Social Network Profiles as Taste Performances

Hugo Liu

Media Laboratory and Program in Comparative Media Studies
Massachusetts Institute of Technology

This study examines how a social network profile's lists of interests—music, books, movies, television shows, etc.—can function as an expressive arena for taste performance. By composing interest tokens around a theme, profile users craft their "taste statements." First, socioeconomic and aesthetic influences on taste are considered, and the expressivity of interest tokens is analyzed using a semiotic framework. Then, a grounded theory approach is taken to identify four types of taste statements—those that convey prestige, differentiation, authenticity, and theatrical persona. The semantics of taste and taste statements are further investigated through a statistical analysis of 127,477 profiles collected from the MySpace social network site between November 2006 and January 2007. The major findings of the analysis include statistical evidence for prestige and differentiation taste statements and an interpretation of the taste semantics underlying the MySpace community—its motifs, paradigms, and demographic structures.

doi:10.1111/j.1083-6101.2007.00395.x

Introduction

The materials of social identity have changed. Up through the 19th century in European society, identity was largely determined by a handful of circumstances such as profession, social class, and church membership (Simmel, 1908/1971a). With the rise of consumer culture in the late 20th century, possessions and consumptive choices were also brought into the fold of identity. One is what one eats; or rather, one is what one *consumes*—books, music, movies, and a plenitude of other cultural materials (McCracken, 2006).

New emphasis on taste and cultural consumption frees identity from some of its traditional socioeconomic limitations (Grodin & Lindlof, 1996). The milieu of cultural interests one creates for oneself can even be transformational, because cultural consumption not only "echoes" but also actively "reinforces" who one *can* be (Csikszentmihalyi & Rochberg-Halton, 1981). In the pseudonymous and text-heavy

online world, there is even greater room for identity experimentation, as one does not fully exist online until one writes oneself into being through "textual performances" (Sundén, 2003).

One of the newest stages for online textual performance of self is the Social Network Profile (SNP). The virtual materials of this performance are cultural signs—a user's self-described favorite books, music, movies, television interests, and so forth—composed together into a *taste statement* that is "performed" through the profile. By utilizing the medium of social network sites for taste performance, users can display their status and distinction to an audience comprised of friends, co-workers, potential love interests, and the Web public. Although social network sites are relatively new, SNP taste performance can be seen as an instance of what sociologist Erving Goffman (1959) termed everyday performance. Successful performers are "aware of the impression they foster." Thus, taste statements need to be crafted so as to stand up to the scrutiny of an audience that is able to "glean unofficially by close observation" (Goffman, 1959, p. 144).

This study adopts a semiotic framework to investigate how taste statements—the high-level outcome of taste performance—can be understood in terms of the expressivity of interest tokens. First, SNPs are described as a medium, and previous work on SNP analysis is reviewed. Second, a review of the taste literature identifies socio-economic and aesthetic influences on taste. Third, according to cultural semiotics, the semantics of taste expression are attributed dually to *paradigms* (e.g., socio-economic paradigms such as "educational capital" and aesthetic paradigms such as "heartwarming versus alienating") and to *syntagmatic rules* governing how interest tokens are combined (e.g., via "expressive coherence"). Fourth, to test research hypotheses about taste influence, taste semantics, and taste statements, an empirical analysis of 127,477 MySpace profiles is undertaken. Following a grounded theory approach, a pilot study found four types of taste statements by qualitative analysis of a small sampling of profiles. Then, applying statistical measures and Principal Components Analysis to all 127,477 profiles, the motifs and paradigms of the MySpace taste community are interpreted, and evidence for taste statements is inferred from the statistics.

Social Network Profiles

Since 2002, hundreds of social network sites have launched with both professional (e.g., LinkedIn) and non-professional (e.g., MySpace) orientations. Unlike most professional sites, non-professional sites typically feature users' interests, so they are the only ones relevant to the present study. Four of the largest non-professional sites in the English-speaking world are MySpace, Facebook, Friendster, and Orkut. While reports vary as to how many millions of users are on these sites (comScore, 2007; Lenhart & Madden, 2007; Nielsen-NetRatings, 2006), MySpace is thought to be the largest (comScore, 2007). Taste data from MySpace are analyzed for the present study.

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

MySpace is an online community that early adopters helped shape into a music-friendly place where hipsters, indie bands, and fans could network and socialize with one another (boyd & Ellison, this issue). It has many features traditionally associated with online communities, such as forums, user groups, network structure, and highly customizable user profiles. Figure 1 depicts the typical layout of a MySpace SNP.

A MySpace profile provides many ways for users to express their tastes. Textually, users can complete forms to provide demographic details and lists of cultural interests; they can also write about themselves in free text. Users can choose photos and customize the look and feel of their profiles using code they find online (Perkel, in press). According to Donath and boyd (2004), a user's friend connections speak to their identity—the public display of friend connections constitutes a social milieu that contextualizes one's identity. The act of "friending" others, and choosing the subset of these friends to display in the so-called "Top 8," constitute identity performances, because they are willful acts of context creation (boyd, 2006). The present study complements scholarship on the identity implications of friend connections by exploring the taste implications of the SNP's lists of cultural interests.

Cultural interests are organized into categories in SNPs. Five of the six categories displayed by MySpace—general interests, music, movies, television, and books—are shared by Friendster, Facebook, and Orkut. The use of these five categories in a SNP originated with Friendster, although the faceted display of interests in a personal profile has a longer history with online dating sites (Fiore & Donath, 2005). It is conventional for users to fill in an interest category with a punctuation-delimited list of interest tokens (cf. Figure 1, left), although social network sites do not explicitly enforce this.

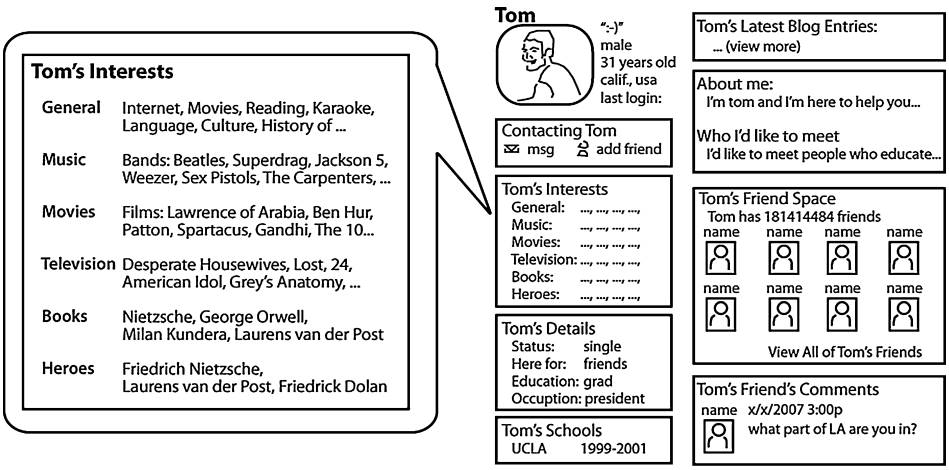

Figure 1 The typical layout of a MySpace profile, filled in with details from MySpace founder Tom Anderson's profile (www.myspace.com/tom), retrieved June 1, 2007

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Previous research has implicated interest tokens as markers of taste and social identity. Liu and Maes (2005) computed statistics on interest token occurrences across 100,000 Friendster and Orkut SNPs and found that tokens exhibited a high degree of clustering. Working with the same dataset, Liu, Maes, and Davenport (2006) showed that clustering interests at a high level of abstraction produced "taste neighborhoods," each with a distinct identity theme (e.g., fashionista). Paolillo and Wright (2005) reported similar findings when they analyzed interest token occurrences across 21,000 user profiles from LiveJournal, a weblog hosting service. Using Hierarchical Cluster Analysis (HCA), they identified nine emergent clusters of interests and interpreted each grouping as corresponding to a subculture or theme (e.g., "goth subculture," "romance"). Then, using Principal Components Analysis (PCA) to articulate the major dimensions of variation in the data, they interpreted each dimension as capturing a paradigm comprised of two opposing motifs (e.g., dimension three: "aesthetic sensibility" versus "general sociability"). From this prior work, a first hypothesis is formulated:

H1: Major dimensions of variation in interest token occurrence data for MySpace SNPs will capture interpretable paradigms and opposing motifs.

Taste and Its Influences

Why does one like what one likes? According to the literature reviewed below, one's tastes are influenced both by socioeconomic and aesthetic factors. Socioeconomic factors—such as money, social class, and education—can shape tastes, because access to cultural goods may require possession of these various forms of capital. Aesthetic factors—such as paradigms of personality (e.g., degree of sarcasm), sentiment (e.g., utopian versus dystopian), and identity (e.g., degree of fashionableness)—define motifs toward which one's tastes may gravitate.

Socioeconomic Influences

Thorstein Veblen's (1899) *Theory of the Leisure Class* was the first work to describe the social motivations for cultivating taste and performing one's tastes for others. Drawing on his observations of the leisure class of his era, Veblen theorized that the tastes of that class, and especially its tendency toward *conspicuous consumption* of costly goods, were driven by the desire to assert and vie for status and honor. The "high class" gentleman "is no longer simply the successful, aggressive male—the man of strength, resource, and intrepidity. In order to avoid stultification he must also cultivate his tastes" (p. 74). Veblen argued that taste has real social utility because it signals prestige and presents an opportunity for differentiation. The fact that luxury goods are costly to procure and leisurely pursuits require ample free time makes the display of taste a more reliable signal of prestige than signals such as verbal declarations, e.g., "I have prestige" (Donath, 1999). This leads to a pair of hypotheses:

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

H2a: MySpace users will craft their SNP lists of interests so as to assert their prestige.

H2b: MySpace users will craft their SNP lists of interests so as to differentiate themselves from their peers.

Veblen also observed that personal tastes gravitate toward class norms; he argues that consumption norms of the high class "will to some extent shape [men's] habits of thought and will exercise a selective surveillance over the development of men's aptitudes and inclinations" (p. 212). From this premise, Pierre Bourdieu's (1984) influential work, *Distinction: A Social Critique of the Judgment of Taste*, argued that taste is central to class identity, and vice versa. Since the members of a socioeconomic class experience the same economic, cultural, and educational boundaries, Bourdieu postulated that the taste norms of the class should greatly shape its members' tastes. To validate his theory, Bourdieu analyzed data from a lifestyle survey of some 1,200 French residents conducted in the 1960s. Correlating survey respondents' tastes to factors such as profession, income, and education, Bourdieu used factor statistics, akin to Principal Components Analysis, to compute two-dimensional "maps" of taste. Each map visualized taste similarity relationships between different demographic groups. Reading these maps, Bourdieu found that the semantics of several major dimensions reflected levels of economic and cultural capital. This was interpreted as strong evidence that individuals' tastes are strongly determined by their socioeconomic circumstance. Bourdieu's thesis leads to the first of a pair of hypotheses about influences on taste:

H3a: Variation in the taste norms of various demographic groups on MySpace can be accounted for by the socioeconomic capital associated with each group.

Aesthetic Influences

Anglo-American scholarship on taste since Bourdieu has largely downplayed socioeconomic determinism, instead relying upon aesthetic groundings. Thornton (1996) observed that tastes of U.K. "club cultures" are governed not by socioeconomic capital but by subcultural capital—access to and knowledge of underground and cult markers of cool.

Gans (1999) suggested that contemporary American society could be stratified into echelons of taste, called "taste publics," but that each echelon is defined by cultural and aesthetic commonality rather than by shared economics. Gans proposed five taste publics: high culture, upper-middle culture, lower-middle culture, low culture, and quasi-folk culture. This organization can be contrasted with the socioeconomic divisions that Bourdieu worked with: bourgeoisie, petite bourgeoisie, and proletariat. Each taste public shares values and common ways of wielding cultural resources, rather than economics or education.

Other scholarship avoids societal characterizations of taste altogether, focusing instead on aesthetic resemblance and affinity. Solomon and Assael (1987) showed that there is a tendency to consume cultural goods in bundles around lifestyles and brand images. They termed these patterns "consumption constellations." Likewise

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

McCracken (1988) argued that underlying aesthetic forces guide consumption. Consumers unconsciously systematize their tastes according to the *Diderot Effect*—"a force that encourages the individual to maintain a cultural consistency in his/her complement of consumer goods" (p. 123). Aesthetics-based accounts of taste compete with socioeconomic explanations. Thus, the second of a pair of hypotheses about influence is proposed:

> **H3b:** Variation in the taste norms of various demographic groups on MySpace can be accounted for by the aesthetics associated with each group.

Semiotics of SNP Taste Speech

Interest tokens referring to books, music, movies, television, and general interests constitute a cultural vocabulary for the language of taste. When observed at a high level of abstraction, a SNP's lists of interests imply a taste statement. According to cultural semiotics (Barthes, 1964/1977), the impression fostered by a taste statement is explainable in terms of 1) the *paradigms* (e.g., socioeconomic or aesthetic themes) governing the semantic interpretation of interest tokens, and 2) the *syntagmatic rules* that were invoked to combine interest tokens into lists.

Paradigms

Interest tokens can be grouped into families of shared connotation, called *motifs* (Todorov, 1981). Two or more mutually exclusive motifs can be grouped together to form a paradigm. Some paradigms are denotative and straightforward. For example, consider that the following two motifs oppose each other and form a paradigm: "black and white movies" (comprised of, e.g., "Citizen Kane," "Casablanca") and "color movies" (comprised of, e.g., "Goodfellas," "The Godfather"). Other paradigms are quite subjective because they involve value judgments. For example, in the paradigm "movies with cult status" versus "movies without cult status," it is far less clear where tokens should be placed in that opposition.

Since a taste community such as MySpace articulates identity according to certain values and concepts that unite or divide its membership, these identity factors should be captured in the community's paradigms. Conversely, if the paradigms of MySpace could be uncovered by analysis of its interest tokens, then such an analysis should directly reflect MySpace's core values and concepts. For example, were a paradigm to capture the opposition "high educational capital" versus "low educational capital," it could be inferred that education level shapes the tastes of MySpace participants.

Syntagmatic Rules

Whereas paradigms define the interpretive semantics of interest tokens, syntagms are actual combinations of interest tokens that form a SNP. Barthes, for example, identified syntagms in the realms of food and fashion—e.g., for food, "Real sequence of dishes chosen during a meal: This is the menu," or for garments, "Juxtaposition in the same type of dress of different elements: skirt-blouse-jacket" (1964/1977, p. 63).

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Syntagms appropriate tokens to form high-level expressions, such as the menu and the outfit in Barthes' examples. In the domain of SNPs, the high-level expression of a syntagm of interest tokens is a taste statement. The particular impression fostered by a taste statement (e.g., "I have status") is owed to the syntagmatic rules that were employed in the combinative process.

The literature describes two syntagmatic rules that are important to taste statements. First, the rule of *group identification* is important to understanding the expression of prestige. After all, prestige is always performed so as to appeal to some group of people. When enacted, the rule prescribes that interest tokens be selected and combined so as to exhibit solidarity with the taste norms of some social group—such as a subculture (e.g., punk, hip-hop), or perhaps the "popular culture" of MySpace. By focusing on the interest tokens favored by a group, one demonstrates knowledge of the group's "inside secrets…whose possession marks an individual as being a member of a group and helps the group feel separate and different from those individuals not 'in the know'" (Goffman, 1959, p. 142). On the other hand, it may be desirable to avoid identification with a group, such as one's circle of friends, in order to express differentiation.

Another syntagmatic rule important for the successful performance of prestige is "expressive coherence." Goffman (1959) described expressive coherence as the recognition of the "fragility" of impressions, as the ability "to 'fill in'…any part that [the performer] is likely given" (p. 73), and as the avoidance of "destructive information" (p. 140). Destructive information consists of "facts which, if attention is drawn to them during the performance, would discredit, disrupt, or make useless the impression that the performance fosters" (p. 141). The rule of expressive coherence should be instrumental to SNP users expressing prestige. Any outlier interest tokens in their profiles—such as the inadvertent mention of something tabooed or distasteful—could constitute destructive information and spoil the impressions the users are trying to foster.

Profiles may exercise various degrees of coherence, such as incoherent, moderately coherent, and very coherent. Sometimes, including a mistake or flaw can enhance a perfectly coherent expression. These "disingenuous mistakes" (Davis, 1992) manipulate the sense of authenticity associated with Goffmanian (1959) "expressions given off." With respect to the clothing code, the Italian designer Nino Cerruti was quoted as dispensing the following advice: "For a man to be elegant, he must dress simply with some mistakes…There is nothing less elegant than to be too elegant" (Hawkins, 1978, p. L8). With respect to a SNP, one can imagine that a perfectly coherent declaration of interests, crafted to express prestige, can only be made more perfect by scattering a few deliberately errant (but not destructive) tokens throughout in order to mask the pretension of the act.

Methods

In order to test the above-proposed research hypotheses about influences on taste, paradigms, and taste statements, an empirical study was conducted of 127,477

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

MySpace profiles. First, a pilot study examined a small sample of profiles to discover relevant types of taste statements. Then, for the main study, natural language processing (NLP) techniques were applied to extract and normalize interest tokens from the full set of SNPs. Statistical measures and Principal Components Analysis were applied to interpret influences on taste, paradigms, and taste statements in the MySpace data.

Data Collection

A Python Web crawling script downloaded the HTML of 198,692 SNPs from MySpace.com from November 2006 through January 2007. This was done without first logging in, in order not to compromise the privacy of MySpace users whose profiles were intended to be viewed only by logged-in users. To create an unbiased sample, 50,000 profiles were randomly selected for download in the first stage. The second stage extracted the IDs of users listed in the "Top 8" friends section from each of the 50,000 downloaded profiles and queued these friends' profiles for download in random order. Some users listed more than eight friend connections in the "Top 8" section; in these cases, only the first eight friends were queued for download. Download was terminated at around 200,000 profiles, as that sample size balanced statistical significance with concerns about computation power required for further analysis.

Of these 198,692 profiles, 30,172 (15%) were identified as band profiles, because they contained a "Member Since" field that was unique to bands. These were discarded. Of 168,520 profiles remaining, a further 41,043 (24%) had to be discarded, because they were either marked as private or were profiles that had been deleted from MySpace. After this, 127,477 user profiles with meaningful data remained to constitute the corpus for this study.

Identification of Taste Statement Types

A pilot study preceded the main study in order to first catalog the types of taste statements that are relevant to MySpace SNPs. In Goffman's performance framework, "the role of expression is conveying impressions of self" (Goffman, 1959, p. 248). A *taste statement*, then, is the highest-level expression produced by a profile's lists of interests. At this highest level, the expression does not pertain to specific interests, but rather, it is a statement about qualities of self. Thus, a qualitative analysis is warranted.

A small sample of 100 profiles (and the "Top 8" friends' profiles associated with each profile) was selected randomly from the full SNP corpus. A subset of each SNP's text was considered in this qualitative analysis—its lists of interests and the "about me" passage. A grounded theory approach (Glaser & Strauss, 1967) was adopted to allow relevant types of taste statements to emerge from perusal of the data. Two passes were made through the data to accomplish the grounded theory tasks of memoing and sorting. The first pass tried to freely articulate the "taste statement" that each profile seemed to express and also noted particular mechanics of taste

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

speech (e.g., paradigms and syntagmatic rules) that seemed instrumental to the expression. The notes produced in the first pass were organized, resulting in the identification of several types of taste statements. In a second pass over the data, profiles were categorized according to the already identified taste statement types. For each profile, both the single best-fitting statement type and all applicable statement types were identified and recorded.

Profile Preprocessing and Normalization

The 127,477 collected profiles required preprocessing and normalization before statistical analysis could be undertaken. First, demographic responses and interest tokens had to be extracted from each profile. For all demographic fields except "occupation," responses came from a fixed list of options, so those fields did not require further processing. The contents of each interest category were tokenized and normalized. Responses to the "occupation" field were just normalized.

To tokenize the contents of each interest category, it was assumed that those contents followed a punctuation-delimited convention. Commas, semi-colons, heart symbols, bullet symbols, new lines, parentheses, and lists were common delimiters. Ambiguous delimiters such as "by" and "and" were handled by pattern-matching regular expressions. For example, "The Stranger by Albert Camus" is delimited around "by." To prevent invalid tokenization (e.g., "by" is not a delimiter in the television token "Step by Step"), only proposed tokens that matched positively against a list of known interest tokens were accepted (e.g., "Step" is not a known television token).

To normalize tokens, orthographic variants such as "csi"/"c.s.i." and "simpsons"/"the simpsons" were reconciled. In such cases, the more statistically frequent orthography was made canonical. To reconcile name variants ("david sedaris"/"sedaris") and abbreviations, a list of suspected equivalent terms was generated, and reconciliations were manually approved or rejected. Efforts to normalize the data had a visible effect on the final statistics. For example, as a result of normalization, "the simpsons" moved up to second place from third place in the television rankings.

Principal Components Analysis

Principal Components Analysis (PCA) is a statistical technique commonly used to identify the major dimensions of variation (i.e., the principal components) in a body of data. As mentioned above, Paolillo and Wright (2005) recently used PCA to identify the principal components (PCs) emergent from interest token data collected from user profiles in LiveJournal blogs. Bourdieu (1984) also used statistics akin to PCA to make sense of data from a lifestyle survey of French residents.

By assigning one PC to the x-axis and another PC to the y-axis, a map can be generated to help visualize the similarity/dissimilarity relationships between the plotted points. Both individual interest tokens and demographic groups (e.g., "occupation= student") may be plotted. Generally speaking, the variation captured by each PC should be interpretable. For example, Paolillo and Wright (2005) interpreted principal

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

component three in their study as bifurcating the space of interest tokens into those pertaining to "aesthetic sensibility" and those pertaining to "general sociability." Bourdieu (1984) interpreted the PCs emergent from the French lifestyle survey as capturing degrees of economic, cultural, educational, and artistic capital.

PCA is computed by performing the Singular Value Decomposition (SVD), a linear algebra transformation, on a covariance matrix. In this study, PCA is performed over two kinds of data points: demographic groups and interest tokens.

To generate a covariance matrix for demographic groups, a demographic category (e.g., "education level") was first chosen. The corpus of 127,477 profiles was sorted into the various demographic groups existing under that category. For example, "high school," "in college," and "post grad" are groups under the education category. Each profile was represented as a vector (salience-weighted list) of interest tokens. The average tastes, or "taste norms," of each demographic group were computed by averaging all the profile vectors contained in that group, producing a single "centroid" vector. Finally, to compute the covariance matrix, the similarity between every pair of centroid vectors was computed, using the cosine similarity calculation (Salton, Wong, & Yang, 1975).

Likewise, to generate a covariance matrix for interest tokens, a category of interest (e.g., "music") was first chosen, and the 20 most frequent interest tokens in that category were identified. For each top token, all profiles containing that token were added to that token's group. A centroid vector was computed for each group to produce the "taste norm" of an interest token. Covariance was computed between every pair of centroid vectors.

Measure of Expressive Coherence

The expressive coherence of a profile's lists of interest tokens is measured as the average of the similarity scores between every pair of tokens in that profile. This measure captures sensitivity to destructive information that is at the heart of the expressive coherence idea (Goffman, 1959), because any single out-of-place token in a profile is potentially dissimilar from all the other tokens.

Before the expressive coherence of particular profiles could be computed, similarity scores between every possible pair of interest tokens had to be computed. This computation utilized all 127,477 profiles for training purposes. Pointwise mutual information (PMI), a co-occurrence statistic, was the measure of similarity used. The PMI (Church & Hanks, 1990) of two interest tokens is proportional to a ratio of probabilities: the probability that a randomly selected profile will contain both tokens, divided by the product of the probabilities that a profile will contain just the first token or just the second token. Intuitively, PMI captures the degree to which two tokens are contextually co-dependent.

Measures of Group Identification

Group identification is the degree of solidarity that exists between a user's tastes and taste norms associated with some social group. Three kinds of social groups are

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

relevant to the present analysis: MySpace's "popular culture," MySpace's subcultures, and a user's "Top 8" friends.

Measuring how much a user identifies with their "Top 8" friends involves three vectors—(a) the vector of the user's interests; (b) the centroid vector averaging the user's friends' interests; and (c) as a baseline, the centroid vector averaging all 127,477 MySpace profiles. The cosine similarity of (a) and (b) produces a score (a*b), and the similarity between (a) and (c) produces (a*c). The ratio of (a*b) over (a*c) will be greater than 1.0 if the user tends to identify with their friends' tastes and will be less than 1.0 if the user tends to dissent from their friends' tastes.

To measure whether a user identifies more with MySpace's "popular culture" or with one of MySpace's subcultures, a rarity heuristic is used, in order to avoid having to manually identify subcultures and profiles belonging to each. A list was made of all unique interest tokens found in the corpus. This list was rank ordered from most frequently to least frequently occurring. A simplifying assumption is made that frequent tokens generally correspond to the taste norms of popular culture, while less frequent and infrequent tokens generally correspond to the taste norms of subcultures. The rarity of a given user's interests is calculated by averaging the popularity ranks of all the tokens in their profile. A profile with a very low rarity score is interpreted as identifying with popular culture. A profile with a very high rarity score is interpreted as identifying with subcultures.

Findings: Pilot Study

A grounded theory (Glaser & Strauss, 1967) qualitative analysis of 100 MySpace profiles yielded four relevant types of taste statements, summarized in Figure 2. Of the 100 profiles in the sample, 21 were either blank or too short to afford analysis of their taste statements. A first pass over the profiles focused on noting the impressions fostered by each profile (Figure 2, rightmost column), and noting particular semiotic details that seemed instrumental to the impression (Figure 2, middle column). These notes were then reviewed and four types of taste statements were identified—those conveying prestige, differentiation, authenticity, and theatrical personas. In a second pass over the profiles, the number of times each taste statement type best fit a profile was recorded ($\text{frequency}_{\text{primary}}$) and the number of times each taste statement type was at all applicable to a profile was recorded ($\text{frequency}_{\text{secondary}}$). These frequencies are shown in the leftmost column and should be considered as fractions of 79 profiles, given that 21 profiles were discarded.

Prestige taste statements were most frequently observed as both primary and secondary statements. This accords with the view that MySpace is the online home of "club culture" (Thornton, 1996), where users are always "dressed to impress." The qualitative coherence of these profiles was judged according to this rubric: "very coherent," "moderately coherent," and "incoherent." Profiles that stated prestige were often very coherent (e.g., A1, A2), but never less than moderately coherent (e.g., C2). The vast majority of prestige taste statements either identified with the

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Taste statement type	Mechanics		Profile excerpts *(and observations)*
Convey prestige *frequency*$_{primary}$: 47 *frequency*$_{secondary}$: 62 i.e.: "I am pop" (popular prestige) "I am alt" (subcultural prestige)	popular identification, subcultural identification, very / moderately coherent	A1	"[music] the clash. the cure. the jam. the kinks. the ramones. the smiths. the stooges. x..." *(tokens highly coherent, 'the' bands all w/ pop centrality; lowercase and alphabetized presentation)*
		A2	"[general] PSYCHOSOMATIC / COMPLEXITY / DREAMS... [music] PERE UBU / EINSTURZENDE NEUBAUTEN / NEGATIVLAND / R HELL & THE VOIDOIDS..." *(rare tokens with high mutual coherence suggests niche centrality; uniform typographic style)*
Convey differentiation *frequency*$_{primary}$: 18 *frequency*$_{secondary}$: 35 i.e.: "I'm unique from friends" "I defy characterization"	dissent from friends' tastes, incoherent	B1	"[music] Bowie & [Notorious] B.I.G. - Radiohead & Rakim - Tupac & Tori [Amos?]...[books] Dostovesky & Dan Brown" *(user presents interests in alliterated pairs; each pair is aesthetically dissimilar and the overall profile has low coherence)*
Convey authenticity *frequency*$_{primary}$: 6 *frequency*$_{secondary}$: 51 i.e.: "I am real / unpretentious" "I am easygoing / relaxed"	breaks from form, atypical details (like song titles), moderately coherent to incoherent, disingenuous mistakes	C1	"[about me] Let's see, what is there to really say about myself. I am a nice person, and I do the best I can with what I am given. [music] Closer, by Nine Inch Nails [books] Just finished the Dan Brown series, loved all of them [television] Not too much... Dr. Phil, 7th Heaven, the amazing race series" *(ingenuous, sincere; interest listings are reticent, undercoded, not very coherent; use of relaxed tokens like 'anything', 'bit of everything')*
		C2	"[music] wilco, feist, arcade fire, bright eyes, beck, radiohead, and also-- old school hip-hop like tupac, top 40s, anything good" *(many mutually coherent indie artists, followed by coherence-breaking broadening; Tupac Shakur, not usually considered 'old school,' renders passage suspect--possible disingenuous mistake)*
Convey theatrical persona *frequency*$_{primary}$: 8 *frequency*$_{secondary}$: 37 i.e.: "I am sexy/sarcastic" "I am a flirt"	"yes/no," emoticons, "lol," atypical orthography, emotional paradigms	D1	"[general] PARTY!!!, beer pong, bars, poker, playing pool, peeing in alleys, TAILGATING!!! [movies] Vince Vaughn, Owen Wilson, Jack Black, rounders, Big Lebowski...[television] espn, american muscle car" *(tokens, some iconic, project a "Frat Boy" machismo; hyperbolic tokens like 'peeing in alleys'; use of all-caps and exclaimations in affected manner)*

Figure 2 Types of taste statements observed as being relevant to MySpace profiles

"popular culture" (e.g., A1) or with a subculture (e.g., A2). Profiles conveying popular prestige seemed to invoke many of the highest frequency interest tokens on MySpace (judged against a site-wide list of top interests), while profiles conveying subcultural prestige invoked more obscure interest tokens. Profile A2, for example, evoked the experimental music subculture.

Differentiation taste statements were also frequently observed. The owner of profile B1 seemed interested in expressing how utterly unique and diverse he or she was. That intention was made clear by his or her use of alliterative pairs of aesthetically diametrical interests. Another kind of differentiation observed was differentiating oneself from one's friends. To judge this, the "Top 8" friends' profiles were briefly examined in the second pass to get a sense of their tastes. Profiles were marked as expressing differentiation from friends when their lists of interests seemed to embody tastes that were substantially different from the tastes of all their "Top 8" friends. No attempt was made to account for the logic of this difference.

Two further statement types—those conveying authenticity and those conveying some theatrical persona—were more frequently considered to be secondary qualities of SNP taste performances, rather than the primary goals of performance, in the qualitative analysis.

Authenticity is important in the eyes of some subcultures, including rap culture (Simpson, 1996) and club culture (Thornton, 1996), so it was not surprising to find

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

that many profiles conveyed this quality. In the semiotics of fashion, authenticity is associated with a relaxed style and the display of slight imperfection (Davis, 1992). Profiles perceived as having an authentic quality were ones that met this criterion. They projected a relaxed feeling, their lists of interests were not overly verbose or coherent, and they often broke from form and convention. For example, C1 breaks form by not adhering to the punctuation-delimited convention for listing interests and also mentions a song name, a rather atypical detail. Not all profiles marked as stating authenticity were necessarily convincing, though. Some profiles tried hard to seem relaxed. Profile C2, for example, began with a coherent list of indie music interests, only to break incoherently toward outlier interests. This, and the fact that rapper Tupac Shakur was identified as an "old school" rapper when by most accounts he was not, suggested that these latter tokens were disingenuous mistakes (Davis, 1992), added to improve the authenticity of the expression.

Finally, some profiles seemed intent on creating and inhabiting a caricature or theatrical persona. For example, two profiles exuded a manic depressive persona, and 14 profiles exuded a sexy persona. To achieve theatricality, tokens such as "lol" and tongue-in-cheek responses of "yes" and "no" to a whole interest category were frequently used. One profile, D1, projected a "frat boy" machismo.

Findings: Main Study

This section reports findings from the empirical analysis of all 127,477 MySpace profiles. First presented are descriptive data about MySpace user demographics and site-wide top interests. Second, the influence of socioeconomic and aesthetic factors on users' tastes is examined. Third, motifs and paradigms that constitute the taste semantics of the MySpace community are interpreted from the results of a Principal Components Analysis of interest token data. Fourth, statistical measures for expressive coherence and group identification are applied to the profile data to infer evidence for taste statements.

Site-Wide Demographics

Users' responses to five demographic fields were extracted from the 127,477 profiles in the MySpace corpus and are compiled in Figure 3. The number of profiles responding to these fields ranged from 57,782 for "religion" to 127,143 for "relationship status." Assuming 95% confidence, the margin of error (i.e., 0.98/sqrt(n)) associated with these response sizes ranged from $\pm 0.4\%$ to $\pm 0.3\%$, respectively. However, additional distorting factors should be considered. "Relationship status" is subject to a forced reporting bias, as it was a mandatory field that only 0.3% of users avoided responding to (by using software hacks). The response "swinger" seemed to be over-represented, possibly due to its comedic value or as a ploy to create plausible deniability about actual status. In addition, "occupation" seemed vulnerable to selective reporting bias, as suspiciously all of the frequently reported occupations had allures attached to them. It is possible that MySpace users preferred

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Relationship Status	[n=127143 (99.7%)]	
single	67730	(53.3%)
in a relationship	36142	(28.4%)
married	15556	(12.2%)
swinger	5325	(4.2%)
divorced	2390	(1.9%)

Religion	[n=57782 (45.3%)]	
christian - other	19654	(34.0%)
catholic	12351	(21.4%)
other	7512	(13.0%)
agnostic	5571	(9.6%)
atheist	4034	(7.0%)
jewish	1657	(2.9%)
buddhist	1559	(2.7%)
protestant	1492	(2.6%)
scientologist	1006	(1.7%)
wiccan	906	(1.6%)
taoist	604	(1.0%)
mormon	574	(1.0)%)
muslim	571	(1.0%)
hindu	291	(0.5%)

Occupation	[n=60243 (47.3%)]	
student	7682	(12.8%)
musician	1238	(2.1%)
artist	1021	(1.7%)
photographer	841	(1.4%)
actor	813	(1.3%)
writer	769	(1.3%)
bartender	641	(1.1%)
producer	632	(1.0%)
model	625	(1.0%)
teacher	605	(1.0%)
dj	601	(1.0%)
sales	559	(0.9%)
graphic designer	437	(0.7%)
actress	417	(0.7%)
designer	411	(0.7%)
dancer	351	(0.6%)
singer	343	(0.6%)
waitress	296	(0.5%)
marketing	267	(0.4%)
comedian	261	(0.4%)
(only top 20 responses shown)		

Views on Children	[n=94480 (74.1%)]	
someday	54849	(58.1%)
proud parent	16137	(17.1%)
i don't want kids	12147	(12.9%)
undecided	8508	(9.0%)
love kids, but not for me	2839	(3.0%)

Education Level	[n=88748 (69.6%)]	
college graduate	25160	(25.9%)
in college	22958	(23.6%)
high school	15270	(15.7%)
some college	14042	(14.4%)
grad / professional school	8513	(8.8%)
post grad	2805	(2.9%)

Figure 3 Site-wide demographics of MySpace, based on a 127,477-profile sample collected from November 2006 through January 2007

to report their moonlighting jobs (e.g., bartender, DJ, dancer) and aspired-to jobs (e.g., actor, artist, singer), rather than their day jobs; this explanation would in fact be consistent with users engaged in taste performance.

Compared to overall U.S. demographics as reported by the U.S. Census Bureau (2007), the breakdowns for "religion" and "education level" corresponded best to the national average. To the extent that "relationship status" could be believed, married people were under-represented compared to the national average. Artistic and entertainment occupations (e.g., musician, artist, photographer, actor, writer) were extremely over-represented, perhaps owing to the aforementioned reasons. Age and gender demographics for MySpace were not gathered for the corpus; arguably, such information would not necessarily have provided an accurate portrayal, due to known distorting factors such as age deception (Donath, 1999).

Site-Wide Top Interests

The frequency distribution of cultural interests captured in the 127,477-profile MySpace corpus provides a portrait of the popular tastes of MySpace during the data collection period of November 2006 through January 2007. After interest tokens were extracted from profiles and normalized into canonical tokens, frequency statistics were compiled for each category of interest. Figure 4 shows the most frequently stated tokens for each of the six categories of interest on MySpace.

Across all six categories of interests, some 240,000 unique interest tokens were mentioned by at least two different users, and 43,000 interest tokens were mentioned by at least 10 different users. From a semiotics viewpoint—in which interest tokens constitute the vocabulary of the SNP taste language—the large size of MySpace's vocabulary of interests suggests that there is great expressive potential in this medium. Music interests accounted for the greatest proportion of this vocabulary, with over 70,000 unique music tokens mentioned by at least two users and 15,000 unique tokens mentioned by at least 10 users. This finding highlights the important

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Books	n=
the da vinci code	2371
the bible	2087
harry potter	1952
catcher in the rye	1750
1984	1354
angels and demons	1129
to kill a mockingbird	937
the alchemist	888
stephen king	881
memoirs of a geisha	796
the great gatsby	764
david sedaris	719
on the road	683
lord of the rings	662
brave new world	622
books	611
yes	579
hunter s. thompson	576
kurt vonnegut	576
anne rice	537
dan brown	532
chuck palahniuk	512
choke	484
fight club	479
a clockwork orange	476
lolita	468

Music	n=
radiohead	5812
the cure	4722
the beatles	4453
nirvana	4280
johnny cash	4273
pink floyd	4261
coldplay	4199
sublime	3898
bob marley	3656
tool	3639
the killers	3620
rock	3569
rap	3443
bjork	3433
jack johnson	3421
the rolling stones	3270
weezer	3211
green day	3191
david bowie	3179
beck	3099
led zeppelin	3063
hip hop	3051
modest mouse	3008
incubus	2955
u2	2912
the smiths	2907

Movies	n=
fight club	5242
pulp fiction	4427
donnie darko	3733
scarface	3553
old school	3488
anchorman	3183
office space	3177
garden state	3130
wedding crashers	3042
star wars	2989
amelie	2976
the notebook	2970
goodfellas	2704
napoleon dynamite	2505
lord of the rings	2446
boondock saints	2412
crash	2376
kill bill	2326
snatch	2320
the matrix	2183
american beauty	2116
requiem for a dream	2114
sin city	2069
lost in translation	2007
american history x	1946
the big lebowski	1934

Television shows	n=
family guy	10836
the simpsons	8167
lost	5439
csi	4115
south park	4013
grey's anatomy	3869
the office	3746
friends	3327
24	3245
seinfeld	3093
sex and the city	2891
house	2765
nip tuck	2691
entourage	2661
the oc	2575
arrested development	2441
futurama	2372
project runway	2304
desperate housewives	2301
scrubs	2221
sopranos	2069
curb your enthusiasm	2060
law and order	2058
six feet under	1986
will and grace	1906
the daily show	1708

Heroes	n=
my mom	5525
my dad	3522
my parents	2759
jesus	2095
god	1259
myself	1147
my family	1134
my friends	854
r.i.p.	769
batman	752
martin luther king jr	732
superman	708
my sister	452
my brother	437
lol	410
spiderman	395
my grandmother	379
my grandfather	373
john lennon	339
malcolm x	338
you	332
ghandi	314
yes	310
bob marley	306
of course	306
oprah	302

General interests	n=
music	15916
movies	6496
reading	6090
dancing	4832
art	4780
photography	4281
traveling	4181
writing	3667
cooking	3043
shopping	3005
friends	3002
sports	2565
snowboarding	2437
travel	2296
hiking	2249
camping	2216
singing	2130
football	2129
sleeping	1932
basketball	1894
books	1832
painting	1777
tattoos	1726
swimming	1700
family	1700
drawing	1693

Figure 4 Site-wide lists of most frequently stated interests, based on a 127,477-profile sampling of MySpace from November 2006 through January 2007

role that musical tastes play in the MySpace community. One can also interpret this more broadly as suggesting that musical interests are playing a more dominant role in the tastes and identities of young people today.

Socioeconomic and Aesthetic Influences

In order to explore the influence of socioeconomic and aesthetic factors on the tastes of MySpace participants as posed in Hypotheses 3a and 3b, PCA was performed over the taste norms associated with each demographic group. The first two principal components (PCs) produced by PCA—PC-1 and PC-2—are considered here, because they explain the greatest amount of variation in the data. PC-1 and PC-2 were used to plot demographic groups in relation to one another on what might be called a "taste map." Each of the four maps shown in Figure 5 visualizes taste similarity (two points close together) and taste dissimilarity (two points far away) relationships between various demographic groups.

Following precedent (Bourdieu, 1984; Paolillo & Wright, 2005), the nature of the variation captured by each PC was interpreted by reading the maps west to east (W-E) to describe PC-1 and north to south (N-S) to describe PC-2. The caveat that such "readings into" are more art than science should be considered.

In the map of top 10 occupations (Figure 5, upper left), a N-S reading contrasts textual/symbolic occupations (i.e., writer, producer, teacher, actor) with sensory-laden ones (artist, photographer, musician). Reading W-E, the possession of creative, artistic, and cultural capitals seem to descend. In the map of education level attained (lower left), implied age seems to ascend reading W-E. N-S was at first difficult to interpret because "high school" and "post grad" were so near each other. Further calculations, where the taste norm of each education grouping was compared with

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

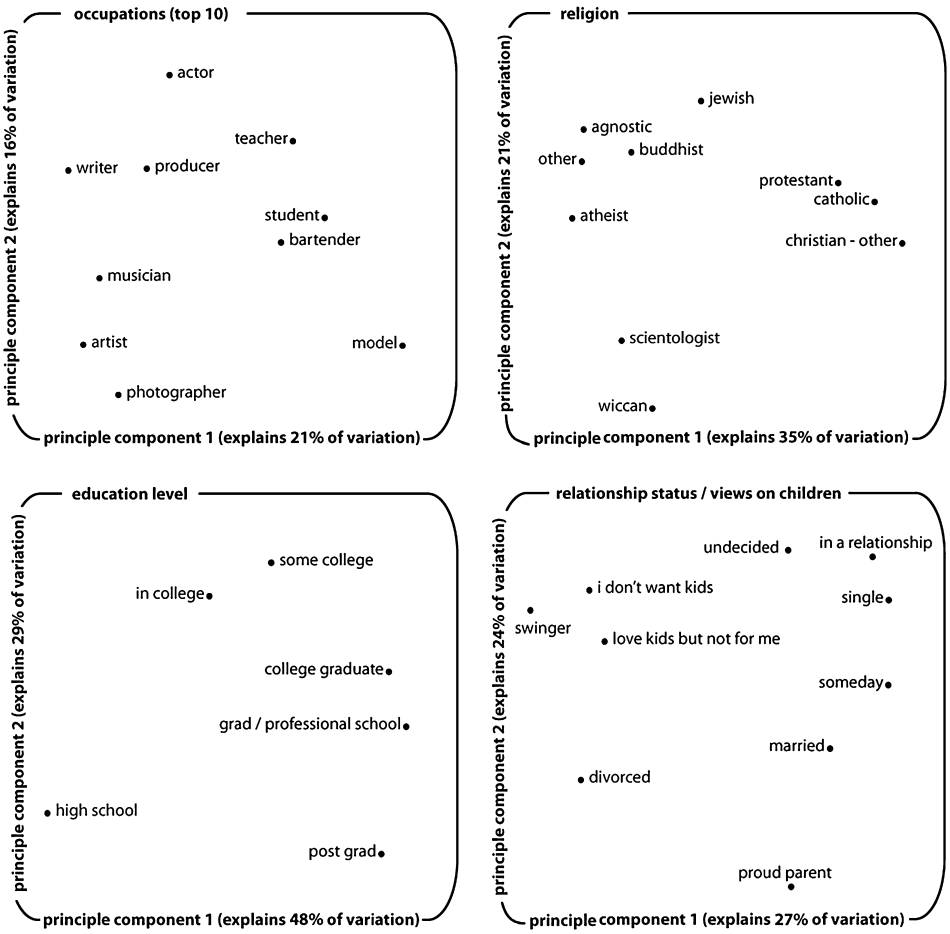

Figure 5 Maps generated by Principal Components Analysis, showing similarity/dissimilarity relationships among the taste norms of various demographic groupings

the taste norm of the "popular culture" on MySpace (as represented by the most frequent tokens), confirmed the interpretation of N-S as mainstream (N) versus non-mainstream (S). These calculations also showed that users in the "some college" and "in college" groupings were the principal demographic groups that constituted mainstream tastes on MySpace.

The map of relationship status and views on children (lower right) captured pessimism and optimism about family ties (W-E). The variation captured by N-S was too ambiguous to interpret. As for religion (upper right), W-E seemed to contrast the secularism of atheists and agnostics with the religiosity of Protestants, Catholics, and other Christians. This, and the fact that PC-1 accounts for a large 35% of all the variation in the tastes of these groups, suggests that one's spirituality and faith either greatly affect, or greatly correlate with, one's cultural tastes. When the

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

taste norm of each religion group was compared with the taste norm of the "popular culture," Jewish and agnostic users were found to be close to the mainstream, while Wiccans and Scientologists were non-mainstream. This insight is captured by a N-S reading of the religion map.

PC-2 of the occupation map (textual-versus-sensory) and PC-1 of the relationship/children map (pessimism-versus-optimism about family ties) provide evidence to support Hypothesis 3b, because they explain demographic variation in aesthetic terms. The expectation for socioeconomic-based variation in the data, as articulated in Hypothesis 3a, was not clearly met, but cannot be fully rejected. PC-1 of the occupation map seemed to capture some combination of creative, artistic, and cultural capital, which did belong to Bourdieu's (1984) notion of socioeconomic capitals; however, none of the PCs in Figure 5 clearly captured the most basic form of capital in the Bourdieuian framework—economic. In addition, a Bourdieuian would have expected that PC-1 or PC-2 of the education map should indicate educational capital (e.g., low capital for "high school," high capital for "post grad"), whereas this was not observed.

Emergent Motifs and Paradigms

Interest token occurrence data were analyzed using PCA. Whereas the taste norms of a demographic group were calculated by averaging the profiles of its members, the taste norm for a single interest token was calculated by averaging all profiles containing that token. The output of PCA generated the maps of Figure 6, which depict similarity relationships among interest tokens. Following Paolillo and Wright's (2005) success in interpreting the PCs of interest tokens data as emergent motifs and paradigms of the LiveJournal community, Hypothesis 1 proposed that the motifs and paradigms of MySpace could similarly be interpreted from the present data.

The PCs of the four maps shown in Figure 6 were interpreted and are summarized below:

- [Music] vague-to-specific (W-E) and cult-to-accessible (N-S)
- [Television] sexy-to-humorous (W-E) and episodic-to-saga structure of show plots (N-S)
- [Books] utopian-to-dystopian (W-E) and sincere-to-satirical (N-S)
- [Cross-category] ironic-to-straightforward (W-E) and heartwarming-to-alienating (N-S)

The relationships captured by each PC dimension were meaningful and interpretable and captured oppositions (e.g., utopian-to-dystopian). Thus, Hypothesis 1 is strongly supported. Each pole (i.e., N, S, E, W) captured a motif largely pertaining to aesthetics, especially to sentiment (e.g., utopic, alienating) and personality (e.g., sexy, sincere). Each PC captured a tradeoff between two opposing motifs, which form a paradigm. In turn, these paradigms seemingly reflect the aesthetic distinctions and debates that are most useful for articulating identities and tastes on MySpace.

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

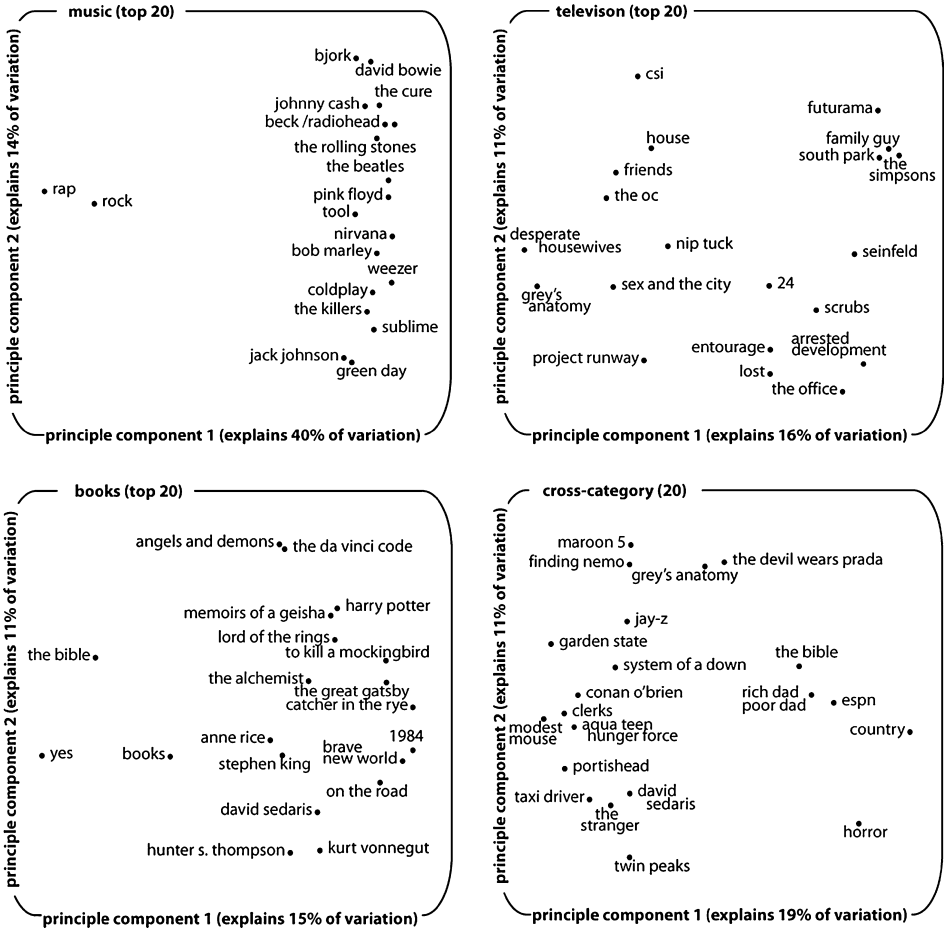

Figure 6 Maps generated by Principal Components Analysis, showing similarity/dissimilarity relationships among top interest tokens

Statistical Inference of Taste Statements

The pilot study provided initial indications of prestige and differentiation taste statements. To test Hypotheses 2a and 2b, statistical measures devised for two indicators of taste statements—expressive coherence and group identification—were applied to the MySpace data.

First, the relationship between a profile's group identification and its expressive coherence was measured (Figure 7). The average of all pairwise similarity scores between the tokens in a profile measured expressive coherence. The average rarity of the tokens in a profile was used as a measure of identification—a profile whose tokens are not rare (high-frequency) is said to identify with the popular tastes on MySpace, whereas a profile with rare tokens is said to identify with subcultural tastes. To boost precision in the coherence measure, a subset of

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

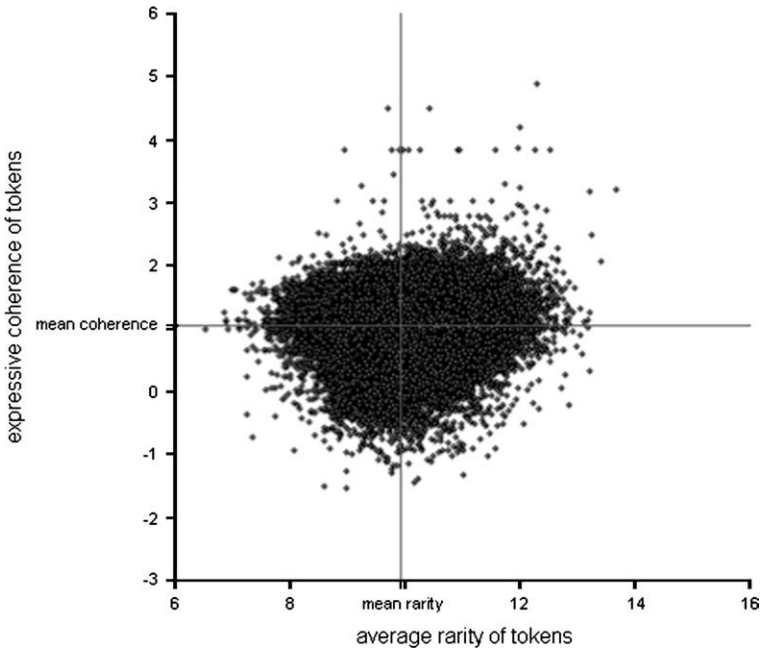

Figure 7 36,132 profiles, plotted by average rarity of their interest tokens (x-axis) versus coherence of tokens (y-axis)

36,132 profiles having six or more interest tokens each was selected from the corpus.

Figure 7 shows that there is a wide distribution for both coherence and average token rarity. The greatest range of coherence is observed near mean rarity, but incoherence begins to taper off as rarity decreases and increases. At the lower right of the mass, there is a slope up: As a profile becomes increasingly obscure, it is more and more unlikely to be incoherent. Since mindfulness about expressive coherence and self-censorship of destructive information signal that a performance is underway (Goffman, 1959), this slope suggests that users with obscure interests are engaged in performance. At the lower left of the mass, a similar slope up suggests that users with popular interests are also engaged in performance. Considering that taste performances are more consistently observed at the two extremes of the x-axis in profiles that identify with either popular culture or subcultures, these performances can be interpreted as seeking popular and subcultural prestige. This finding strongly supports Hypothesis 2a.

To examine the degree to which users identified with or dissented from their friends' tastes, the statistical similarity between a user's interests and the aggregated interests of the user's "Top 8" friends was measured. This was then compared to a baseline—the similarity between the user's profile and the overall tastes of MySpace. To improve precision in these measurements, a subset of 24,979 profiles

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

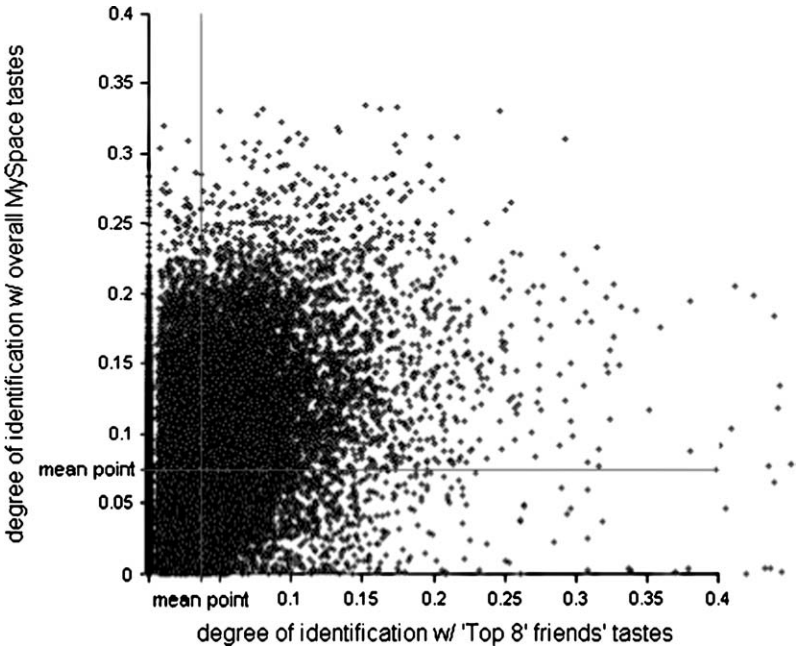

Figure 8 24,979 profiles, plotted by degree of identification with "Top 8" friends' tastes (x-axis) versus degree of identification with overall MySpace tastes (y-axis)

was identified in the corpus, such that each profile contained six or more interests, and all of their "Top 8" friends' profiles were also in the corpus. Figure 8 plots the profiles along these two measures. The center of mass is located at (0.036, 0.074)—this indicates with statistical significance that on average, MySpace users tended to differentiate themselves from their friends, rather than identifying with their friends' tastes. This finding supports Hypothesis 2b, although not conclusively.

Discussion

The above findings from empirical analyses of MySpace profile data clearly indicate that SNPs and their lists of cultural interests are being fashioned into taste statements and utilized for taste performances. In addition, the findings confirmed that much of the variation across the taste norms of various demographic groupings was owed to aesthetic factors.

The pilot study suggested that prestige and differentiation were the two primary types of taste statements expressed by MySpace SNPs, with authenticity and theatrical persona as secondary qualities. Further evidence that users were asserting prestige and differentiation was inferred from statistical measures of profiles' expressive coherence and group identification. Two groups of users—those whose profiles aligned with the most popular tastes and those whose profiles aligned with obscure and subcultural tastes—happened to demonstrate the most consistent

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

level of expressive coherence, a Goffmanian (1959) indicator of taste performance. The collocation of the performance stance and identification with popular/subcultural tastes most certainly suggests that prestige was the goal of these performances.

Another graph comparing user identification with "Top 8" friends versus user identification with overall MySpace tastes showed with statistical significance that users' interests tended to be more dissimilar from their friends' tastes than would be expected to happen by chance. Several explanations are possible for this observation: 1) users tended to friend those who complemented rather than overshadowed their unique tastes; 2) users tended to only select friends who did not overshadow them for the "Top 8;" or 3) users maintained an awareness of their friends' profiles and crafted their own profiles so as to be unique. The third explanation most directly supports the finding of differentiation taste statements, while all three explanations are consistent with Simmel's (1908/1871b) observation that differentiation is a basic social drive, of which taste performance is only a small part.

Principal Components Analysis was applied fruitfully to gauge socioeconomic and aesthetic influences on the tastes of MySpace. Aesthetic paradigms such as pessimistic versus optimistic views about family ties and textual versus sensory-laden occupations offered evidence that taste differences across demographic groups were aesthetic in nature, according with the Diderot Effect (McCracken, 1988). Evidence for socioeconomic factors was inconclusive. Although economic capital did not explain variation for any of the demographics examined, other forms of capital were implicated, including cultural capital, which was emphasized by Gans (1999). The maps of education and religion also captured a paradigm of mainstream tastes versus non-mainstream tastes, which could also be interpreted as possession of social capital or subcultural capital (Thornton, 1996).

Finally, PCA was applied to interest token occurrence data, revealing the many motifs and paradigms that frame MySpace users' conceptions of taste and identity. The paradigms that emerged had much to do with personality and sentiment. Different kinds of paradigms arose within different interest categories, suggesting that each interest category has its own expressive affordances. For example, books were useful for expressing utopian and dystopian worldviews. Television interests, the data suggested, afforded the expression of sexiness or humorousness, but the data also implied that on MySpace, it is unusual to be thought of as both sexy and humorous, as these were paradigmatic. As music interests were the most diverse category, having by far the greatest number of unique tokens of any category, it was unexpected that the 20 most frequent music interests were not at all diverse. According to PC-1 in the map of music interests in Figure 6, a large 40% of the variation among the top 20 tokens could be explained simply by identifying "rock" and "rap" as odd ones out. Since the other tokens do contain some rock bands, one explanation is that "rock" and "rap" are ostracized because they are vague labels. If this interpretation is correct, then MySpace, a site where music factors greatly into taste expression, values specificity of music interest above all else.

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Conclusions

Whereas previous studies (boyd, 2006; Donath & boyd, 2004) implicated social network site users' friend connections and friending behavior in identity portraiture and performance, this study explored the expressive potential of lists of cultural interests contained in social network profiles and presented evidence to suggest that these lists of interests can function as taste performances. Considering that on MySpace, interest tokens were found to be markers of rich motifs like irony, alienation, utopia, and satire, the social network profile's lists of interests might actually be more useful as an indicator of one's aesthetics than as a factual declaration of interests.

Many of the methods that were employed in this study—semiotics, natural language processing, Principal Components Analysis, and statistical similarity measures—are still novel within the context of the subject matter of taste and performance. Their use here has helped to foreground some of their benefits and limitations. These methods scale well to enormous data sets, they can illuminate latent dimensions of data (such as motifs and paradigms in MySpace) that might not have been considered otherwise, and they could potentially be deployed to track aesthetic and social trends in an online community in real time. The limitations of these large-scale computational and statistical methods include the loss of some transparency—not always being able to understand how a generalization was reached or how it can be mapped back onto specific examples; and the loss of some precision—not being able to model all the technical factors and data interactions that explain a conclusion.

Whereas this study analyzed the patterns of the MySpace community at roughly one point in time, the next wave of taste insights might only be achieved by a recurrent, comparative critique across multiple communities and cultures, including in languages other than English. For example, do the popular tastes of Friendster lag those of MySpace? The aesthetics of MySpace are ironic rather than straightforward, dystopian rather than utopian, sexy and humorous, sincere and satirical. If and when the demographics of social network sites equal the demographics of society in general, insights such as these will take on new social importance.

References

Barthes, R. (1977). *Elements of Semiology* (A. Travers & C. Smith, Trans.). New York: Hill and Wang. (Original work published 1964.)

boyd, d. (2006). Friends, Friendsters, and MySpace Top 8: Writing community into being on social network sites. *First Monday,* **11**(12). Retrieved August 20, 2007 from http://www.firstmonday.org/issues/issue11_12/boyd/

Bourdieu, P. (1984). *Distinction: A Social Critique of the Judgment of Taste* (R. Nice, Trans.). Cambridge: Harvard University Press. (Original work published 1979.)

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, **16**(1), 22–29.

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

comScore. (2007, July 31). *Social networking goes global*. Retrieved July 31, 2007 from http://comscore.com/press/release.asp?press=1555

Csikszentmihalyi, M., & Rochberg-Halton, E. (1981). *The Meaning of Things: Domestic Symbols and the Self*. Cambridge: Cambridge University Press.

Davis, F. (1992). *Fashion, Culture, and Identity*. Chicago: University of Chicago Press.

Donath, J. (1999). Identity and deception in the virtual community. In P. Kollock & A. S. Smith (Eds.), *Communities in Cyberspace* (pp. 29–59). London: Routledge.

Donath, J., & boyd, d. (2004). Public displays of connection. *BT Technology Journal*, **22**(4), 71–82.

Fiore, A., & Donath, J. (2005). Homophily in online dating: When do you like someone like yourself? *Proceedings of the Conference on Human Factors in Computing Systems* (pp. 1371–1374). New York: ACM Press.

Gans, H. (1999). *Popular Culture and High Culture: An Analysis and Evaluation of Taste*. New York: Basic.

Glaser, B. G., & Strauss, A. (1967). *Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine.

Goffman, E. (1959). *The Presentation of Self in Everyday Life*. Garden City, NY: Doubleday.

Grodin, D., & Lindlof, T. (1996). The self and mediated communication. In D. Grodin & T. Lindlof (Eds.), *Constructing the Self in a Mediated World* (pp. 3–12). Thousand Oaks, CA: Sage.

Hawkins, T. (1978, December 8). For men. *Los Angeles Times*, L8.

Lenhart, A., & Madden, M. (2007, January 3). Social networking websites and teens: An overview. *Pew Internet Project Data Memo*. Retrieved July 30, 2007 from http://pewinternet.org/pdfs/PIP_SNS_Data_Memo_Jan_2007.pdf

Liu, H., & Maes, P. (2005). InterestMap: Harvesting social network profiles for recommendations. *Proceedings of the IUI'2005 Beyond Personalization Workshop* (pp. 54–59). Retrieved August 1, 2007 from http://www.grouplens.org/beyond2005/papers.html/

Liu, H., Maes, P., & Davenport, G. (2006). Unraveling the taste fabric of social networks. *International Journal on Semantic Web and Information Systems*, **2**(1), 42–71.

McCracken, G. (1988). *Culture & Consumption*. Bloomington: Indiana University Press.

McCracken, G. (2006). *Plenitude*. Bloomington: Indiana University Press.

Nielsen-NetRatings. (2006, May 11). *Social networking sites grow 47 percent, year over year, reaching 45 percent of web users*. Retrieved July 30, 2007 from http://www.nielsen-netratings.com/pr/pr_060511.pdf

Paolillo, J. C., & Wright, E. (2005). Social network analysis on the Semantic Web. In V. Geroimenko & C. Chen (Eds.), *Visualizing the Semantic Web* (2nd ed., pp. 229–242). Berlin: Springer.

Perkel, D. (in press). Copy and paste literacy? Literacy practices in the production of a MySpace profile. In K. Drotner, H. S. Jensen, & K. Schroeder (Eds.), *Informal Learning and Digital Media: Constructions, Contexts, Consequences*. Newcastle, UK: Cambridge Scholars Press.

Salton, G., Wong, A., & Yang, C. S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, **18**(11), 613–620.

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Simmel, G. (1971a). How is society possible? In D. N. Levine (Ed.), *On Individuality and Social Forms* (pp. 6–22). Chicago: University of Chicago Press. (Original work published 1908.)

Simmel, G. (1971b). Group expansion and the development of individuality. In D. N. Levine (Ed.), *On Individuality and Social Forms* (pp. 251–293). Chicago: University of Chicago Press. (Original work published 1908.)

Simpson, T. A. (1996). Construction of self and other in the experience of rap music. In D. Grodin & T. Lindlof (Eds.), *Constructing the Self in a Mediated World* (pp. 107–123). Thousand Oaks, CA: Sage.

Solomon, M. R., & Assael, H. (1987). The forest or the trees—a gestalt approach to symbolic consumption. In J. Umiker-Sebeok (Ed.), *Marketing and Semiotics* (pp. 189–218). Berlin: Mouton de Gruyter.

Sundén, J. (2003). *Material Virtualities: Approaching Online Textual Embodiment*. New York: Peter Lang Publishing.

Thornton, S. (1996). *Club Cultures: Music, Media, and Subcultural Capital*. Middletown, CT: Wesleyan University Press.

Todorov, T. (1981). *Introduction to Poetics*. Minneapolis: University of Minnesota Press.

United States Census Bureau. (2007). *The 2007 statistical abstract*. Retrieved February 21, 2007 from http://www.census.gov/compendia/statab/

Veblen, T. (1899). *The Theory of the Leisure Class*. New York: Dover Publications.

About the Author

Hugo Liu (hugo@media.mit.edu) is a research affiliate of the Media Laboratory and the Program in Comparative Media Studies at MIT, where he has taught courses on artificial intelligence and philosophy of aesthetics. His taste research explores the new empowerments and insights into food, fashion, lifestyle, and consumer culture afforded by statistical and computational modeling.

Address: MIT Media Laboratory, 20 Ames Street, Cambridge, MA, 02139 USA

10836101, 2007, 1, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2007.00395.x, Wiley Online Library on [01/01/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License