Stereotypes: Static Abstractions or Dynamic Knowledge Structures?

Article *in* Journal of Personality and Social Psychology · December 2006

DOI: 10.1037/0022-3514.91.5.814 · Source: PubMed

3 authors:

Leonel Garcia-Marques
University of Lisbon
122 PUBLICATIONS **749** CITATIONS
SEE PROFILE

Ana Sofia Santos
University of Lisbon
60 PUBLICATIONS **74** CITATIONS
SEE PROFILE

Diane Mackie
University of California, Santa Barbara
130 PUBLICATIONS **7,152** CITATIONS
SEE PROFILE

Journal of Personality and Social Psychology
2006, Vol. 91, No. 5, 814–831

0022-3514/06/$12.00 DOI: 10.1037/0022-3514.91.5.814

Stereotypes: Static Abstractions or Dynamic Knowledge Structures?

Leonel Garcia-Marques and A. Sofia C. Santos
Universidade de Lisboa

Diane M. Mackie
University of California, Santa Barbara

Stereotypes have been assumed to be long-lasting knowledge structures that persist even in the face of contrary evidence. However, there is almost no within-participant research relevant to this assumption. The authors describe 4 studies ($N = 267$), the first 3 of which assessed within-participant stereotype stability over a few weeks with measures of stereotypic trait verification, typicality ratings of exemplar sets, and exemplar retrieval. In the 4th study, the authors manipulated context stability. Overall, results showed only low-to-moderate stereotype stability. The stability obtained was a function of the perceived centrality of traits or exemplars and of context constancy. The authors discuss the implications of these results for abstractionist, exemplar, mixed, and connectionist models and identify possible mechanisms that underlie within-participant stereotype instability.

Keywords: malleability of knowledge structures, context sensitivity, stereotypes

> A stereotype is a *fixed* impression, which conforms very little to the fact that it pretends to represent, and results from our defining first and observing second. (Katz and Braly, 1935, p. 181, emphasis added)

> [S]tereotypes have been regarded as rigid because they are believed to be persistent over time. (Ashmore & Del Boca, 1981, p. 18)

Are stereotypes stable over time? For many, including the authors, the answer to this question seems, at first sight, quite obvious. Just think of the "Princeton trilogy" studies, for example (Gilbert, 1951; Karlins, Coffman, & Walters, 1969; Katz & Braly, 1933). Although the level of consensus decreased somewhat across studies, successive generations of Princeton University students conveyed only slightly more benevolent versions of basically the same stereotypes.

But are stereotypes stable over time within the same individual? Again, our gut feeling, and we suppose we are not alone, suggests a positive answer. But, in fact, we simply lack the relevant empirical evidence (for one exception, see unpublished longitudinal study, described in Rothbart & John, 1993). At best, we might derive theoretical expectancies from what we currently know about knowledge structures, including stereotypes. In particular, we might look to studies of nonsocial categorization, which have already raised the question of within-individual representational stability. In this article, we first briefly describe abstractionist, exemplar, mixed models, and connectionist perspectives on the question of within-individual stability of stereotypes. We then review relevant findings from studies of nonsocial common concepts and present four studies. The first three were designed to assess the stability of social category representations in three crucial domains: stereotypic property verification, graded structure, and exemplar retrieval. In the fourth study, we directly manipulated context stability and evaluated its effects on stereotype stability.

Classic Abstractionist Perspectives

According to early abstractionist positions, stereotypes play an important role in achieving cognitive economy (e.g., Crocker, Fiske, & Taylor, 1984; Fiske, 1980; Fiske & Taylor, 1984; Taylor, 1981). Fulfilling such a role demands cognitive structures that are both constant and persistent. In fact, according to these views, the need for cognitive stability coupled with the scarcity of cognitive resources forces social information processors to neglect much of the detail about individual members of social groups or categories. Perceivers do best by judiciously ignoring the least relevant characteristics of individual targets and going beyond the information given: in short, by becoming chronic abstractionists. Such abstractionist tendencies should in turn make stereotypes self-perpetuating and highly resistant to change (e.g., Hamilton & Trolier, 1986; MacArthur, 1982; Snyder, 1981). In fact, the many ingenious ways stereotypes resist change are among our discipline's most well-documented findings and popular class anecdotes. Early abstractionist views thus envisaged stereotypes, like mental representations of other objects, as enduring mental entities that exhibit an impressive degree of constancy in the face of environmental turmoil. Theoretically, then, stereotypes, at least under ideal measurement conditions, should exhibit high reliability across relatively extended periods of time within the same individual. Such views inspired decades of theory and research on stereotypes in the social cognition tradition (Abelson, 1994).

Leonel Garcia-Marques and A. Sofia C. Santos, Faculdade de Psicologia e de Ciências da Educação, Universidade de Lisboa, Lisboa, Portugal, Diane M. Mackie, Department of Psychology, University of California, Santa Barbara.

This research was supported in part by Grant POCTI/PSI/47252/2002 from the Sapiens Program from Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa to Leonel Garcia-Marques and A. Sofia C. Santos and by Public Health Service Grant MH63762 to Diane M. Mackie. We thank Lawrence Barsalou, Markus Brauer, and David L. Hamilton, who provided valuable commentary and criticism regarding a previous version of the article.

Correspondence concerning this article should be addressed to Leonel Garcia-Marques, Faculdade de Psicologia, e de Ciências da Educação, Alameda da Universidade, 1649-013, Lisboa, Portugal. E-mail: garcia_marques@sapo.pt

The Priority of the Specific: The Exemplar View

What perceivers actually reported about social categories soon triggered challenges to the abstractionist stance (i.e., Kahneman & Miller, 1986; Linville, Fischer, & Salovey, 1989; E. R. Smith, 1988). People can retrieve much more information about specific social category exemplars—including the contingency, range, and variability of group members' attributes (Linville & Fischer, 1993)—than should be included in any fully developed abstract stereotype. Thus it appears that the information gained about specific exemplars is never completely discarded but remains available to be sampled when required. According to the exemplar perspective, stereotypes, group judgments, and group generalizations are the result of an exhaustive parallel search of a retrieved subset of stored exemplar representations. How much within-individual stability should we expect according to this view? Not much, because exemplar retrieval is at least partially guided by the context the judgment is made in (Garcia-Marques & Mackie, 1999; Kahneman & Miller, 1986; E. R. Smith, 1988). When the judgment context changes, so too does the relative impact of specific exemplars (Sia, Lord, Blessum, Thomas, & Lepper, 1999). Thus according to exemplar views, within-individual stability of stereotypes should be relatively low—a clear disavowal of the abstractionist position.

The Eclectic Edge: Mixed Models of Social Categorization

Contemporary views have more often sought the conditions under which group judgments reflect either abstractions or exemplars, rather than pitting one perspective against the other in a quest for the ultimate truth (McGuire, 1983). Many current researchers adopted hybrid or mixed model views that suggest that group concepts or stereotypes include both abstractions and specific exemplars (Babey, Queller, & Klein, 1998; Brewer, 1988; Fiske & Neuberg, 1990; Hamilton & Sherman, 1994; Judd & Park, 1988; Sherman, 1996; Zarate & Smith, 1990). Actually, most mixed models confer primary status on abstractions and assign exemplars more of a last-resource role in judgments (e.g., Babey et al., 1998; Brewer, 1988; Fiske & Neuberg, 1990; Hastie & Park, 1986; for a similar argument, see E. R. Smith, 1998). Abstractions play the primary role because they allow cognitive economy, but abstractions are complemented with exemplar information either because abstractions take time to develop (e.g., Sherman, 1996) or because abstractions are sometimes too narrow to accommodate the idiosyncrasies of atypical group members (e.g., Babey et al., 1998). How much within-individual stability would mixed models lead us to expect? It would depend, of course, on the relative weight of abstraction and exemplar components. The abstraction component should be relatively stable over time (or it loses its cognitive efficiency edge), but if exemplar retrieval changes with the context, then only moderate stability might be expected.

Patterns That Connect: The Connectionist Perspective

The connectionist perspective is a relatively new contender in the literature, but a number of connectionist models of stereotyping have already emerged (Kashima, Woolcock, & King, 1998; Queller & Smith, 2002; E. R. Smith & DeCoster, 1998; Van Rooy, Van Overwalle, Vanhoomissen, Labiouse, & French, 2003). According to this perspective, stereotypes are represented by dynamic activation patterns that occur in networks formed by simple and undifferentiated nodes. Each node receives positive or negative activation from neighboring nodes according to their respective connection weights. These connection weights derive from the previous history of activation from neighboring nodes. As all information is represented in the same network by different activation patterns, connectionist representations are superimposed and/or distributed (but see Van Rooy et al., 2003). Note that the information learned by such a network is not stored and consequently cannot be retrieved. It must be reconstructed from connection weights in response to input activation cues. Although network learning is preserved in connection weights, this reconstructive process is strongly affected by the immediate context because both the immediate context and previous learning are being represented at the same time in the same network. It is obvious that the dynamic nature of connectionist representations is at odds with abstractionist predictions about within-individual stability of stereotypes. Of course, whereas activation patterns can change very rapidly from moment to moment, connection weights change only very slowly and keep the network from behaving erratically. Even so, from the connectionist perspective, stereotypes are certainly not rigid knowledge structures and within-individual stability in stereotyping is expected to be only moderate at best. In sum, although classic abstractionist perspectives would predict a high degree of within-individual stability in social category representation and use, mixed model and connectionist approaches suggest greater variability, and exemplar models predict maximal context-driven variability.

The Empirical Case of Common Concepts

Empirical evidence from studies assessing the stability of common concepts and categories also found much more instability than a classic abstractionist position would suggest. The same individual on two different occasions (24 hr or 2 months apart) exhibited only modest reliability in defining and characterizing common concepts (Barsalou, Spindler, Sewell, Ballato, & Gendel, 1987; Bellezza, 1984a, 1984b), retrieving exemplars from common categories (Bellezza, 1984c), classifying instances into categories (McCloskey & Glucksberg, 1978), and rating the typicality of instances relative to their parent categories (Barsalou, Sewell, & Ballato, 1986). Other research showed that common categories are largely context sensitive, in that the immediate linguistic context biases both how typical an instance is judged to be as well as how fast it can be accessed (Roth & Shoben, 1983).

Stability in Social Category Representation

Thus the preponderance of theoretical expectation and relevant empirical evidence suggests a considerable degree of variability in social category representation, even within the same individual. Should we expect mental representations of social categories to show the same degree of fluidity as nonsocial categories? Stereotypes share crucial cognitive features with other mental representations and so might be expected to show the same degree of malleability and context sensitivity. At the same time, there are reasons to think that social stereotypes might be more abstract in

nature than common categories and, therefore, less fluid. After all, stereotype content is often developed from hearsay without direct intergroup contact (Linville & Fischer, 1993); such information comes in an abstract linguistically encoded form from the start. In addition, stereotypes bias information processing in a number of self-perpetuating ways, increasing stereotype stability (see Hamilton & Sherman, 1994; E. R. Smith & Mackie, 1995, for reviews). For these reasons, stereotype representations may well be more stable than those of common concepts.

Thus, we assessed as one of the primary goals of the first three studies to be reported here whether stereotypes evidenced the same type of context sensitivity and malleability demonstrated for mental representations of nonsocial categories by using longitudinal methodology (for a review, see Barsalou, 1987, 1989; Barsalou & Medin, 1986). The studies reported here belong to the very few that directly examine the within-participant stability of social stereotypes (for a precedent, see Rothbart & John, 1993).

Study 1

How likely is a given individual to select the same attributes to characterize a social category at two different points in time (the property verification task, Barsalou et al., 1987)? This seems a particularly appropriate way to assess stereotype stability because attribute or property selection was the first procedure used to study stereotypes empirically (i.e., the adjective checklist of Katz & Braly, 1933) and has remained quite popular (for a recent review see Dovidio, Brigham, Johnson, & Gaertner, 1996). The checklist was of course the procedure used in the series of studies that assessed stereotypes in different generations of the same student population to infer the temporal persistence of stereotypes (Gilbert, 1951; Karlins et al., 1969; Katz & Braly, 1933). Our primary goal in Study 1 was to assess stereotype stability as reflected in stability of selection of attributes as stereotypic within individuals over time.

We also assessed whether all stereotype content showed equal levels of stability or instability across time. Most current theories of representation conceive of conceptual cores, comprising more central or important, and more stable, attributes or features. Some proposed that conceptual cores are definitional (Armstrong, Gleitman, & Gleitman, 1983; Osherson & Smith, 1981; E. E. Smith & Medin, 1981). According to core-plus-identification views, mental representations of categories contain definitional cores and identification procedures based on typicality. To preserve the existence of definitional cores, the core-plus-identification view allows for categorization instability in terms of identification procedures—such procedures focus on the attributes that proved helpful in previous categorization tasks and reflect idiosyncratic experience, resulting in less stability. The definitional cores, however, reflect natural and logical invariants and should be perfectly stable. Other authors proposed that conceptual cores contain intuitive theories (Murphy & Medin, 1985). On the other hand, Barsalou (1982, 1987, 1989) saw conceptual cores as based on categorization experience. Certain properties become core properties because they are processed in conjunction with a category on so many occasions that they become automatically associated with it. Regardless of the mechanism(s) that might produce conceptual cores, the question of whether attribute centrality or importance moderates stereotype instability is a crucial one to examine in the case of social categories.

Finally, we also addressed the question of perspective. Barsalou and his colleagues (1987; see also Barsalou & Sewell, 1984) asked participants to generate the most important attributes of a number of categories from different points of view (e.g., from the perspective of the self, a suburban housewife, a country redneck, etc.) and found similar degrees of within-participant instability. We chose to manipulate point of view in a way that closely corresponds to an important variable in the social stereotype literature—the difference between cultural and individual stereotypes. Devine and Elliot (1995) suggested that cultural stereotype content—what "people in general" think—seems to be more stable than individual or personal stereotypes. However, because their data were not longitudinal, it remains to be seen whether cultural stereotypes evidence greater stability than do personal stereotypes in a within-participant sense.

Method

Participants and Design

Participants were 46 University of Lisbon students (26 female and 20 male) who volunteered for the study at the request of the researcher. The design of the study was a 3 (social category: gypsy, gay, and African immigrant) × 2 (perspective: cultural and individual) × 2 (session: 1 and 2) within-participants factorial.

Pretesting the Adjective Checklist

A different group of 31 psychology sophomores were asked to give descriptions of three social groups (gypsy, gay, and African immigrant).[1] Participants were instructed to generate a list of attributes for each social category on the basis of cultural stereotypes. From these data and for each social group, the nine most frequently mentioned attributes were selected (excluding those overlapping in meaning). The nine traits more frequently identified are in bold in Appendixes A, B, and C for gypsy, gay, and African immigrant social categories, respectively. Whenever possible, attribute antonyms were added to the list. Because in this free-response task participants almost always generated personality traits, the final list included only this type of characteristic. This task produced a final list of 43 personality traits.

Procedure

All participants were tested twice as a group, with the second session following the first session by 2 weeks. To identify a participant's answers across sessions, we asked each participant to indicate his or her birthday date and that of his or her mother, assuring anonymity. Participants were given a booklet containing the instructions and experimental materials.

The instructions for the individual stereotype trait selection task read as follows:

[1] In previous research (Santos, 2001), we used occupational groups (medical doctors, computer programmers, and disco bouncers) as targets. The level of instability we obtained in stereotype assembling was considerable, but it could plausibly be argued that these stereotypes are weak or relatively benevolent. We chose the target social categories because they were consensually identified by pretest participants as strong, pervasive, and clear-cut stereotypes in contemporary Portuguese society and thus provided a conservative test of the fluidity hypothesis.

> Society is composed of many different groups about whom we usually have some knowledge. In fact, the ease with which we form relatively well-defined impressions about the individuals and social groups who surround us greatly simplifies our social life. These impressions about groups are often generic, and don't apply to every member of the group but only to a percentage of them. For instance, when we say that computer programmers are intelligent we are not saying that every computer programmer is intelligent, but that a significant percentage of them are. These generic impressions are, obviously, simplifications. They are not judgments based in objective data. And even when we recognize differences between groups, that doesn't mean that one group is better or worse than another group. In this study, you will be asked to give impressions about some social groups. Naturally, we don't all have the same ideas about such groups. There are no right or wrong answers. We are interested in your personal impressions, your intuitions, and your gut reactions, and not so much in what you think it is proper to say.

The instructions for the cultural stereotype version of the trait selection task read as follows:

> Society is composed of many different groups about whom people in general have some knowledge. In fact, the ease with which people form relatively well-defined impressions about the individuals and social groups that surround them greatly simplifies their social life. On many occasions, either through hearsay or direct contact, we find out something about the impressions that people in general have about social groups. In this study, you will be asked to give your opinion about what people in general think about some social groups. Naturally, the impressions that people in general have about social groups may or may not reflect your personal beliefs. So give your answer based on what you know to be the culturally shared beliefs people in general have about those social groups, whether or not you believe those ideas to be true.

In both perspective conditions, participants then had to choose and write down from the full list of 43 traits the 5 that best described each of the target groups (following Katz & Braly, 1933). Participants then made ratings of trait centrality and importance of the traits they chose. Evaluations were made on two types of 9-point scales, one ranging from 1 (*not at all central*) to 9 (*very central*), and the other ranging from 1 (*not at all important*) to 9 (*very important*). Ratings of centrality and importance were highly correlated (correlations across group replications ranged from .62 to .77, $p < .001$, at least) and were averaged to form a single centrality index. At the second session approximately 2 weeks later, participants were again given the same perspective instruction they received in the first session and completed the checklist again (but they did not make new centrality ratings), and they were fully debriefed and thanked.

Results and Discussion

Aggregate Sample (Within-Item) Stability

When the checklist methodology is used, stereotype stability is typically assessed by the correspondence between the attributes chosen to describe the social category across different studies (e.g., Devine & Elliot, 1995). We followed this procedure to compare attributes chosen across the two sessions (see Appendixes A, B, and C). Across sessions, agreement was very high (the within-item correlations varied from .91 to .97, including both cultural and individual stereotypes). A similar degree of agreement was found by Rothbart and John (1993; average group stereotype across session 4 years apart, $r = .95$).[2] Note that, if we used only this analysis to assess stereotype stability, as previous studies have done, we would find results that vindicated the abstractionist position.

Within-Participants Stability

To determine the degree of overlap in the attributes used to describe social groups by any one participant across sessions, we used a common-element correlation (Bellezza, 1984a, 1984b, 1984c). To compute this value, by participant, we divided the number of common attributes generated in both sessions by the square root of the product of the total number of attributes generated in first session and the total number of attributes generated in second session (the geometric mean). This measure of correlation represents the proportion of common to total items and varies between the values of 0 and 1.

The overlap values indicated that there was only moderate correspondence between the category's attributes selected in the two sessions. Mean overlap scores ranged in value from .48 to .60 (see Table 1), indicating that only approximately half a participant's trait selections for a category in one session were also chosen in the second session. A similar result was obtained by Rothbart and John (1993; average personal stereotypes across sessions 4 years apart, $r = .50$).

These levels of within-participant stability in the content of social categories were generally similar to those found with nonsocial categories, using a similar experimental paradigm. In fact, when we computed the 5% confidence intervals of these common-elements correlations (following Schmidt, 1996), they overlap those found in the common objects category domain (see Table 1). Thus, despite the factors noted earlier that might have conferred even greater stability on social stereotypes, social stereotypes were no more stable than common objects concepts. These results stand in stark contrast to abstractionist expectations and reverse the picture we obtained when stability was assessed within-item aggregating across participants.

A 3 (social category) $\times$ 2 (perspective) analysis of variance (ANOVA) on the mean overlap scores revealed only a significant effect for social group, $F(2, 88) = 4.06$, $p = .021$, $MSE = .046$, with less overlap in across-session responses about the African immigrant group ($M = .49$) than about the gypsy group ($M = .58$), with responses about the gay group falling in between ($M = .53$). Note that although we anticipated higher levels of stability for cultural relative to individual stereotypes, we did not find that to be the case.

Impact of Centrality on Stability

As a first assessment of whether centrality affected attribute stability, we compared the average centrality of traits common across sessions versus traits unique to sessions. The results from *t* tests for dependent samples are shown in Table 2.

[2] Rothbart and John (1993) did not assess the stability across time in the choice of traits that best describe a social group but the stability across time of ratings of how characteristic of a social group a set of traits is. Rothbart and John's (1993) methods thus approximate a mix of the different aspects of stereotypes we studied separately in Studies 1 and 2. Although we included the reference to Rothbart and John (1993) in Study 1, their results also compare and converge with the results of Study 2.

Table 1
Within-Participant Trait Overlap Across Sessions for Personal and Cultural Stereotypes by Social Category

Social category	Personal stereotype		Cultural stereotype	
	M	5% CI	*M*	5% CI
Gypsy	.56	.36–.76	.60	.41–.79
Gay	.56	.36–.76	.49	.37–.71
African immigrant	.50	.38–.72	.48	.26–.60
Common taxonomic categories (from Barsalou et al., 1987)	.43	.24–62		

Note. CI = confidence interval.

Mean centrality was generally higher for common than for unique traits although the magnitude of these differences was relatively small and the differences did not always reach conventional statistical significance. The traits that were stable across sessions were considered more central and important, providing some evidence for the core attributes idea.

To explore the role of centrality further, we performed a median split on the centrality ratings of the traits chosen by each participant and computed separate common-element correlations. To distinguish between central and peripheral traits, we computed, for each participant, the median of the centrality ratings made in the first session. Traits with ratings below the median were considered peripheral traits, and traits with ratings above the median were considered central traits. Traits with evaluations equal to the median were excluded. The total number of attributes selected in the first session was used to compute the common-element correlations. Mean overlap scores by condition are shown in Table 3.

Impact of Centrality on Accessibility

Traits high in centrality were more stable across sessions than traits low in centrality. These highly central traits appear to be responsible within individuals for the moderate level of stability found across sessions. Barsalou (1982, 1987, 1989) referred to such attributes as *context independent* and proposed that, rather than being definitional, these are simply properties that have been processed frequently with the category. If that is the case, they should be highly accessible and therefore occur early in participants' protocols. That is, a trait's degree of centrality should negatively predict its ordinal output position.

Table 2
Differences in Mean Evaluations of Centrality and Importance (Averaged) for Common and Unique Traits by Social Category, Study 1

Variable	Gypsy	Gay	African immigrant
Personal			
Common traits	5.92	6.03	5.96
Unique traits	5.47	5.29	5.11
t	1.63, *df* = 43	2.08, *df* = 43	2.12, *df* = 43
p	.110	.042*	.038*
Cultural			
Common traits	6.24	5.75	6.12
Unique traits	5.44	5.46	6.20
t	2.02, *df* = 43	0.74, *df* = 42	−0.26, *df* = 43
p	.049*	.462	.79

Note. Asterisks indicate differences that reach conventional statistical significance.

Table 3
Degree of Within-Participant Trait Overlap by Social Category, Type of Stereotype, and Trait Centrality Study 1

Stereotype and trait	Gypsy	Gay	African immigrant
Personal			
Low centrality	.38	.37	.54
High centrality	.64	.59	.69
Cultural			
Low centrality	.44	.32	.50
High centrality	.78	.51	.49

Note. A 3 (social category) × 2 (perspective) × 2 (centrality/importance) analysis of variance was performed on the mean overlap scores. Only a significant effect of centrality emerged, $F(1, 12) = 7.98$, $p = .015$, $MSE = .180$, indicating that, as expected, traits high in centrality were significantly more stable ($M = .62$) than traits low in centrality ($M = .43$).

The output position of each trait in each participant's protocol for each social category for each type of belief was correlated with the trait's centrality, by condition. One Pearson correlation was computed for each participant's protocol for each social group and for each type of belief. Each correlation was then transformed to Fisher *Z* scores and back to Pearson coefficients so that the average correlation could be calculated. Correlations in all conditions (see Table 4) indicated a tendency for highly central traits to be retrieved early, as expected.

Impact of Stability on Accessibility

Are the traits that are more accessible in the first session more stable across sessions? If so, then traits with earlier output positions in the first session should be more likely to be also generated in the second session than traits with later output positions. Bellezza (1984a) found that exemplars retrieved earlier were more likely to be common to both sessions than exemplars retrieved later. Barsalou et al. (1987) also reported that common traits appeared earlier in participants' protocols than did unique properties.

To explore this idea, we performed a median split on the output position of the traits chosen by each participant and computed separate common-element correlations. To distinguish between traits with early and late output positions, we computed for each participant the median on the basis of the output positions of the traits chosen in the first session. Traits with output positions below the median were considered early traits, and traits that appeared

Table 4
Average Correlation Between Trait Centrality and Trait Output Position by Social Category and Stereotype

Stereotype	Gypsy	Gay	African immigrant
Personal	−.12	−.18	−.13
Cultural	−.25	−.29	−.19

after the median were considered late traits. Traits at the median position were excluded. The total number of attributes selected in the first session was used to compute the common-element correlations. Mean overlap scores by condition are shown in Table 5.

A 3 (social category) × 2 (perspective) × 2 (output position) ANOVA on the mean overlap scores revealed a significant effect of output position, $F(1, 43) = 17.30$, $p = .0001$, $MSE = .138$. As expected, traits with early output positions were significantly more stable ($M = .60$) than traits with late output positions ($M = .46$). More accessible traits were more stable across sessions. Output position also interacted with perspective, $F(1, 43) = 8.25$, $p = .006$, $MSE = .083$. Accessible traits were more stable ($M = .63$) and less accessible traits less stable ($M = .42$) for cultural stereotypes than for individual stereotypes (.57 and .50, respectively).

The results of Study 1 revealed about the same moderate level of trait stability in social category representations as other researchers have found with nonsocial categories (for reviews, see Barsalou, 1987, 1989; Barsalou & Medin, 1986). When a participant was asked to choose five traits that best describe a group from a list of 43 traits, the chance that he or she would include one trait across time was approximately equal to the probability of including it at only one time. This result held for both personal and cultural stereotypes. These results are clearly at odds with early abstractionist assertions, which might predict perfect reliability. Moreover, had we looked only at across-participants within-item agreement, as previous research did (Devine & Elliot, 1995), our conclusions would be quite the opposite. Although these results clearly clash with abstractionist positions about stereotype stability, they fit quite well with exemplar, mixed model, and connectionist alternative views.

Not all kinds of stereotype content were equally unstable, however. In fact, more central attributes were found to be significantly more stable than peripheral ones, though the stability of even highly central traits was still far from what a full-fledged abstractionist perspective would lead us to expect. Traits with early output positions were also more stable within participants than traits with late output positions, indicating that accessible traits within sessions were also accessible across sessions for a given individual.

Most of the instability of stereotype content across sessions was due to attributes considered to be less central and/or important. Such traits appear to be relatively inaccessible and thus have late output positions. These results are consistent with theories that postulate the existence of a category conceptual core, whether this core is definitional in nature (Armstrong et al., 1983; Osherson & Smith, 1981) or more experience-based (Barsalou, 1982, 1987, 1989). We defer further consideration of this issue to the General Discussion.

Table 5
Within-Participant Stability by Social Group, Type of Stereotype, and Trait Output Position

Stereotype and output position	Gypsy	Gay	African immigrant
Personal			
Early trait	.57	.63	.51
Late trait	.54	.47	.51
Cultural			
Early trait	.72	.58	.59
Late trait	.43	.43	.40

Study 2

Having demonstrated only moderate stability across time in stereotype content, we next assessed the degree of stability in the graded structure of social stereotypes. Graded structure (also known as typicality or exemplar goodness) reflects the extent to which different exemplars represent the category. Within a category, graded structure refers to the gradient of representativeness that exists across category members.

Graded structure has been found in a wide variety of categories, and a natural category without such structure has yet to be identified. Common taxonomic categories (e.g., fruit, birds, furniture; Rosch and Mervis, 1975), formal categories (e.g., odd numbers, square root; Armstrong et al., 1983), and ad hoc categories and goal-derived categories (Barsalou, 1983, 1985) all show graded structure. E. E. Smith and Medin (1981) and Medin and Smith (1984) have shown that graded structure predicts performance in a number of fundamental categorization tasks.

The graded structure of nonsocial categories has been shown to be dynamic. Depending on the population, individual, or context, the perceived typicality of category exemplars can vary widely (e.g., Barsalou, 1987). Barsalou and Medin (1986; see also Barsalou, 1985, 1987) argued that changes in the graded structure of a category provide further evidence for the fluid nature of category representation. Because an exemplar's typicality rating reflects its similarity to the category representation, changes in category representation would cause instability in category–exemplar similarity and thus instability in typicality ratings and thus graded structure (Barsalou, 1989).

Barsalou and Medin (1986) explored the extent to which any individual produces the same graded structure for a category across time but in the same context. To assess agreement, they correlated participants' graded structure on 1 day with their graded structure in the same context a few weeks later. Participants' exemplar typicality ratings in one session accounted for only about 64% of the variance in the later session. Because individuals showed much higher stability after only 1 hour's delay, the instability at longer delays could not be attributed to measurement error or some other source of random variability. Instead, it appeared to reflect changes in the individual's representation of the category between the sessions.

Is graded structure important for stereotypes? The answer is clearly yes. Such a typicality gradient is a major component of a very influent exemplar theory of stereotype representation and change (Rothbart & John, 1985). According to this theory, differences in typicality help to perpetuate stereotypes because typical exemplars have much higher probability of being retrieved and used in categorization-based tasks (Rothbart & Lewis, 1988; Rothbart, Sriram, & Davis-Stitt, 1996). In addition, changes in graded structure (i.e., an increase in the perceived typicality of atypical exemplars) are considered the most reliable signs of stereotype change (Maurer, Park, & Rothbart, 1995; see also Garcia-Marques & Mackie, 1999). On this basis, we again expected that graded structure in stereotypes might exhibit higher within-participant stability than the graded structure of common objects concepts.

Our second study assessed the extent to which social categories show within-participant stability in graded structure across time. By including manipulations of exemplar typicality, we could also assess whether typicality moderated this stability, replicating the moderating effects of attribute centrality found in Study 1. We did not manipulate belief type. Given that in Study 1 we found few differences in personal and cultural stereotypes, we asked only about cultural stereotypes because many participants feel more comfortable reporting cultural than individual stereotypes (Devine & Elliot, 1995).

Method

Participants and Design

Participants were 32 (23 female and 9 male) University of Lisbon students who volunteered for the study. The design was a 2 (social category: gypsy, gay, and African immigrant) × 2 (session: 1 and 2) within-participants factorial.

Pretest

Sixty participants first generated short descriptions of the first five exemplars, either real or imagined, that came to mind for each social category. Forty other participants then rated the typicality of each exemplar description on a 9-point rating scale ranging from 1 (*extremely uncharacteristic of the group*) to 9 (*extremely characteristic of the group*).[3] Participants were told to use the cultural stereotype—the views that people in general had about these groups—in making their ratings. On the basis of these ratings, we then selected the eight most consensually agreed atypical, eight moderately typical, and eight typical exemplars for each category for use in the study.

Procedure

Participants received a booklet with each of the 24 exemplars descriptions typed on a single page, in one set order. Participants judged the typicality of each exemplar with the following instructions:

> Society is composed of many different groups about whom we usually have some knowledge. Most of us have a certain mental image about what is typical or characteristic of each group, and, generally, we are able to identify examples that are typical of the group. Some members can be more similar to what is typical of the group than others. In this study, some examples of a social group's members will be presented and you will be asked to evaluate how good an example each exemplar is of his or her group. We are not so much interested in your personal opinions but in what are considered true by people in general. So it is absolutely essential that you assume the point of view of people in general while judging the typicality of exemplars. There are no right or wrong answers. We are interested in your first impressions, intuitions, and gut reactions, and not so much in what you think it is proper to say.

Participants were given as much time as necessary to complete the task. Participants were asked to rate the typicality of each exemplar of each of the three target groups on the 9-point rating scale. Participants performed the same tasks again approximately 4 weeks later and were fully debriefed after the session.

Results and Discussion

Within-Participants Correlation by Social Group

To assess within-participant agreement, we calculated for each participant Pearson correlation coefficients between typicality ratings in Session 1 and the same typicality ratings made in the same context 4 weeks later. The 32 participant correlations, computed for each social group, were transformed to Fisher *Z* scores and back to Pearson coefficients so that average correlations could be calculated.

Across the three social groups, we found that the average correlation was around .67. This moderate level of agreement across time was not higher than the level found with nonsocial categories (Barsalou et al., 1986), such that the respective 5% confidence intervals overlap. Table 6 shows the average correlation scores by social group.

Within-Participants Correlation by Exemplars Levels of Typicality

To distinguish between atypical, moderately typical, and typical exemplars, we computed, for each exemplar, the 33rd and 66th percentile of the distribution of typicality judgments made by all participants in the first session. Exemplars with a typicality rating below the 33rd percentile were considered atypical, exemplars between 33rd and 66th percentiles (inclusively) were considered moderately typical, and exemplars with a typicality rating above the 66th percentile were considered typical.

To assess the degree of correlation between the graded structures generated in the first and second sessions, we calculated within-participants Pearson correlation coefficients by degree of typicality. Average within-participant correlation scores increased from .54 for atypical exemplars, through .58 for moderately typical exemplars, to .62 for typical exemplars, indicating that within-participant stability of typicality ratings increased with exemplar typicality.

In sum, participants showed only moderate reliability in the exemplar typicality ratings they generated 4 weeks apart. These results indicate levels of instability in the graded structure of social categories similar to those found with nonsocial categories. Stability increased linearly with exemplar typicality, a finding that diverged from Barsalou et al.'s (1986) finding that typicality instability peaked with moderately typical exemplars. This may suggest that slightly different mechanisms underlie perceived typicality of social and nonsocial exemplars. Alternatively, the slightly different methodology used in the two studies may be responsible for this difference. Regardless, it is clear that instability in graded structure does not come only from perceptions of atypical exemplars—considerable instability was found even in ratings of typical exemplars.

Study 3

As a third assessment of fluidity in social category representation, we examined the reliability of retrieval of social category exemplars. In the nonsocial domain, Bellezza (1984a) demon-

[3] Instead of asking participants to judge how typical each exemplar is, some researchers (Barsalou, 1985) ask participants to rate how good an example a given examplar is of its category because typicality might bias participants toward using frequency of instantiation (people's estimates of how often they have encountered an exemplar as a category member). However, asking how bad or good something is might also have evaluative connotations.

Table 6
Within-Participant Correlation Across Sessions by Social Category

Social category	Within-participant correlation	
	M	5% CI
Gypsy	.72	.55–.89
Gay	.58	.35–.81
African immigrant	.70	.52–.88
Common taxonomic categories (Barsalou et al., 1986)	.70	.54–.86

Note. CI = confidence interval.

strated some representational fluidity when he asked participants to generate category instances in two recall tests 1 week apart. He found that only about 69% of the category instances generated in the first session were also generated in the second session.

Does exemplar retrieval from stereotypical social category representations show a similar degree of instability? This question is critical for exemplar-based categorization models (e.g., E. R. Smith & Zárate, 1992) because these models argue that judgments about a category's most characteristic attributes are based on the attributes of exemplars activated at the time (i.e., Bodenhausen, Schwarz, Bless, & Wanke, 1995). Mixed model perspectives also of course understand instability in stereotypes as deriving from instability in exemplar retrieval rather than from changes in the abstracted components of the representation. Thus if exemplar retrieval within-participants exhibited a much higher degree of stability than the level already found for property verification (Katz & Braly's (1933) task) or graded structure, exemplar and mixed model views would be hard pressed to account for such findings. Abstractionist views would of course have little to say about the reliability of exemplar retrieval.

Thus, on the basis of the results of Studies 1 and 2, we expected a similar degree of instability in exemplar retrieval as we found in stereotype attribute assignment (Study 1) and graded structure (Study 2). We again assessed this index of representational stability at various levels of exemplar typicality.

Method

Participants and Design

Participants were 65 University of Lisbon students (40 women and 25 men) who volunteered for the study in return for partial course credit. The design was a 3 (gypsy, gay, and African immigrant) × 2 (Session 1 and Session 2) within-participants factorial.

Procedure

Participants were tested in small group sessions of up to 10 members. General instructions were the same as in Study 1. Participants were debriefed after Session 2.

First exemplar retrieval task. Participants were asked to generate descriptions of five (more or less) different members of each of the three social categories, always in the same order: gypsy, gay, and African immigrant. Participants described each exemplar, which could be a real or imaginary person, in a few sentences.

Typicality judgments. After completing the 15 descriptions, participants evaluated each exemplar's typicality on a 9-point rating scale ranging from 1 (*extremely uncharacteristic of the group*) to 9 (*extremely characteristic of the group*).

Second exemplar retrieval task. Approximately 2 weeks later, participants returned and completed the exemplar retrieval task again. At the end of the second session, each participant was given the descriptions he or she had generated in the first session and asked to say which across-sessions descriptions depicted the same category member. Thus, we coded a second session description as repeated across sessions whenever the participant who generated it identified it that way. Note that this procedure set a conservative standard for instability, working against our hypothesis.

Results and Discussion

Number of Exemplars Generated

Although participants were asked to generate about five exemplars, some generated more than this and some fewer. An ANOVA on the mean number of exemplars retrieved by session and category (see Table 7, Columns 1 and 2) revealed a significant main effect for social category, $F(2, 128) = 13.47$, $p = .0001$, $MSE = .609$. Participants produced more exemplars for the gypsy target group ($M = 4.80$) than for the other two social categories, $F(1, 64) = 17.69$, $p = .0001$, $MSE = .925$, which did not differ one from another ($M = 4.38$ vs. $M = 4.36$), $F(1, 64) = .118$, $p = .731$, $MSE = .292$.

Within-Participants Reliability

To determine the degree of overlap in the exemplars retrieved by any one participant across the two sessions, we computed common-element correlations for each social group by participant (Bellezza, 1984a, 1984b, 1984c). The number of common exemplars (as identified by the participant) retrieved in both sessions was divided by the square root of the product of the total number of exemplars retrieved in the first session and the total number of exemplars retrieved in the second session. Mean overlap scores ranged from .55 to .62 (see Table 7, Column 3), indicating the modest reliability with which social category exemplars are retrieved from memory. Again, this level of within-participant instability was about the same as that found with nonsocial categories, using an equivalent experimental paradigm (Bellezza, 1984a),

Table 7
Mean Number of Exemplars Generated and Across Session Stability by Social Category

Social category	Session 1	Session 2	Within-participant overlap	
			M	5% CI
Gypsy	4.83	4.77	.55	.38–.72
Gay	4.37	4.38	.62	.47–.77
African immigrant	4.34	4.37	.55	.38–.72
Common taxonomic categories (Bellezza, 1984a)			.69	.51–.87

Note. CI = confidence interval.

such that their respective 5% confidence intervals overlap (see Table 7).

Typicality Judgments of Retrieved Exemplars

Paralleling the procedures used in Study 1, we assessed whether the exemplars that were common versus unique across sessions differed in typicality. To do so, arithmetic typicality means were computed for the set of common exemplars and also for the set of unique exemplars. The results of *t* tests for dependent samples are shown in Table 8.

Results supported the findings from Study 1: Exemplar typicality was a good predictor of stability in exemplar production, at least in two of the groups. For the gypsy and gay categories, common exemplars (generated in both sessions) were judged as significantly more typical than uniquely generated instances.

Within-Participants Reliability Concerning High and Low Typical Exemplars

Each participant's median typicality rating in the first session was computed. Exemplars with typicality ratings below the median were considered atypical exemplars and exemplars with typicality ratings above the median were considered typical exemplars. Exemplars with ratings equal to the median were excluded. Common-element correlations were then computed as described earlier.

A 3 (social category) × 2 (exemplar) ANOVA revealed a significant effect of typicality, $F(1, 27) = 18.81$, $p = .0002$, $MSE = .158$, showing that highly typical descriptions (.77) were more likely to be generated across sessions than less typical descriptions (.50; see Table 9).

Correlation of Degree of Typicality Attributed to Exemplars and Output Position

Following the same reasoning applied to traits in Study 1, we expected highly typical exemplars to be more accessible and therefore to be generated earlier than less typical exemplars, causing exemplar typicality and exemplar output position to be negatively related. The output position of each exemplar generated by each participant in the first session was correlated with that exemplar's typicality rating for the relevant social group. Each correlation computed was transformed to Fisher *Z* scores and back to Pearson coefficients so that average correlations could be calculated. Correlations in all conditions indicated that, as expected, typical exemplars tend to be retrieved early ($r = -.59$ for gypsy exemplars, $r = -.62$ for gay exemplars, and $r = -.36$ for African immigrant exemplars).

Table 8
Mean Typicality of Common and Unique Exemplar Descriptions by Social Category, Study 3

Exemplar descriptions	Gypsy	Gay	African immigrant
Common	6.32	6.02	6.37
Unique	5.51	4.28	5.94
t (difference)	1.98, *df* = 64	4.47, *df* = 64	1.17, *df* = 64
p	.052	.000	.243

Table 9
Within-Participant Stability by Social Category and Exemplar Typicality

Exemplar typicality	Gypsy	Gay	African immigrant
Low	.52	.46	.53
High	.65	.88	.78

Within-Participant Reliability for Accessible and Less Accessible Exemplars

Following Bellezza (1984a), we also tested the idea that accessible exemplars (those output early) in the first session were more likely to be generated in both sessions. To distinguish exemplars with early and late output positions, we computed, for each participant, the median output positions of the exemplars generated in the first session. Exemplars with output positions below the median were considered accessible, and exemplars with output positions above the median were considered less accessible. Common exemplars with evaluations equal to the median were excluded. The total number of exemplars generated in the first session was used to compute the common-element correlations. Mean stability scores by condition are shown in Table 10.

A 3 (social category) × 2 (exemplar) ANOVA indicated that, as expected, exemplars with early output positions had a significantly higher stability (.63) than exemplars with late output positions (.51), $F(1, 64) = 11.33$, $p = .001$, $MSE = .126$.

In sum, the results of the exemplar retrieval task replicated the results of Studies 1 and 2, with social category representations revealing levels of instability similar to those found with nonsocial categories. Once again, exemplar typicality affected stability, with more representative exemplars showing higher stability across time. Corroborating these results, within-participants reliability revealed more typical exemplars to be more stable over time. Once again, however, instability was not confined to the atypical exemplars: even the most typical exemplars showed some level of instability.

Study 4

Although we found a consistent picture across the three previous studies, it is possible that the degree of instability we obtained was, at least partially, due to measurement error or to participants' deliberate monitoring of their responses. If participants felt that, for some reason, they should vary in the second session the responses they had given to identical instructions 3 or 4 weeks earlier, the level of instability we obtained might represent little more than an artifact derived from repeated measurement.

Table 10
Within-Participant Stability by Social Category and Output Position

Output position	Gypsy	Gay	African immigrant
Early	.56	.71	.64
Late	.51	.56	.47

Our primary goal in the fourth study was to assess this possibility. We did so in two different ways. First, to show that stereotype instability is a result of context sensitivity and not simply a byproduct of measurement error, we manipulated the consistency of the context in which participants identified the traits associated with a given social stereotype. We did so by having participants perform an exemplar typicality rating task immediately before completing the stereotype trait selection task used in Study 1 both during Session 1 and Session 2. The rated exemplars were either equivalent or nonequivalent in typicality across sessions. If the attributes participants chose across sessions reflected only random variation, this manipulation should have no effect. In contrast, if stereotype fluidity reflects context sensitivity, variation in stereotype attribute choice would be driven by the context manipulation.

Second, to assess the possibility that the variation in responses obtained across sessions in Studies 1–3 was due to participants' deliberate attempts to change their responses, we compared their spontaneous responses with their attempts to accurately recall the choices they had made during the first session. Finally, we also wanted to assess the subjective or perceived stability of stereotypes to see whether our participants were aware of the fluidity in their social category representations or whether they perceive illusory consistency in their stereotypes (Ross & Conway, 1986).

Method

Participants and Design

Participants were 124 University of Lisbon students (71 female and 53 male) who volunteered for the study at the request of the researchers in return for partial course credit. The design of the study was a 2 (session: 1 and 2) × 2 (social category: gypsy and African immigrant) × 2 (exemplar: typical or atypical, rated in Session 1) × 2 (exemplar typicality: equivalent or nonequivalent across sessions) × 2 (task: stereotype attribute selection or recall in Session 2) mixed factorial design, with the first two factors being within-participants.

Procedure

All participants were tested twice as a group with the second session following the first session by 1 month. Participants were tested in small group sessions of up to 10 people. In Session 1, all students performed a typicality judgment task, completed the stereotype attribute selection task described in Study 1, and evaluated the centrality of each trait they selected. In Session 2, the same students performed a second typicality judgment task and then either repeated the stereotype attribute selection task or attempted to recall their responses from Session 1. Participants who repeated the trait selection task also estimated the number of traits chosen in Session 2 that they had also chosen in Session 1.

Manipulation of context. To manipulate the context in which participants chose stereotype attributes, participants were first required to complete an exemplar typicality rating task. In both Session 1 and Session 2, participants first rated the typicality of three exemplars from the social category of gypsy and three from the social category of African immigrant. Participants received a booklet with each exemplar description typed on a single page in one set order. Instructions were as described in Study 2. Participants rated the typicality of each exemplar of each of the two groups on a 9-point rating scale ranging from 1 (*extremely uncharacteristic of the group*) to 9 (*extremely characteristic of the group*). The three-exemplar sets were either typical (two typical exemplars and one neutral exemplar) or atypical (two atypical exemplars and one neutral exemplar). Exemplars were selected from the pretest previously described in Study 2.[4]

We manipulated the equivalence of the exemplars rated for typicality across the two sessions. In the equivalent condition, the three exemplars rated in Session 2 were different from but equivalent in typicality to the ones judged in Session 1 (that is, participants saw either typical or atypical exemplars in both sessions). In the nonequivalent condition, the three exemplars rated in Session 2 were also different from, as well as nonequivalent in typicality to, those rated in Session 1 (that is, participants who first rated typical exemplars now were asked to rate atypical exemplars and vice versa).

Stereotype attribute selection task. The stereotype attribute task always immediately followed the typicality rating task. As described in Study 1, participants had to choose from a list of 43 personality traits the five traits that best described each of the target groups. After choosing the traits, participants rated the centrality to the relevant stereotype of each of their chosen traits. Evaluations were made in a 9-point scale ranging from 1 (*not at all central*) to 9 (*very central*). All participants completed the attribute choice task immediately after the typicality rating task in Session 1. In Session 2, approximately 4 weeks later, half the participants repeated the attribute choice task and then estimated the number of traits that they had just chosen, which they had also chosen in the first session. These estimates served as a measure of subjective overlap.

Memory task. The other half of the participants in Session 2 (all of whom had completed the stereotype attribute task in Session 1) were asked to reproduce the choices they had made in the first session as best they could.

Results and Discussion

Within-Participant Stability

As in Study 1, we computed common-element correlations (overlap scores) from the choices of each participant across sessions (Bellezza, 1984a; McNemar, 1969). We then computed a 2 (session: 1 and 2) × 2 (social category: gypsy and African immigrant) × 2 (exemplar: typical or atypical, rated in Session 1) × 2 (exemplar typicality: equivalent or nonequivalent across sessions) × 2 (task: stereotype attribute selection or recall in session 2) mixed model ANOVA on these overlap scores. Only two significant main effects emerged. The first was an equivalence main effect, $F(1, 113) = 5.27$, $p = .024$, $MSE = .044$, such that participants who rated two sets of exemplars equivalent in typicality across sessions showed more stability in their stereotype attribute choices across session ($M = .51$) than participants who rated exemplars who differed in typicality across sessions ($M = .45$). The second significant main effect was a second session task effect, $F(1, 113) = 54.00$, $p = .000$, $MSE = .044$, showing that participants who performed the stereotype attribute selection task a second time were much more consistent ($M = .59$) than participants who were intentionally trying to reproduce their first session choices ($M = .38$).

The superior consistency of participants who performed the stereotype attribute selection task a second time over those who tried to recall their choice may seem surprising, but such results are far from unprecedented. Repeating the selection task a second time may be considered as an implicit memory test of the choices made

[4] A group of pretest participants generated descriptions of exemplars for each social category. Forty other participants then rated the typicality of each exemplar description on a 9-point rating scale. On the basis of these ratings, we then selected the most consensually agreed-on atypical, neutral, and typical exemplars for each category for use in the study.

the first time. As such, the comparison between participants who repeated the selection task or who try to recall the choices previously made can be equated with comparison between an implicit and an explicit memory test of the choices made in the first session after a 3 weeks interval. The amazing superiority of implicit over explicit memory tests over extended periods of time is one of the dissociative features of these two types of tests (E. R. Smith, 1990). For instance, Kolers (1976) showed that 1 year after some practice in reading inverted text, participants would read faster the inverted pages they had been presented with than new inverted text pages even though they were totally incapable of recognizing them.

These results thus argue against taking variation in participants' identification of stereotypic content across sessions as a function of either measurement error or deliberate changes in responding. On one hand, the equivalence main effect showed that variation in the stereotype attribution task reflected meaningful context sensitivity and not simply measurement error. On the other hand, the fact that participants trying to remember their first session choices were much less consistent than those repeating the stereotype attribute task casts severe doubt on the plausibility of alternative explanations that depend on participants' explicit memory of their original responses.

Impact of Centrality on Stability

As in Study 1, to explore the role of centrality on choice stability, we performed a median split on the centrality ratings of the traits chosen by each participant and computed separate overlap scores. We then computed a 2 (session) × 2 (social category) × 2 (Session 1 exemplar typicality) × 2 (exemplar typicality equivalence across sessions) × 2 (Session 2 task) × 2 (trait) mixed model ANOVA on the overlap scores. To avoid redundancy with the previous analysis, we report only effects involving centrality. Replicating Study 1, a main effect emerged for this factor, $F(1, 83) = 65.45$, $p = .000$, $MSE = .119$. Participants showed more overlap across sessions for central ($M = .58$) than for peripheral attributes ($M = .29$). Of more interest, two 2-way interactions qualified this main effect.

The first was a significant Equivalence × Centrality interaction, $F(1, 83) = 4.46$, $p = .038$, $MSE = .119$. This effect was due to the fact that exemplar typicality equivalence across sessions had a much greater impact on the stability of central attributes chosen across sessions ($M_{\text{equivalent}} = .66$ vs. $M_{\text{nonequivalent}} = .50$) than it did on the stability of peripheral traits ($M_{\text{equivalent}} = .29$ vs. $M_{\text{nonequivalent}} = .28$). This finding suggests that fluidity in central compared with peripheral traits may reflect different processes. Whereas fluidity in central trait choices may reflect context sensitivity rather than measurement error, variation in peripheral trait choices may reflect the opposite. Future research should address this possibility. Moreover, note that context sensitivity is not always an obstacle to stereotype stability. For instance, when the context remained constant, stereotype stability across sessions was very high ($M = .89$).

The second significant interaction was between second session task and centrality, $F(1, 83) = 12.88$, $p = .001$, $MSE = .119$. This effect was due to the fact that the stability of central traits was much higher ($M = .79$) than for peripheral traits ($M = .38$) for participants who repeated the stereotype attribute task than for participants trying to reproduce their first session choices ($M_{\text{central}} = .36$ vs. $M_{\text{peripheral}} = .21$). This effect attests to the critical differences between naturally performing the stereotype task and performing it by trying to explicitly remember previous responses. In turn, this suggests that any within-participant fluidity that occurs in reported stereotype content is unlikely to have been influenced by explicit memory for earlier reactions.

Subjective Stereotype Stability

Participants who performed the stereotype attribute task twice were also asked to estimate the number of traits chosen in the second session that they also chose in the first session. We used these estimates to compute "subjective" common-element correlations (subjective overlap scores). These subjective overlap scores could therefore be contrasted with "objective" overlap scores (on the basis of the actual chosen traits across sessions). To compare subjective and objective overlap scores, we computed a 2 (session) × 2 (social category) × 2 (Session 1 exemplar typicality) × 2 (exemplar typicality equivalence across sessions) × 2 (overlap) mixed model ANOVA. Again to avoid redundancy with previous analyses, we report only effects related to the overlap factor. A main effect emerged, $F(1, 60) = 7.63$, $p = .008$, $MSE = .031$, attesting to the fact that participants overestimated the degree of overlap ($M = .65$) in their responses across sessions relative to the actual degree of overlap ($M = .59$).[5] This effect may help to explain why sometimes our intuitions regarding stereotype consistency are much stronger than our data. Whereas we may be inconsistency avoiders at heart, our cognitive system may be able to accommodate more variability than we give it credit for. Note that these results echo those found in the perceived attitude stability domain. Participants who are experimentally induced to significantly change their attitudes on critical issues report little or no attitude change at all (Bem & McConnell, 1970; Goethals & Reckman, 1973; Ross & Shulman, 1973). It may be the case that the illusion of stability is not incompatible with highly context-sensitive knowledge structures—it may even be a requirement of such structures. When we perceive our knowledge as instable or variable, we perceive it as having little validity and cease regarding it as a valuable guide for behavior (Bem, 1972; Kelley, 1973).

General Discussion

The primary goal of these studies was to assess within-participant stability in the content and use of social category representations over time. Stability was assessed across sessions, 2 to 4 weeks apart, in a within-participant manner. We assumed natural variations in the idiosyncratic judgment contexts of our participants, such that particular moods, particular thoughts, recent experiences, and so forth would vary quite broadly across sessions. Such natural variability provided the appropriate conditions to test abstractionist intuitions concerning a high degree of stability in stereotyping against alternative positions.

The results of the first three studies reported here provide considerable evidence of instability in the representation of both

[5] A noninterpretable four-way interaction also emerged, $F(1, 60) = 4.94$, $p = .030$, $MSE = .015$.

personally and culturally held social categories. First, we found low within-participant stability in the selection of traits as stereotypic of a social category (Study 1), in the typicality ratings of category members (Study 2), and in category exemplars retrieved (Study 3). Second, the level of within-participant instability we obtained was only partially accounted for by feature or exemplar centrality or importance. Instability greatly decreased for more central or important trait attributes (Study 1) and for more typical group members (Studies 2 and 3). However, even for more central attributes and typical exemplars, considerable instability was still present. Third, although accessible attributes (Study 1) and accessible exemplars (Study 3) were more stable across time, even they showed considerable instability. Accessibility and centrality, and accessibility and typicality, were both moderately correlated.

The results of the fourth study are critical for three reasons. First, they provide data against the idea that fluidity in stereotype content across sessions is simply a byproduct of measurement error. Stereotype stability was a function of context constancy (particularly in the case of traits deemed central), suggesting that stereotype fluidity is likely to reflect sensitivity to the natural variation in the contexts in which stereotypes are constructed and used, rather than reflecting measurement error. Second, the results are equally discouraging for the notion that participants may consciously avoid making the same choices across sessions. The fact that our participants showed poor explicit memory for their first session choices suggests that explaining variation in stereotype construction as response monitoring or response editing is implausible. Finally, the results of Study 4 also suggest that within-participant stereotype consistency may be illusorily overestimated.

To end this section, a final caveat is in order. Note that although we found considerable evidence for within-participants stereotype instability, we also found considerable evidence for between-sessions stability (see Study 1). We in no way claim that the question of stereotype stability should be addressed only from an intraindividual perspective. Both intraindividual and between-sample perspectives are important and complementary. We simply argue that stereotype research has addressed the question of stereotype stability in a way that pays too little attention to the intraindividual side of the coin. With our research, we attempted to contribute a more balanced approach.

Dynamic Alternatives to Enduring Abstractionism

This pattern of results is clearly inconsistent with early abstractionist views—and perhaps with lay theories about stereotypes—but is quite consistent with more modern views of stereotype representation (exemplar, mixed models, or connectionist frameworks). In fact, despite the unique nature of social stereotypes, the general level of context sensitivity found in these studies is quite comparable with the results obtained from studies of common object concepts. Thus, although some of the content of the stereotypes studied here seemed relatively context independent, exhibiting greater accessibility and stability across time and context, a lot of our participants' knowledge about the stereotypes was quite fluid. In this final section, we suggest some mechanisms that need to be integrated into views of social category representation to better account for their apparent fluidity.

Partial Retrieval and/or Variable Activation

According to some exemplar approaches to categorization (e.g., Barsalou, Huttenlocher, & Lamberts, 1998; Kahneman & Miller, 1986; Nosofsky & Palmeri, 1997), knowledge structures are not represented by abstractions but instead by the whole class of stored exemplars. At any particular moment, and dependent on the demands of the cognitive task at hand, a subset of exemplars is retrieved so that a category judgment can be made or the processing of new exemplars facilitated. (Alternatively, exemplar retrieval may correspond to a continuous activation function, see Hintzman, 1986.) Context sensitivity is to be expected from such exemplar-based perspectives because category judgments and exemplar processing depend on the idiosyncratic nature of the small subset of exemplar retrieved. Models of attitudes as temporary constructs use a similar reasoning, contending that the stored information relevant to any given attitude is likely to be both extensive and contradictory and that at any given point in time, only a small portion of this information is retrieved and used to compute an attitude (Lord & Lepper, 1999; Wilson & Hodges, 1992). Partial retrieval thus accounts for stereotype instability. Although exemplar models would seem most compatible with partial retrieval, it is quite possible to envisage an abstraction-based model in which only part of the concept-relevant stored knowledge is retrieved depending on the situation (see Barsalou, 1999, 2002). Thus a partial retrieval temporary-abstractions model would also exhibit the required property of context sensitivity.

Situated Knowledge

If people learn about categories and concepts in a succession of episodes and if category knowledge becomes grounded in situational knowledge, then it makes sense to assume that different situations activate different categorical knowledge (Yeh & Barsalou, 1996, 2002). For stereotypes, it is quite plausible to predict that the kind of stereotypic knowledge that is activated by the presence of a given gay person or gypsy should be quite different depending on the context in which information was gathered. Some exemplar models possess such a plasticity-inducing mechanism (e.g., Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1987), but again, abstraction-based models could integrate it (Barsalou, in press).

Compound Retrieval Cues

The context dependent nature of categorical knowledge may derive from the compound nature of the retrieval cues that are used to retrieve such knowledge. According to global matching models (e.g., Hintzman, 1986), for example, self-assembled or context available cues are spontaneously integrated, forming compound retrieval cues that can be matched against stored memories for an output judgment (Dosher & Rosedale, 1989; McKoon & Ratcliff, 1992; Ratcliff & McKoon, 1988). Thus even if stereotypes are stable representations, the process of assembling or retrieving them in any given situation may be unreliable, in the sense that it is inherently dependent on the momentarily available situation (Medin & Ortony, 1989; E. R. Smith, 1989).

Source of Activation Confusion

When a stereotype is constructed or retrieved, activation is a consequence of matching (attributes or group members more fre-

quently associated with other group cues become more activated). However, contextually activated attributes may also affect stereotype construction because we are usually aware only of the consequences of memory trace activation and not of its source (Ayers & Reder, 1998; Reder & Schunn, 1996). Thus, information that is not usually part of a stereotype may inadvertently be incorporated into one because of its momentary salience.

Summary

All four of these proposed mechanisms describe processing mechanisms rather than representational formats. In that sense, they are compatible even with abstraction-based representations. And thus, it is quite possible to argue for the necessity of maintaining a role for abstractions in our conceptions of stereotypes but, at the same time, to acknowledge the situated and context sensitive nature of stereotypic representations. Like our data, these proposed cognitive mechanisms are incompatible, however, with assumptions of invariant retrieval of nonsituated stereotypic knowledge that dominated early social cognition-based research on stereotypes and that still heavily contaminate lay and intuitive theories of stereotypes. These four mechanisms are not necessarily incompatible and only further research can allow us to better articulate them and to choose among them.

Our data are much more consistent with the idea that stereotypes, despite being widely shared and highly resilient, are nevertheless temporary constructs that are assembled in a flexible and context sensitive way to meet situational requirements and the perceiver's goals. Across time and contexts, stereotypes may differ markedly, as a function of the variability of those times and contexts in which stereotypes are brought to mind. This view converges broadly with exemplar, mixed model, and connectionist perspectives on social categories, and also with recent positions taken in the cognitive literature (Barsalou, in press).

Our findings suggest that considerable flexibility and fluidity in stereotype knowledge is possible. If we are in fact so capable of cognitive sensitivity and flexibility, why does the social information processor seem so consistency prone, avoiding dissonance, inconstancy, and incongruence? It may be that our apparent cognitive stability derives not from impoverished inputs (abstractions) to our cognitive system but from impoverished inputs constrained by our social environments. The world is a big and diverse place, but we choose to live in encapsulated noneventful social worlds where like-minded people live out similar lives. It is therefore possible that we have mistaken social consistency for cognitive consistency. If so, such an acknowledgement allows for a new perspective on stereotypes, one that recognizes the situated nature of social cognition and conceives information processing as a socially distributed network process (Clancey, 1997; Higgins, 2000; Wegner, 1995).

References

Abelson, R. P. (1994). A personal perspective on social cognition. In P. G. Devine, D. L. Hamilton, & T. M. Ostrom, *Social cognition: Impact on social psychology* (pp. 15–39). San Diego, CA: Academic Press.

Armstrong, S. L., Gleitman, L. R., & Gleitman, H. (1983). On what some concepts might not be. *Cognition, 13,* 263–308.

Ashmore, R. D., & DelBoca, F. K. (1981). Conceptual approaches to stereotypes and stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 1–35). Hillsdale, NJ: Erlbaum.

Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review, 5,* 1–21.

Babey, S. H., Queller, S., & Klein, S. B. (1998). The role of expectancy violating behavior in the representation of trait knowledge: A summary-plus-exception model of social memory. *Social Cognition, 16,* 287–339.

Barsalou, L. W. (1982). Context-independent and context-dependent information in concepts. *Memory and Cognition, 10,* 82–93.

Barsalou, L. W. (1983). Ad hoc categories. *Memory and Cognition, 11,* 211–227.

Barsalou, L. W. (1985). Ideals, central tendency, and frequency of instantiation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 629–654.

Barsalou, L. W. (1987). The instability of graded structure: Implications for the nature of concepts. In U. Neisser (Ed.), *Concepts reconsidered: The ecological and intellectual bases of categories* (pp. 101–140). New York: Cambridge University Press.

Barsalou, L. W. (1989). Intra-concept similarity and its implications for inter-concept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 76–116). New York: Cambridge University Press.

Barsalou, L. W. (1999). Perceptual symbols systems. *Behavioral and Brain Sciences, 22,* 577–660.

Barsalou, L. W. (2002). Being there conceptually: Simulating categories in preparation for situated action. In N. L. Stein, P. J. Bauer, & M. Rabinowitz (Eds.), *Representation, memory, and development: Essays in honor of Jean Mandler* (pp. 1–16). Mahwah, NJ: Erlbaum.

Barsalou, L. W. (in press). Situated simulation in the human conceptual system. *Language and Cognitive Processes*.

Barsalou, L. W., Huttenlocher, J., & Lamberts, K. (1998). Basing categorization on individuals and events. *Cognitive Psychology, 36,* 203–272.

Barsalou, L. W., & Medin, D. L. (1986). Concepts: Static definitions or context-dependent representations? *Cahiers de Psychologie Cognitive, 6,* 187–202.

Barsalou, L. W., & Sewell, D. R. (1984). Constructing representations of categories from different points of view. *Emory Cognition Technical Report #2* (pp. 1–66), Emory University, Atlanta, GA.

Barsalou, L. W., Sewell, D. R., & Ballato, S. M. (1986). Assessing the stability of category representations with graded structure. Unpublished manuscript, Emory University, Atlanta, GA.

Barsalou, L. W., Spindler, J. L., Sewell, D. R., Ballato, S. M., & Gendel, E. M. (1987). Assessing the stability of category representations with property generation. Unpublished manuscript, Emory University, Atlanta, GA.

Bellezza, F. S. (1984a). Reliability of retrieval from semantic memory: Common categories. *Bulletin of the Psychonomic Society, 22,* 324–326.

Bellezza, F. S. (1984b). Reliability of retrieval from semantic memory: Noun meanings. *Bulletin of the Psychonomic Society, 22,* 377–380.

Bellezza, F. S. (1984c). Reliability of retrieval from semantic memory: Information about people. *Bulletin of the Psychonomic Society, 22,* 511–513.

Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1–62). New York: Academic Press.

Bem, D. J., & McConnell, H. K. (1970). Testing the self-perception explanation of dissonance phenomena: On the salience of premanipulated attitudes. *Journal of Personality and Social Psychology, 14,* 23–31.

Bodenhausen, G. V., Schwarz, N., Bless, H., & Wanke, M. (1995). Effects of atypical exemplars in racial beliefs: Enlightened racism or generalized appraisals? *Journal of Experimental Social Psychology, 31,* 48–63.

Brewer, M. B. (1988). A dual process model of impression formation. In

R. Wyer & T. Srull (Eds.), *Advances in social cognition* (Vol. 1, pp. 1–36). Hillsdale, NJ: Erlbaum.

Clancey, W. J. (1997). *Situated cognition: On human knowledge and computer representations.* New York: Cambridge University Press.

Crocker, J., Fiske, S. T., & Taylor, S. E. (1984). Schematic bases of belief change. In J. R. Eiser (Ed.), *Attitudinal judgment* (pp. 197–226). New York: Springer-Verlag.

Devine, P. G., & Elliot, A. J. (1995). Are racial stereotypes *really* fading? The Princeton trilogy revisited. *Personality and Social Psychological Bulletin, 21,* 1139–1150.

Dosher, B., & Rosedale, G. (1989). Integrating retrieval cues as a mechanism for priming in retrieval from memory. *Journal of Experimental Psychology: General, 16,* 519–555.

Dovidio, J. F., Brigham, J. C., Johnson, B. T., & Gaertner, S. L. (1996). Stereotyping, prejudice and discrimination: Another look. In C. N. Macrae, C. Stangor, & M. Hewstone (Eds.), *Stereotypes and stereotyping* (pp. 276–319). New York: Guilford Press.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology, 38,* 889–906.

Fiske, S. T., & Neuberg, S. L. (1990). A continuum of impression formation from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). San Diego, CA: Academic Press.

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition.* Reading, MA: Addison-Wesley.

Garcia-Marques, L., & Mackie, D. M. (1999). The impact of stereotype-incongruent information on perceived group variability and stereotype change. *Journal of Personality and Social Psychology, 77,* 979–990.

Gilbert, G. M. (1951). Stereotype persistence and change among college students. *Journal of Abnormal and Social Psychology, 46,* 245–254.

Goethals, G. R., & Reckman, R. F. (1973). The perception of consistency in attitudes. *Journal of Experimental Social Psychology, 9,* 491–501.

Hamilton, D. L., & Sherman, J. W. (1994). Stereotypes. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., Vol. 2, pp. 1–68). Hillsdale, NJ: Erlbaum.

Hamilton, D. L., & Trolier, T. K. (1986). Stereotypes and stereotyping: An overview of the cognitive approach. In J. Dovidio, & S. L. Gaertner (Eds.), *Prejudice, discrimination, and racism* (pp. 127–163). New York: Academic Press.

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review, 93,* 258–268.

Higgins, E. T. (2000). Social cognition: Learning about what matters in the social world. *European Journal of Social Psychology, 30,* 3–39.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Judd, C. M., & Park, B. (1988). Out-group homogeneity: Judgments of variability at the individual and group levels. *Journal of Personality and Social Psychology, 54,* 778–788.

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153.

Karlins, M., Coffman, T. L., & Walters, G. (1969). On the fading of social stereotypes: Studies in three generations of college students. *Journal of Personality and Social Psychology, 13,* 1–16.

Kashima, Y., Woolcock, J., & King, D. (1998). The dynamics of group impression formation: The tensor product model of exemplar-based social category learning. In S. Y. Read & L. C. Miller (Eds.), *Connectionist models of social reasoning and social behavior* (pp. 71–110). Mahwah, NJ: Erlbaum.

Katz, D., & Braly, K. W. (1933). Racial stereotypes of 100 college students. *Journal of Abnormal Social Psychology, 28,* 280–290.

Katz, D., & Braly, K. W. (1935). Racial prejudice and racial stereotypes. *Journal of Abnormal Social Psychology, 30,* 175–193.

Kelley, H. H. (1973). The processes of causal attribution. *American Psychologist, 28,* 107–128.

Kolers, P. A. (1976). Reading a year later. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 554–565.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44.

Linville, P. W., & Fischer, G. W. (1993). Exemplar and abstraction models of perceived group variability and stereotypicality. *Social Cognition, 11,* 92–125.

Linville, P. W., Fischer, G. W., & Salovey, P. (1989). Perceived distributions of characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology, 57,* 165–188.

Lord, C. G., & Lepper, M. R. (1999). Attitude representation theory. In M. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 31, pp. 265–343). New York: Academic Press.

MacArthur, L. Z. (1982). Judging a book by its cover: A cognitive analysis of the relationship between physical appearance and stereotyping. In A. H. Hastorf & A. M. Isen (Eds.), *Cognitive social psychology* (pp. 149–212). New York: Elsevier/Elsevier Science.

Maurer, K. L., Park, B., & Rothbart, M. (1995). Subtyping versus subgrouping processes in stereotype representation. *Journal of Personality and Social Psychology, 69,* 812–824.

McCloskey, M. E., & Glucksberg, S. (1978). Natural categories: Well defined or fuzzy sets*? Memory and Cognition, 6,* 462–472.

McGuire, W. L. (1983). A contextualist theory of knowledge: Its implications for innovation and reform in psychological research. *Advances in Experimental Social Psychology, 16,* 1–47.

McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 1155–1172.

McNemar, Q. (1969). *Psychological statistics* (4th ed.). New York: Wiley.

Medin, D. L., & Ortony, A. (1989). Psychological essentialism. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 179–196). New York: Cambridge University Press.

Medin, D. L., & Schaffer, M. M. (1978). A context theory of classification learning. *Psychological Review, 85,* 207–238.

Medin, D. L., & Smith, E. E. (1984). Concepts and concept formation. *Annual Review of Psychology, 35,* 113–138.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92,* 289–316.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13,* 87–109.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review, 104,* 266–300.

Osherson, D. N., & Smith, E. E. (1981). On the adequacy of prototype theory as a theory of concepts. *Cognition, 9,* 35–58.

Queller, S., & Smith, E. (2002). Subtyping versus bookkeeping in stereotype learning and change: Connectionist simulations and empirical findings. *Journal of Personality and Social Psychology, 82,* 300–313.

Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming. *Psychological Review, 95,* 385–408.

Reder, L. M., & Schunn, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In L. M. Reder (Ed.), *Implicit memory and metacognition* (pp. 45–77). Mahwah, NJ: Erlbaum.

Rosch, E. H., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605.

Ross, M., & Conway, M. (1986). Remembering one's own past: The construction of personal histories. In R. M. Sorrentino & E. T. Higgins

(Eds.), *The handbook of motivation and cognition: Foundations of social behavior* (122–144). New York: Guilford Press.

Ross, M., & Shulman, R. F. (1973). Increasing the salience of initial attitudes: Dissonance versus self-perception theory. *Journal of Personality and Social Psychology, 28,* 138–144.

Roth, E. M., & Shoben, E. J. (1983). The effects of context on the structure of categories. *Cognitive Psychology, 15,* 346–378.

Rothbart, M., & John, O. P. (1985). Social categorization and behavior episodes: A cognitive analysis of the effects of intergroup contact. *Journal of Social Issues, 41,* 81–104.

Rothbart, M., & John, O. P. (1993). Intergroup relations and stereotype change: A social-cognitive analysis and some longitudinal findings. In P. M. Sniderman, P. E. Tetlock, & E. G. Carmines (Eds.), *Prejudice, politics and the American dilemma* (pp. 32–59). Stanford, CA: Stanford University Press.

Rothbart, M., & Lewis, S. (1988). Inferring category attributes from exemplar attributes: Geometric shapes and social categories. *Journal of Personality and Social Psychology, 55,* 861–872.

Rothbart, M., Sriram, N., & Davis-Stitt, C. (1996). The retrieval of typical and atypical category members. *Journal of Experimental Social Psychology, 32,* 309–336.

Santos, A. S. C. (2001). *A instabilidade dos estereótipos: Experiências em cognição social com uma metodologia de teste-reteste longitudinal.* [*Stereotype instability: Experiments in social cognition with a test–retest methodology*]. Unpublished master's thesis, Universidade de Lisboa, Lisboa, Portugal.

Schmidt, F. L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods, 1,* 115–129.

Sherman, J. W. (1996). Development and mental representation of stereotypes. *Journal of Personality and Social Psychology, 70,* 1126–1142.

Sia, T. L., Lord, C. G., Blessum, K. A., Thomas, J. C., & Lepper, M. R. (1999). Activation of exemplars in the process of assessing social category attitudes. *Journal of Personality and Social Psychology, 76,* 517–532.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard University Press.

Smith, E. R. (1988). Category accessibility effects in a simulated exemplar-based memory. *Journal of Experimental Social Psychology, 24,* 448–463.

Smith, E. R. (1989). Procedural efficiency: General and specific components and effects on social judgment. *Journal of Experimental Social Psychology, 25,* 500–523.

Smith, E. R. (1990). Content and process specificity in the effects of prior experiences. In T. K. Srull & R. S. Wyer, Jr. (Eds.), *Advances in social cognition* (Vol. 3, pp. 1–59). Hillsdale, NJ: Erlbaum.

Smith, E. R. (1998). Mental representations and memory. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (4th ed., Vol. 1, pp. 391–445). New York: McGraw-Hill.

Smith, E. R., & DeCoster, J. (1998). Knowledge acquisition, accessibility, and use in person perception and stereotyping: Simulation with a recurrent connectionist network. *Journal of Personality and Social Psychology, 74,* 21–35.

Smith, E. R., & Mackie, D. M. (1995). *Social psychology.* New York: Worth Publishers.

Smith, E. R., & Zárate, M. A. (1992). Exemplar-based model of social judgments. *Psychological Review, 99,* 3–21.

Snyder, M. (1981). On the self-perpetuating nature of social stereotypes. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 183–212). Hillsdale, NJ: Erlbaum.

Taylor, S. E. (1981). A categorization approach to stereotyping. In D. L. Hamilton (Ed.), *Cognitive processes in stereotyping and intergroup behavior* (pp. 83–114). Hillsdale, NJ: Erlbaum.

Van Rooy, D., Van Overwalle, F., Vanhoomissen, T., Labiouse, C., & French, R. (2003). A recurrent connectionist model of group biases. *Psychological Review, 3,* 536–563.

Wegner, D. M. (1995). A computer network model of human transactive memory. *Social Cognition, 13,* 319–339.

Wilson, T. D., & Hodges, S. D. (1992). Attitudes as temporary constructions. In L. L. Martin & A. Tesser (Eds.), *The construction of social judgments* (pp. 37–65). Hillsdale, NJ: Erlbaum.

Yeh, W., & Barsalou, L. W. (1996). The role of situation in concept learning. In G. W. Cottrell (Ed.), *Proceedings of the eighteenth annual conference of the cognitive science society* (pp. 469–474). Hillsdale, NJ: Erlbaum.

Yeh, W., & Barsalou, L. W. (2002). *The situated nature of concepts.* Unpublished manuscript, Emory University, Atlanta, GA.

Zárate, M. A., & Smith, E. R. (1990). Person categorization and stereotyping. *Social Cognition, 8,* 161–185.

Appendix A

Percentage of Choices of the Five Traits That Best Describe the Gypsy Group (Katz–Braly's (1933) Task, Study 1)

Descriptor	Cultural stereotype		Descriptor	Individual stereotype	
Dishonest	78.30	82.60	**Distrustful**	65.20	54.30
Trouble-maker	76.00	82.60	**Trouble-maker**	60.10	41.30
Distrustful	54.30	39.10	**Happy**	43.40	30.40
Ill-mannered	45.70	34.80	**Superstitious**	34.80	34.80
Greedy	28.30	13.00	Emotional	30.40	34.80
Happy	28.30	19.60	**Ill-mannered**	23.90	19.60
Ignorant	26.10	32.60	**Dishonest**	21.70	6.50
Superstitious	26.10	26.10	Hardworking	21.70	19.60
Disrespectful	23.90	30.40	Conceited	21.70	30.40
Show-off	19.60	19.60	**Ignorant**	19.60	19.60
Poor	15.20	8.70	Shy	19.60	28.30
Emotional	13.00	8.70	Strong	13.00	19.60
Strong	10.90	8.70	Dynamic	13.00	13.00
Ugly	8.70	6.50	**Greedy**	10.90	6.50
Rich	6.50	6.50	**Disrespectful**	10.90	6.50
Hardworking	6.50	10.90	Rich	10.90	6.50
Conceited	6.50	10.90	Poor	8.70	13.00
Lazy	6.50	8.70	Aloof	6.50	4.30
Aloof	4.30	4.30	Friendly	6.50	15.20
Unintelligent	4.30	4.30	Faithful	6.50	6.50
Shy	4.30	4.30	Show-off	4.30	21.70
Friendly	4.30	2.20	Sad	4.30	4.30
Sad	2.20	2.20	Vulgar	4.30	10.90
Naive	2.20	0.00	Honest	4.30	4.30
Vulgar	2.20	8.70	Unfaithful	4.30	2.20
Passive	2.20	2.20	Sensitive	4.30	4.30
Faithful	2.20	0.00	Ugly	2.20	0.00
Insensitive	2.20	0.00	Lazy	2.20	2.20
Honest	0.00	2.20	Unintelligent	2.20	0.00
Cultured	0.00	0.00	Naive	2.20	0.00
Attractive	0.00	0.00	Passive	2.20	0.00
Peaceful	0.00	0.00	Cultured	2.20	4.30
Dynamic	0.00	2.20	Peaceful	2.20	0.00
Unfaithful	0.00	2.20	Fragile	2.20	2.20
Sensitive	0.00	0.00	Generous	2.20	0.00
Discreet	0.00	0.00	Intelligent	2.20	4.30
Fragile	0.00	0.00	Insensitive	0.00	0.00
Unpretentious	0.00	0.00	Attractive	0.00	2.20
Polite	0.00	0.00	Discreet	0.00	0.00
Generous	0.00	0.00	Unpretentious	0.00	0.00
Sophisticated	0.00	0.00	Polite	0.00	0.00
Respectful	0.00	0.00	Sophisticated	0.00	2.20
Intelligent	0.00	0.00	Respectful	0.00	0.00
	$r_{\text{within-items}} = .97$			$r_{\text{within-items}} = .91$	

Note. Items in bold represent the nine attributes most frequently mentioned by participants.

(*Appendixes continue*)

Appendix B

Percentage of Choices of the Five Traits That Best Describe the Gay Group (Katz–Braly's (1933) Task, Study 1)

Descriptor	Cultural stereotype		Descriptor	Individual stereotype	
Show-off	58.70	60.90	**Sensitive**	76.10	65.20
Polite	47.80	39.10	**Emotional**	41.30	47.80
Disrespectful	43.50	45.70	**Polite**	41.30	43.50
Conceited	37.00	32.60	Peaceful	34.80	39.10
Emotional	34.80	30.40	Friendly	30.40	21.70
Unfaithful	30.40	21.70	**Fragile**	28.30	10.90
Sensitive	30.40	41.30	Honest	26.10	10.90
Fragile	26.10	32.60	Shy	19.60	37.00
Vulgar	19.60	8.70	Discreet	19.60	21.70
Peaceful	19.60	17.40	**Show-off**	15.20	15.20
Shy	17.40	15.20	Respectful	15.20	13.00
Unintelligent	15.20	13.00	Happy	13.00	6.50
Friendly	13.00	8.70	Sad	13.00	8.70
Insensitive	8.70	6.50	**Sophisticated**	13.00	6.50
Sophisticated	8.70	13.00	Intelligent	13.00	6.50
Dishonest	6.50	6.50	**Conceited**	10.90	15.20
Ignorant	6.50	4.30	Cultured	10.90	15.20
Sad	6.50	2.20	Attractive	10.90	8.70
Happy	4.30	6.50	Generous	10.90	10.90
Aloof	4.30	4.30	Strong	8.70	2.20
Naive	4.30	10.90	Faithful	6.50	17.40
Discreet	4.30	6.50	Unpretentious	6.50	4.30
Respectful	4.30	2.20	Hardworking	4.30	4.30
Trouble-maker	2.20	4.30	**Unfaithful**	4.30	2.20
Ill-mannered	2.20	0.00	Trouble-maker	2.20	0.00
Superstitious	2.20	0.00	Distrustful	2.20	6.50
Ugly	2.20	6.50	**Disrespectful**	2.20	4.30
Hardworking	2.20	2.20	Vulgar	2.20	6.50
Passive	2.20	4.30	Dishonest	0.00	0.00
Faithful	2.20	4.30	Ill-mannered	0.00	0.00
Honest	2.20	2.20	Greedy	0.00	2.20
Cultured	2.20	2.20	Ignorant	0.00	0.00
Attractive	2.20	2.20	Superstitious	0.00	0.00
Unpretentious	2.20	4.30	Poor	0.00	0.00
Distrustful	0.00	4.30	Ugly	0.00	0.00
Greedy	0.00	0.00	Rich	0.00	0.00
Poor	0.00	0.00	Lazy	0.00	0.00
Strong	0.00	0.00	Aloof	0.00	2.20
Rich	0.00	2.20	Unintelligent	0.00	0.00
Lazy	0.00	2.20	Naive	0.00	2.20
Dynamic	0.00	0.00	Passive	0.00	0.00
Generous	0.00	0.00	Insensitive	0.00	2.20
Intelligent	0.00	2.20	Dynamic	0.00	0.00
	$r_{\text{within-items}} = .96$			$r_{\text{within-items}} = .92$	

Note. Items in bold represent the nine attributes most frequently mentioned by participants.

Appendix C

Percentage of Choices of the Five Traits That Best Describe the African Immigrant Group (Katz–Braly's (1933) Task, Study 1)

Descriptor	Cultural stereotype		Descriptor	Individual stereotype	
Ignorant	63.00	52.20	**Hardworking**	56.50	56.50
Poor	56.50	45.70	**Poor**	52.20	50.00
Trouble-maker	39.10	41.30	**Strong**	32.60	32.60
Hardworking	37.00	41.30	Happy	28.30	21.70
Unintelligent	28.30	32.60	**Ignorant**	28.30	23.90
Passive	26.10	28.30	Friendly	23.90	21.70
Ill-mannered	21.70	15.20	Respectful	23.90	6.50
Vulgar	21.70	21.70	Distrustful	19.60	17.40
Dishonest	19.60	21.70	Sad	19.60	10.90
Disrespectful	19.60	19.60	**Passive**	19.60	6.50
Strong	19.60	32.60	Honest	17.40	13.00
Ugly	19.60	19.60	Peaceful	17.40	17.40
Lazy	17.40	19.60	Shy	13.00	15.20
Naive	15.20	13.00	Naive	13.00	13.00
Distrustful	10.90	23.90	**Trouble-maker**	10.90	17.40
Superstitious	10.90	6.50	**Ill-mannered**	10.90	10.90
Shy	10.90	6.50	**Vulgar**	10.90	10.90
Happy	8.70	2.20	Show-off	8.70	6.50
Sad	8.70	6.50	Discreet	8.70	10.90
Peaceful	8.70	2.20	Unpretentious	8.70	13.00
Aloof	6.50	0.00	Generous	8.70	8.70
Friendly	6.50	4.30	Dynamic	6.50	10.90
Honest	6.50	4.30	Disrespectful	4.30	6.50
Discreet	4.30	2.20	Emotional	4.30	17.40
Greedy	2.20	2.20	Conceited	4.30	8.70
Emotional	2.20	2.20	Lazy	4.30	4.30
Conceited	2.20	0.00	Aloof	4.30	0.00
Unpretentious	2.20	4.30	Faithful	4.30	0.00
Respectful	2.20	4.30	Sensitive	4.30	8.70
Intelligent	2.20	0.00	Fragile	4.30	4.30
Show-off	0.00	10.90	Greedy	2.20	0.00
Rich	0.00	0.00	Superstitious	2.20	10.90
Faithful	0.00	0.00	Ugly	2.20	4.30
Insensitive	0.00	0.00	Insensitive	2.20	2.20
Cultured	0.00	0.00	Intelligent	2.20	4.30
Attractive	0.00	0.00	Dishonest	0.00	2.20
Dynamic	0.00	0.00	Rich	0.00	0.00
Unfaithful	0.00	0.00	**Unintelligent**	0.00	0.00
Sensitive	0.00	0.00	Cultured	0.00	0.00
Fragile	0.00	0.00	Attractive	0.00	0.00
Polite	0.00	0.00	Unfaithful	0.00	0.00
Generous	0.00	0.00	Polite	0.00	2.20
Sophisticated	0.00	0.00	Sophisticated	0.00	0.00
	$r_{\text{within-items}} = .94$			$r_{\text{within-items}} = .92$	

Note. Items in bold represent the nine attributes most frequently mentioned by participants.

Received January 26, 2005
Revision received February 7, 2006
Accepted February 7, 2006 ■