{\tilde{\alpha}(l', x, y, t')\} < E\{\tilde{\alpha}(l', x', y, t')\}.
\]

Theorem 3 and Corollary 2 highlight both the benefits and the drawbacks of passing the minimal test code of a highly taken-for-granted type. On the one hand, reliance on defaults means that audience members will assign a high grade of membership to the producer and find its offerings to have high intrinsic appeal. But this restricts the producer’s ability to demonstrate fit with a clashing type in the future. Audience members will rely on prior defaults in the case of partial perception and immediately assume a poor fit with the schema and their tastes for offerings of the clashing type.

**Contrast and Taken for Grantedness**

Our rendering of the typecasting argument has potentially broad implications. Exploring them requires attention to related processes underlying taken for grantedness (legitimation).

The original theory of density-dependent legitimation holds that growth in
the number of producers associated with a category increases its taken for grantedness (Hannan and Freeman 1989). This original formulation, however, did not address the idea that different producers might contribute differentially to the taken for grantedness of a category. To incorporate this possibility and generalize the theory, Hannan et al. (2007) shifted attention from density to what they call contrast. The contrast of a type refers to the average grade of membership among those with positive GoM in the type (from the vantage of a focal audience member). In other words, a high-contrast condition approximates a binary distinction (full membership versus non-membership). We define this notion formally as follows.

**Definition 9.** The contrast of a type (to an agent) is the average grade of membership in the meaning given by the agent to the type taken over all those objects to which the agent assigns the type label. (LOT D.3.4)

\[ c(l, y, t) = \frac{\text{card}\{\mu_i(l)(y, t)\}}{|\text{supp}\{\mu_i(l)(y, t)\}|}. \]

Categories with high contrast stand out sharply against the background. Such contrast reflects a tendency for members of the audience to see the cluster of producers in similar ways. This, in turn, promotes the emergence of consensus among audience members about the meaning of the category (intensional consensus). Hannan et al. (2007) posit that a label’s legitimation for an audience increases monotonically with the degree of intensional consensus about the label. Then it follows in the new theory that legitimation (the average level of taken for grantedness over all pairs of producers and audience members) increases with average contrast in the audience.

Here we focus on another path, one that links legitimation to contrast at the level of the agent (the audience member). The contrast of a type for an audience member is the average of the nonzero grades of membership that the audience member assigns to the objects to which he assigns the type label.

For an individual audience member, high type contrast means that the producers to whom the audience member assigns the type label generally fit the concept schema well (because high contrast means that the audience member
assigns either high or very low GoM in the meaning of the label to the objects in the domain). Cases of poor fit to schemas will generally be viewed as exceptions to the general rule in such situations. Such generic fit causes the audience member to come to take for granted that any producers to which she applies the type label will have schema-consistent features. The probability that beliefs about schema conformity will become defaults therefore increases with contrast.

**Postulate 1.** A type’s expected taken-for-grantedness for an audience member normally increases monotonically (possibly with some delay) with its contrast.

$$\mathfrak{N} \exists u \forall s [(0 < s < u) \land (c(l, y, t+s) > c(l', y, t'+s)) \land (G(l, y, t+s) = G(l', y, t'+s)) \rightarrow E\{G(l, y, t + u) > E\{G(l', y, t' + u)\}\}].$$

With these preliminaries in hand, we return to our main substantive focus, the effects of typecasting.

**Form Emergence**

The intuition underlying our rendition of the typecasting theorem—that beliefs about a producer’s membership in one concept constrain beliefs about fit with others—can be usefully extended to shed light on other aspects of the dynamics of types and concepts. In this section, we demonstrate this by considering how this process of induction and typecasting relates to key findings on identity-based form emergence. We continue here to build the model at the level of the audience member.

In their seminal study of the disk-array producers, McKendrick and Carroll (2001) found that a label for an emerging category gains more legitimation from *de-novo* entrants (those with no prior history for the audience member) than from *de-alio* entrants (already existing producers who the audience member regards as diversifying from some other categories). They reasoned that audience members perceive de-novo entrants as being more focused on the activities associated with the label than their diversified counterparts and thus
contribute more to audience members’ understandings of what it means to be a type member. This finding had a serious impact on the thinking of organizational theorists. However, so far it has not been integrated into the formal theoretical framework of categories.

**Definition 10. De-novo and de-alio entrants.**

A. *De-novo entrant:* \( \text{DE-NOVO}(l, x, y, t) \leftrightarrow (l \in I(x, y, t)) \land \forall t'[(t' < t) \rightarrow (l(x, y, t) = \emptyset)] \).

B. *De-alio entrant:* \( \text{DE-ALIO}(l, x, y, t) \leftrightarrow (l \in I(x, y, t)) \land \forall s[(s < t) \rightarrow (l \notin I(x, y, s))] \land \text{MTST}(\sigma(l', y, t), t_j^r) \rightarrow \Box_y (f_{j,x,t} \in t_j^r) \).

C. *The number of de-novo entrants:* \( e_n(l, y, t, t^r) = |\{x \mid (t \leq u < t^r) \land \text{DE-NOVO}(l, x, y, u)\}| \).

D. *The number of de-alio entrants:* \( e_a(l, y, t, t^r) = |\{x \mid (t \leq s \leq t^r) \land \text{DE-ALIO}(l, x, y, s)\}| \).

To capture McKendrick and Carroll’s core insight, we focus on the case of de-alio entrants with clashing memberships. Just as the typecasting dynamic rested on the assumption that distinct genres correspond to different schemas defined for characteristics such as skill sets and social relations, the story about de-novo and de-alio entrants rests on the idea that the different producer types are associated with *clashing* schemas. A producer has de-novo status in a label to an audience member who applies the label at the time point and has not previously applied any label to that producer. A producer has de-alio-clashing membership in a label if the audience member applies the label at the time point and also continues to apply a label assigned earlier in a clashing concept and believes that the producer passes the minimal test for the clashing concept.

When an audience member assesses the fit of a de-alio entrant from a clashing concept to a focal type, the process of induction that drives the typecasting dynamic (as stated in Theorem 3) should operate. Membership in the clashing concept reduces the (believed) fit of the de-alio entrant in the focal type.
This puts the de-alio entrant at a disadvantage as compared with a comparable de-novo entrant (when there is random availability of beliefs and a common probability of forming a schema-conforming belief for each producer).

**Theorem 4.** De-novo entrants presumably have higher expected grades of membership in audience members’ types than do de-alio entrants with memberships in clashing concepts.

\[ \mathbb{P} [\Phi(t, t') \land \Psi(t, t') \land \text{DE-NOVO}(l, x, y, t) \land \text{DE-ALIO}(l, x', y, t') \rightarrow \text{E}\{\mu_{li}(x, y, t')\} > \text{E}\{\mu_{li}(x', y, t')\}]. \]

This grade-of-membership disadvantage for de-alio entrants also results in a disadvantage in terms of the appeal of their offerings to audience members (according to the rule chain supporting Theorem 4 and Definition 3).

**Corollary 3.** The offerings of de-novo entrants presumably have greater intrinsic appeal than do those of de-alio entrants with memberships in clashing concepts.

\[ \mathbb{P} [\Phi(t, t') \land \Psi(t, t') \land \text{DE-NOVO}(l, x, y, t') \land \text{DE-ALIO}(l, x', y, t') \rightarrow \text{E}\{\tilde{\alpha}(l, x, y, t')\} > \text{E}\{\tilde{\alpha}(l, x', y, t')\}]. \]

Corollary 3 might appear to contradict empirical findings that de-alio producers tend to outperform de-novo producers when they initially enter a new market (e.g., Hannan and Freeman 1988; Mitchell 1994; Carroll, Bigelow, Seidel, and Tsai 1996; Hannan, Carroll, Dobrev, and Han 1998; Klepper and Simons 2000). However, the performance advantage of de-alio entrants can be largely attributed to differences in capabilities and endowments. De-alio producers generally enjoy significantly greater resources and experience than their de-novo counterparts, which enables them to better engage their targeted audience (Hannan et al. 2003; Hsu et al. 2009). So, while the higher grade of memberships of de-novo entrants in a category leads to their greater intrinsic appeal, the superior capacities for audience engagement that de-alio entrants enjoy often counteracts and overwhelms this advantage.
McKendrick and Carroll (2001) suggest that the extent to which audience members perceive a set of entrants as having a type focus contributes to the taken for grantedness of the type. This idea of perceptual focus can be analyzed in terms of contrast, as defined in the previous section. Because de-novo entrants have a higher expected grade of membership in an audience member’s concept, they naturally contribute more to type contrast.

**Lemma 1.** De-novo entrants presumably contribute more to the contrast of audience member’s type than de-alio entrants from clashing concepts.

Let the number of entries in two labels be the same over the relevant period.

$$
\mathbb{P} [\Phi(t, t') \land \Psi(t, t') \land (t' > t) \land (e_{n}(l, y, t, t') > e_{n}(l', y, t, t')) \land (c(l, y, t) \geq c(l', y, t)) \\
\rightarrow \mathbb{E} \{c(l, y, t')\} > \mathbb{E} \{c(l', y, t')\}].
$$

The core insight of the de-novo/de-alio story can now be seen to be an implication of the preceding argument.

**Theorem 5** (McKendrick–Carroll 2001). The expected (possibly delayed) contribution of a set of entrants to the taken for grantedness of a type for an audience member presumably is higher for de-novo than de-alio entrants.

Let the number of entries in two labels be the same over the relevant period.

$$
\mathbb{P} [\Phi(t, t') \land \Psi(t, t') \land (e_{n}(l, y, t, t') > e_{n}(l', y, t, t')) \land (c(l, y, t) \geq c(l', y, t)) \\
\rightarrow \exists u[(u \geq t') \land (\mathbb{E} \{G(l, y, u)\} > \mathbb{E} \{G(l', y, u)\})]].
$$

**Discussion**

This paper aims to build a formal theory of legitimation for a multi-category context and to use it to sharpen our understanding of the processes underlying typecasting and form emergence. We claim that induction of fit to schemas lies at the core of the issues considered by these two theories. Audience members rely on defaults regarding schema-relevant features to define the concept memberships of producers in the face of partial information. Following Hannan et
al. (2007), we proposed that the concept of taken-for-grantedness is linked to induction: what defines the taken-for-grantedness of a type is the degree to which audience members automatically fill in defaults for producers who exhibit some minimal criteria for membership for that type. In the case of a single category, one implication is that audience members presumably assign a higher GoM to a producer for a concept versus mere type. As a result, they also presumably find the producer more appealing in the former case.

Generalizing this story to the case of multiple memberships sets the foundation for modeling the typecasting dynamic. When audience members associate a producer with one concept at an earlier point in time, the (believed) fit of that producer in that concept will be enhanced, but its fit in other types will be reduced. A similar process of induction can explain the findings by McKendrick and Carroll (2001) on form emergence. De-alio entrants are generally believed to have worse fit with a concept schema because they already belong to a clashing concept. Hence de-alio entrants get assigned lower GoMs in a concept than their de-novo counterparts, and they contribute less to the taken for grantedness of a type.

The intuition underlying both theories relies on a counterfactual: with complete information about the producers under comparison, audience members would assign equal grades of membership in the relevant types. Both theories suggest that, by relying on defaults to cope with incomplete information, audience members become biased in systematic ways in their perceptions.

Our formalization also shows the importance of the relationships among categories. Without some clash between the codes of the relevant categories, there would be no reason to expect such systematic bias in beliefs. Zuckerman et al. (2003) argue that lay theories of skill provide the foundation for typecasting dynamics to emerge. For classification structures such as types and concepts to be restrictive, audience members must believe that the features or skills necessary to be a member of one differ from those required of others. (If they must do so, then they have little reason to use different labels and different schemas for them.) They must also believe that significant type-specific investments are
necessary to acquire the core features of each, meaning that a so-called principle of allocation tradeoff operates (Hsu 2006). Reliance on these assumptions leads audience members to presume that a producer who has already demonstrated a fit with one category must lack the features necessary for others.

Clearly, a formal translation of this insight would require attention to the relationship between the membership schemas for different categories. But to capture the core intuition, we needed to formalize this in a way that audience assumptions (rather than their direct perceptions) would drive the main dynamic. More specifically, we needed audience members to assume incompatibility of membership for producers who might actually possess features consistent with the types in question. Thus, clash in codes between concepts should matter when some features are unobserved. Our treatment of induction provides a sensible way to model this by allowing us to focus on code clash outside of an audience member’s minimal test criteria.

This thought process also led us to consider whether induction and code clashes play in form emergence. Their natural language theorization did not explicitly consider the relationship between the categories de-alio producers span. Yet, our framework suggests that clashes between spanned categories must also be present for bias against de-alio producers to emerge. We argue that the presumption of code clashes outside the minimal test criteria provides a way to understand the intuition underlying this theory.

A core tool in our analysis is the use of modal models. Following Hannan et al. (2007), we proposed that audience members rely on defaults when they encounter producers who pass their minimal tests for a type. In the case of highly taken-for-granted types, these minimal test codes are very small and processes of induction are common. However, for less taken-for-granted types, a large test means the audience member will not assume much about conformity with the concept schema. In such cases, the perception operator largely applies, and only partial membership will be assigned when an audience member lacks a belief about the values of some schema-relevant features. Together, these modalities capture in a very specific way what is distinctive about membership in
highly legitimated types and membership in multiple market types and concepts (categories and forms).

We developed our model at the level of the agent by considering the audience member’s application of labels and assessment of fit to his/her own schema for the label. As we noted above, these results have implications at the level of the audience as a whole. If the members of an audience come to substantial agreement about the meaning of a set of labels, they will generally make similar assessments of fit of producers to schemata and engage in induction based on similar observations. Hence, the line of argument we presented in this paper applies *mutatis mutandis* to a comparison of categories and forms, the audience-level parallels of types and concepts.

Audience members do not always come to agree on the meaning of a label, however. Kennedy, Lo, and Lounsbury (forthcoming), for example, observe that multiple meanings of the label “nanotechnology” have co-existed and competed with one another since the early 1990s; one view is largely associated with science fiction and the radical predictions of Eric Drexler, while a competing view defines nanotech in terms of “materials with special but explainable properties.” In such cases, individual audience members could hold strong defaults coinciding with one versus the other schema. At the level of overall audience, however, disagreement and contestation over meanings would likely prevent the label from attaining the status of a taken-for-granted form (as appears to be the case for this example). Capturing such dynamics (and better integrating them with other theory fragments on categorization) requires a language and framework for characterizing variation in audience-level consensus over the meaning of a label (see Hannan et al. 2007 for one approach to doing so).

In a similar spirit, our theory focuses on how clashes in category concepts can lead to typecasting for individual producers who span categories. Yet, the picture becomes more complicated when one considers dynamics at the category level. As noted earlier, research on category spanning suggests that increased spanning activities shifts categorical meanings and weakens boundaries; at high levels, it clouds the meanings associated with labels. (Rao et al. 2005; Negro
et al. 2010). Widespread spanning might also encourage the formation of new categories, as boundaries get redrawn around hybrid members (Kennedy, Lo, and Lounsbury forthcoming). As these possibilities suggest, an important next step would extend our framework to consider dynamics at an aggregate level.

More broadly, this integration project pushed us to identify, articulate, and check the consistency of principles that likely apply to other identity-based theories of market categorization. Theories about market categorization often concern counterfactuals—the way an agent’s position in the market’s role structure is decoupled from its actual features and exerts independent constraints on its opportunities (Zuckerman et al. 2003; Hsu et al. 2009). Our model building highlights how distinct theories conceptualize this issue in similar ways with regards to the constraints of classificatory memberships. In particular, we demonstrate that induction shapes these counterfactual dynamics. In doing so, we provide a path for articulating similar processes for theories of social structuration more generally. For example, similar inductive processes likely feature in processes involving status (Podolny 1993; 1994), reputation (Fombrun and Shanley 1990; Rao 1994), and role structures more generally (White 1981; Leifer 1988). Attention to this fundamental social process might provide a way for researchers in these diverse areas of research to better integrate their findings.

Another promising area for integration involves research on form emergence that emphasizes the influence of producers, media outlets, and social activists on the creation and promotion of new categories (e.g., Rosa et al. 1999; Lounsbury and Rao 2004; Schneiberg et al. 2008; King et al. forthcoming). To give one example, Kennedy’s (2008) work on media co-mentions finds that firms can influence how audience members perceive them by the choice of which rival firms to mention in their press releases. By referencing a small number of rivals, firms in new markets help emphasize similarities among those firms and create the perception that a new, meaningful category exists. Translated to our framework, this suggests that a set of producers that consistently claim each other (and not other firms) as rivals likely produce high type contrast—they will stand out sharply against the market background as a coherent new category. They can
thus be expected to have higher taken-for-grantedness than a comparable set of producers with more heterogeneous claimed-comparison sets.

In addition to guiding integration efforts, the theoretical framework proposed here can be used to generate new theoretical insights. For example, a natural extension of the process of induction and typecasting would consider the domain-level implications of variation in the degree of taken for grantedness, or institutionalization, of categories. Research on systems of categorization has typically focused on systems that are already well established and highly institutionalized. One notable exception is Ruef and Patterson’s (2009) work on credit evaluations in the U.S. during 1870–1900. This research finds that the institutionalization of a broad system of categories shapes how key audience members perceive category boundaries and react to boundary spanning. In particular, credit reporters attended more to such boundaries and penalized more strongly boundary spanners as the prevailing system of categorization in the U.S. became more firmly established.

Yet, even beyond explaining the consequences of spanning, the theory developed in this paper suggests that institutionalization might influence whether such organization with broad niches are perceived to bridge category boundaries in the first place. By increasing reliance on defaults about membership, institutionalization increases the tendency of audience members to typecast producers into a single type. As a result, the overall contrast of categories within a domain will increase with the overall level of institutionalization of categories.

This framework could also be used to generate new substantive insights regarding the relationships between codes for different market types. Available treatments of multiple memberships either do not explicitly consider relationships among the categories or focus exclusively on oppositional categories. Yet, real-life social structures contain richer sets of inter-category relationships. Our framework provides a concrete way to conceptualize relationships between category codes and provides an avenue for future research development on multiple-category memberships. For example, some categories appear to overlap in expected features and seem complementary, while others do not directly clash
but contain features that are largely unrelated. An important direction for future research would develop a deeper understanding of how such inter-category relationships shape market dynamics.
Appendix: Proofs

Testing what follows from the premises in a stage of a theory in the nonmonotonic logic we use operates on representations of arguments in the form of “rule chains.” The links in these chains are strict rules, definitions, auxiliary assumptions, and causal stories. The chains start with the subject of the argument and terminate with the purported conclusion of the argument (the consequence to be derived). In nonmonotonic inference, different rule chains—each representing an argument embodied in the state of the theory—might lead to opposing conclusions. The testing procedure determines whether any inference can be drawn at all and, if so, which one. Such testing requires standards for assessing whether a pair of relevant rule chains is comparable in specificity and for determining specificity differences for comparable chains. In the case of this paper, the available premises and definitions all point in the same direction; we do not see any rule chains that point to opposing conclusions. Thus all that is required is that we establish a rule chain that connects the antecedent and consequent in a claimed theorem.

Theorem 1

Proof. A rule chain yielding this theorem can be constructed as follows. Under the baseline probability model, the simplification stated in $\Phi(t)$, and the absence of induction, the expected ratio of beliefs that feature values fit the schema for the label to positive beliefs about the relevant feature values are the same for the two situations being compared, the expected ratio equals $\rho$. Given the restriction to a flat schemas, this implies that $E(\mu_{(l)}(x, y, t)) = E(\mu_{(l)}(x', y, t'))$ in the absence of induction. So the only systematic difference between these cases must be due to induction. In particular, if the expected number of inductions of schema satisfaction is greater for one situation than the other, then the expected grade of membership is higher for that situation.

Let the random variable that records the number of inductions be denoted
by \( \text{in}(l, x, y, t) \). By the law of total probability,

\[
E\{\text{in}(l, x, y, t)\} = E\{\text{in}(l, x, y, t) \mid \text{min. test for } l \text{ passed}\} \cdot \text{Pr}\{\text{min. test for } l \text{ passed}\} \\
+ 0 \cdot (1 - \text{Pr}\{\text{min. test for } l \text{ passed}\}),
\]

because no induction takes place if the minimal test is not passed. Under the baseline probability model stated in Auxiliary Assumptions 1 and 2 and \( \Phi[t, t'] \), the probability that \( x \) passes \( y \)'s minimal test for \( l \) at time point \( t \) equals \( (\pi \rho)^J \).

Because inductions can only apply to features outside the test code (of which there are \( I - J \) for the label \( l \) at time \( t \)) for which the audience member does not have a belief about the value of the feature. The probability of not having a belief on a feature is \( 1 - \pi \). So the expected number of inductions, conditional on passing the minimal test for \( l \) at \( t \) is \((I - J)(1 - \pi)\). Similar calculations yield \( E\{\text{in}(l', x', y, t')\} = (I' - J')(1 - \pi) \cdot (\pi \rho)^{J'} \). The rule chain supporting the theorem requires that the expected number of inductions for \( l \) at \( t \) exceeds that for \( l \) at \( t' \), which requires that \((I - J)(1 - \pi) \cdot (\pi \rho)^J > (I' - J')(1 - \pi) \cdot (\pi \rho)^{J'} \). Dropping the common multiplier \((1 - \pi)\) and noting that \( I = I' \) under the restrictions given in the antecedent, we check whether \((I - J)\kappa^J > (I' - J')\kappa^{J'} \), where \( \kappa = \pi \rho \). After rearranging terms, we must show that

\[
\frac{I - J}{I' - J'} > \kappa^{J' - J}.
\]

By Definition 5, \( g(l, x, y, t) > g(l', x', y, t') \) yields \((I - J)/I > (I' - J')/I' \) and the antecedent states that \( l = l' \). Together these conditions imply that \( J' > J \). This latter inequality in turn implies that \((I - J)/(I - J') > 1 \) and \( \kappa^{J' - J} < 1 \) (because \( \kappa = \pi \rho \) and the antecedent in the formula stating the lemma states that both \( \pi \) and \( \rho \) lie between zero and one). So the expected number of inductions is higher for \( l \) at \( t \), which implies that \( x \)'s expected fit to \( y \)'s meaning of \( l \) is higher at \( t \) than at \( t' \).
Theorem 2

Proof. Under the baseline probability model, the simplification stated in $\Phi(t)$, and the absence of induction, the expected fit of the two producers to the agent’s schemas for the two labels is the same (because the codes are flat and the probabilities that the agent forms a belief about the values of each relevant feature and the probabilities that any such belief is one of conformity with the applicable schema are the same). When the label marks a type but not a concept, then there is less reliance on defaults than when the label marks a category, for which reliance on defaults is pervasive. This means that the agent will fill in by default a higher proportion of feature values as schema-conforming in the latter case.

We set $g$ in Definition 4 to unity for purposes of the proof. This implies that the minimal test code for $l$ has zero elements, $J = 0$. Then, the expected number of inductions (as given in the proof of Lemma 1) equals $I(1 - \pi)$, because the minimal test if passed by mere application of the label. And the expected proportion of features on which schema conformity is induced equals $I(1 - \pi)/I = 1 - \pi$. For the type, not-concept $l'$, the expected number of inductions is $(I' - J')(1 - \pi)(\pi \rho)^{J'}$ and $J > 0$ (or else $l'$ would be a concept). The expected proportion of features on which schema conformity is induced equals $(I' - J')(1 - \pi)(\pi \rho)^{J'}/I'$.

This difference in the expected number of inductions makes $E\{p^+(l, x, y, t)\} > E\{p^+(l', x', y, t')\}$. By the definition of flat schemas (Definition 5), the consequent in the theorem follows immediately. \hfill \Box

Theorem 3

Proof. In the absence of induction, the expected fits of the producers is the same for each label under the assumptions stated in the definition of $\Psi[t, t']$, because the audience member’s schemas for the labels are flat and the probability that a schema-relevant feature will be observed is the same as is the probability that a positive belief will be one of schema conformity for each label. Induction can
produce both increased fit (when the feature value induced fits the schema) and reduced fit (when the induction goes the other way).

In the case of the first term in the consequent (fit to \( l \)), the result follows from the assumptions that (1) the audience member believes that \( x \) passes the minimal test at all points in the interval and (2) no information is available about such a belief for the second producer, \( x' \). It then follows that induction normally increases the expected proportion of features with believed conformity to the agent’s schema and thereby increase the expected fit to \( l \) for the producer \( x \) relative to \( x' \).

Under the assumption that clashes outnumber non-clashes, the expected net effect of induction is to reduce the expected proportion of features with believed conformity to the agent’s schema and thereby lower the expected fit of \( x \) to the second label (\( l' \)) relative to that of \( x' \).

**Theorem 4**

*Proof.* According to Definition 10, a producer is a de-novo entrant in a label if the audience member applies the label at that time point and does not apply any label to the producer at any earlier time point. A producer is a de-alio-entrant (from a clashing concept) if the audience member applies the focal label to the producer and has earlier applied to it the label of a clashing concept. In such a comparison, the rule chain that supports Theorem 2 applies; and this rule chain yields the conclusion.

**Lemma 1**

*Proof.* This lemma follows is an immediate implication of Theorem 4, which tells that each de-novo entrant has higher expected GoM in the audience member’s type than does a de-alio entrant. Definitions 11 and 12 tell that an entrant with higher GoM increases contrast more than does one with lower GoM, which implies that a de-novo entrant adds more to contrast. Addition over entries preserves this inequality, given the stipulation that the number of de-novo entries
is at least as great as the number of de-alio entries.

\[ \Box \]

**Theorem 5**

*Proof.* The rule chain supporting this implication employs Lemma 3 and Postulate 1. Because Lemma 3 yields a difference in the expected contrasts at the end of the period \([t, t']\) and Postulate 1 states a delayed effect of contrast on expected taken for grantedness, we express the desired theorem as holding for expected taken for grantedness at some time at or after \(t'\) (reflecting the possible delay in the effect of contrast on taken for grantedness).

\[ \Box \]
References


---


