CERTIFICATION OF APPROVAL

I certify that I have read Who are you? Categorization and social repercussions by Jordan Seliger, and that in my opinion this work meets the criteria for approving a thesis submitted in partial fulfillment of the requirement for the degree Master of Arts in Psychology: Mind, Brain, and Behavior at San Francisco State University.

Avi Ben-Zeev, Ph.D.
Professor of Psychology

Mark W. Geisler, Ph.D.
Professor of Psychology
The current research was designed to examine the influence of racial-phenotypic information on social categorization. Classic studies on essentialist-based categorization have shown that people tend to believe that group membership is not based solely on physical appearance, but by an alleged underlying essence (Medin & Ortony, 1989; Rips, 1989). However, Hampton, Estes, and Simmons (2007) have proposed a causal homeostasis account, which suggests that if one’s outward appearance has changed sufficiently to look like a different category, people will believe that the internal makeup (or essence) has changed as well. The current research directly tests these models of categorization on social categories, as well as examines potential social repercussions of crossing seemingly immutable category boundaries.
ACKNOWLEDGEMENT

I would like to acknowledge the Complex Cognition Laboratory and the Cognitive Psychophysiology Laboratory, to whom I am indebted to for the culmination of this project. I would also like to thank my wife, Emily for supporting me through this endeavor.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>List of Appendices</td>
<td>xiii</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Pilot Study</td>
<td>7</td>
</tr>
<tr>
<td>Method</td>
<td>8</td>
</tr>
<tr>
<td>Results</td>
<td>9</td>
</tr>
<tr>
<td>Study 1</td>
<td>10</td>
</tr>
<tr>
<td>Method</td>
<td>12</td>
</tr>
<tr>
<td>Results</td>
<td>16</td>
</tr>
<tr>
<td>Discussion</td>
<td>20</td>
</tr>
<tr>
<td>Study 2</td>
<td>22</td>
</tr>
<tr>
<td>Method</td>
<td>24</td>
</tr>
<tr>
<td>Results</td>
<td>26</td>
</tr>
<tr>
<td>Discussion</td>
<td>29</td>
</tr>
<tr>
<td>General Discussion</td>
<td>29</td>
</tr>
<tr>
<td>References</td>
<td>33</td>
</tr>
</tbody>
</table>
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Categorization results from Study 1</td>
<td>37</td>
</tr>
<tr>
<td>2. Uniquely human trait ratings from Study 1</td>
<td>38</td>
</tr>
<tr>
<td>3. Human nature trait ratings from Study 1</td>
<td>39</td>
</tr>
<tr>
<td>4. Uniquely human trait ratings from Study 2</td>
<td>40</td>
</tr>
<tr>
<td>5. Planned contrasts on uniquely human trait ratings from Study 2</td>
<td>41</td>
</tr>
<tr>
<td>6. Human nature trait ratings from Study 2</td>
<td>42</td>
</tr>
</tbody>
</table>
LIST OF APPENDICES

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Appendix A: Stimulus Images</td>
<td>43</td>
</tr>
<tr>
<td>2. Appendix B: Stimulus Vignettes</td>
<td>44</td>
</tr>
</tbody>
</table>
Introduction

Recently, Julie Chen, host of the popular television show, The Talk, revealed that she has been pressured to undergo a double eyelid surgery (see Kaw, 1991), a form of ‘ethnic cosmetic surgery’ (Davis, 2003), which involves widening the shape of the eyes, resulting in a more Caucasian appearance. For Julie Chen and many others, the explicit rationale is not to pass as White but to achieve a more Western/European ideals of beauty. De facto, these surgeries enable individuals to pass as belonging to higher status ethnic groups while reducing signs of otherness (Davis, 2003; Haiken, 1997).

Race is a highly essentialized construct (Haslam, Rothschild, & Ernst, 2000), meaning people tend to view group membership in so-called natural categories (i.e., perceived as biological rather than social in origin, such as race) as fixed due to an alleged essence, termed *origin essentialism*. For example, people tend to believe that a White person cannot become Asian and vice versa. In the case of the Asian category, eye shape is a highly diagnostic feature (Brown, Dane, & Durham, 1998); therefore, it is unclear whether physical passing affects racial categorization. To what extent does appearing more white as a result of ethnic cosmetic surgery affect similarity versus categorization. Do those who have known Julie Chen to be Asian now categorize her as White? Do people base their racial categorization primarily on observable features, or rather on a supposition of a deep, unchanging property, independent of physical appearance,
responsible for category membership? In fact, the cognitive literature has observed dissociations between similarity and categorization (see Rips, 1989). However, there is another camp in the cognitive literature that suggests that if one’s appearance “passes” for a new category, then underlying causal properties responsible for category membership will change as well (see Hampton, Estes, & Simmons, 2007). Racial cosmetic surgery raises questions about the nature of social categories, such as race, and their relationship to physical features, which tend to be diagnostic of racial categories. For instance, we ask whether physical passing affects racial categorization. A cosmetic change to a diagnostic racial feature person’s appearance changes to the extent that she/he “passes” as White.

While it seems likely that people would base categorization on physical appearance alone, previous research examining people’s beliefs about social categories suggests the opposite. That is, people tend to believe that racial categories are highly natural and stable (Haslam, Rothschild, & Ernst, 2000). The naturalness and stability of race relies on people’s lay beliefs of a deep, underlying property, or essence, responsible for category membership (Medin & Ortony, 1989). Essentialism, differs from similarity-based theories (e.g., Rosch & Mervis, 1975), in that category membership is not necessarily dependent on surface features, but rather an unobservable, causal characteristic shared among members of a specific group. As mentioned above, a major tenant of psychological essentialism is that category membership is immutable. Thus,
belonging to a category that is highly essentialized, such as race, cannot be changed, regardless of altering one's appearance.

While the categorization literature has established that appearance can change and result in a match to a new category as a form of maturation, this does not fit the way people conceptualize race. That is, because racial categories are seen as immutable, a changing one's appearance to look like another race may be seen as unnatural and, thus, a form of mutation. Violating the status quo in this manner may also result in backlash, not unlike that observed when women violate gender-based hierarchies (Rudman, Moss-Racusin, Phelan, & Nauts, 2012).

Rips (1989) provided support for essentialist theorizing by demonstrating that categorization was not solely a function of similarity to a prototype or group of category members. Specifically, Rips showed that after reading about an insect-like creature that had accidently been transformed to look like a bird-like creature, people believed that the creature was still an insect-like creature despite being more similar to a bird-like creature. The main finding that after an accidental mutation, people dissociated similarity ratings from categorization, suggests that people are not basing their categorizations on appearance alone, but rather on an unobservable essence. That is, looking and acting more like a bird is not sufficient for reclassification because the causal essence remained unchanged. This type of categorization, termed origin essentialism, involves basing
categorization primarily on a creature's pre-transformation state, rather than its post-transformation appearance.

In contrast to the essentialist account of categorization, Hampton, Estes, and Simmons (2007) proposed causal homeostasis as an alternative. Causal homeostasis states that there are causal links between properties and, when making category judgments, causal features are weighted more heavily. One way Hampton, Estes, and Simmons demonstrated this effect by providing participants with vignettes similar to those Rips used. Presumably, if people believe that a causal feature, such as an essence placeholder like genetic information, remained intact, category membership would not change. To test this, participants in one condition, read about a transformation in which the mutated creatures still had viable offspring, while in another condition, information about offspring was omitted. In the former, Hampton et al. replicated Rips' findings; people believed that the causal feature (i.e., genetic information) was unchanged, category membership was unchanged as well. In contrast, when genetic information was omitted, people were more likely to base categorization on surface features. That is, if outward appearance has changed to the extent that one "passes" for the new category, then people will believe that the deep, causal properties responsible for category membership have changed as well, resulting in reclassification.

The categorization literature has established these competing theories, but they have not been examined in the social cognitive literature. It is an open
question how these theories would apply to social categories especially considering the conflict between a seemingly unnatural transformation and the desire to look White. In the case of Julie Chen, people may believe that she is still Asian because she was born Asian, in line with the origin essentialism hypothesis. In contrast, people may believe that she is White because she has changed sufficiently to “pass” for her new category, in line with the causal homeostasis hypothesis.

As mentioned above, the competing hypotheses have only been tested on natural kinds, such as birds and insects, and not on social categories. Applying the previously used paradigms to humans raises interesting questions about the volition of the transformation. Specifically, attempting to move from one racial category to another may be seen as unnatural because racial groups tend to be viewed as discrete and immutable (Haslam, Rothschild, & Ernst, 2000). As a result, an individual who undergoes a transformation to looks like a member of a different race may lead to uncertainty of the person’s essence and, thus, negative social repercussions. Additionally, if the transformation in voluntary, we hypothesize that the backlash incurred will be exaggerated because, not only is the person violating the natural order, but also the person is doing so willfully. Although the context is rather different, Rudman, Moss-Racusin, Phelan, and Nauts (2012) showed similar backlash effects for women who violate the status quo suffer from backlash as a way to maintain gender-based hierarchies. Their
findings, as well as our predictions, are well situated in system justification theory (Jost & Banaji, 1994), which maintains that people are motivated to maintain the status quo, which includes beliefs that races are highly natural, immutable categories. Specifically, we predict that people may engage in a subtle form of dehumanization, known as infrahumanization, when viewing an individual who has crossed a category boundary.

Infrahumanization involves the denial of uniquely human traits that differentiate humans from animals and is observed when people make assessments of outgroups (Haslam, 2006; Leyens, Demoulin, Vaes, Gaunt, & Paladino, 2007; Leyens et al., 2000). Leyens et al. (2001) argue that because differences in social categories, such as race, tend to be explained by differential attribution of essences, the denial of a human essence may explain infrahumanization between groups. However, the features that are fundamentally human, known as human nature traits, are also thought to be highly essentialized, and the denial of these traits result in a different form of dehumanization (Haslam, 2006). While subtle dehumanization tends to be thought of the denial of uniquely human traits, different social contexts have also elicited the denial of human nature traits in a manner similar to previous infrahumanization research (Bain, Park, Kwok, & Haslam, 2009). In order to capture the broad scope of the possible forms of dehumanization, the current research examined the possible denial of
both uniquely human and human nature traits of an individual who blurs the lines of racial categories.

The current research has two main foci: (a) To test the competing categorization hypotheses (i.e., the origin essentialism hypothesis and the causal homeostasis hypothesis) adapting transformation paradigms to social categories (Study 1); (b) To examine whether voluntarily undergoing racial cosmetic surgery would result in more negative social repercussions in the form of infrahumanization (Studies 1 and 2). Specifically, Study 1 presented people with images and a vignette that described either an Asian woman who changed her appearance to be White, or a White woman who changed her appearance to be Asian. Participant categorized the woman, rated what category her appearance was most similar to, and completed a measure of infrahumanization. Study 2 sought to test whether any infrahumanization effects observed in Study 1 would be present in any cosmetic surgery, or if the backlash is reserved for violating social category boundaries.

Pilot Study

The purpose of the pilot study was to ensure that participants in the population from which we would be sampling from in Study 1 agreed on the categorization and similarity ratings of the target images to be used in Study 1. That is, there are no systematic differences between the category the target images belong to and what category they are most physically similar to. Additionally, we
wanted to confirm that there are no differences on these measures between the Asian target and the White target.

**Method**

**Participants.** Eighty participants were recruited from Amazon’s Mechanical Turk. Thirty-five identified as Asian and 45 identified as White on a self-report measure.

**Materials.** Stimulus faces were created using FACES software with the only difference between the images being the shape of the eyes (IQ Biometrix, Inc.). Target images are available in Appendix A.

Categorization and similarity ratings were obtained by using measures adapted from Rips (1989). Categorization was measured by asking participants “How would you categorize this woman’s race?” Participants replied using a four-point scale (1 = Asian; 2 = More Asian than White; 3 = More White than Asian; 4 = White). Similarity was measure by asking participants “What race would you physically describe this woman as being most similar to?” Participants replied using the same four-point scale (1 = Asian; 2 = More Asian than White; 3 = More White than Asian; 4 = White).

**Procedure.** Participants were randomly assigned to view one of the stimulus images (either an Asian or a White woman) and rated her on the categorization and the similarity measure described above. The order of the categorization and similarity items as well as the order of the points on the scales was
counterbalanced across participants. To make direct comparisons between the
different stimuli, the values of the ratings were recoded such that lower values
indicate ratings that correspond to the target’s intended race. For example, the
White target rated as “White” would be a value of “1” as would the Asian target
being rated as “Asian.”

Results

These ratings were submitted to a 2 (participant ethnicity: Asian vs.
White) x 2 (target race: Asian vs. White) x 2 (rating type: categorization vs.
similarity) mixed factorial ANOVA with the last factor measured within-subjects.
Most importantly, there was not a significant main effect of rating type, \( F(1, 76) =
1.618, p = .207, \eta^2_p = .021 \), suggesting that participants’ categorization ratings (\( M
= 1.741, SE = .094 \)) did not differ from their similarity ratings (\( M = 1.821, SE =
.089 \)). Additionally, there was no significant main effect of target race, \( F(1, 76) =
.516, p = .475, \eta^2_p = .007 \), indicating that participants rated the Asian target (\( M =
1.843, SE = .120 \)) and the White target (\( M = 1.719, SE = .124 \)) similarly. Finally,
there was not a significant rating type by target race interaction, \( F(1, 76) = .005, p
= .994, \eta^2_p < .001 \). There was, however, a significant participant ethnicity by
target race interaction, \( F(1, 76) = 9.616, p = .003, \eta^2_p = .112 \), indicating that Asian
participants rated the Asian target as closer to the intended category (\( M = 1.526,
SE = .175 \)) than they did the White target (\( M = 1.938, SE = .191 \)); similarly, White
participants rated the White target as closer to the intended category (\( M = 1.500,
SE = .159) than they did the Asian target (M = 2.159, SE = .163). This interaction highlights the importance of including participant ethnicity as a factor in order to examine ingroup and outgroup effects in Study 1. No other main effects or interactions were significant, all Fs ≤ .320, all ps ≥ .573. Overall, participants rated the target images in accordance with their intended categories.

**Study 1**

In order to examine both the categorization and infrahumanization of an individual whose appearance is changed to look like another race, Study 1 adapted Rips’ transformation paradigm to human categories. Specifically, we utilized double eyelid surgery because this procedure is common among Asian women to have more White-looking eyes (Davis, 2003; Kaw, 1993), and the eyes are the most important feature for categorizing an individual as Asian (Brown, Dane, & Durham, 1998). Participants viewed an image that depicted either an Asian woman, categorized her race, and rated which race she was most similar to. Participants were randomly assigned to read a vignette that described either a voluntary or involuntary racial cosmetic surgery that changed the woman’s appearance to look White. Participants viewed an image of the same woman with Caucasian eyes, categorized her, and rated which race she was most similar to based on the post-transformation image. We utilized difference scores (post-transformation – pre-transformation) on the categorization and similarity measures to gauge category movement. We hypothesize that we will observe a
dissociation between categorization and similarity in support (i.e., higher
difference scores for the similarity measure compared to the categorization
measure) of the origin essentialism hypothesis. However, if the casual
homeostasis hypothesis is supported, we will not observe a dissociation between
categorization and similarity (i.e., difference scores will be the same on the
similarity and categorization measures). Additionally, participants rated the
woman on the extent to which she possessed 10 uniquely human traits (5 positive
and 5 negative) and 10 human nature traits (5 positive and 5 negative) (Bastian &
Haslam, 2010; Haslam, Bain, Douge, Lee, & Bastian, 2005). We predict that there
will be the denial of these traits, such that participants will believe she possess
these traits to a lesser extent, when the transformation is voluntary. To examine
any possible ingroup and outgroup effects, the current study recruited participants
who identified as either Asian or White. Additionally, we varied the
transformation direction such that participants were randomly assigned to either
an Asian-to-White or a White-to-Asian transformation condition. Thus the design
to test the competing categorization hypotheses is a 2 (participant ethnicity: Asian
vs. White) by 2 (transformation direction: Asian-to-White vs. White-to-Asian) by
2 (volition: voluntary vs. involuntary) by 2 (rating type: categorization vs.
similarity) mixed factorial Analysis of Variance (ANOVA) with the last factor
serving as a repeated measure. We do not predict any main effects or interactions
for the participant ethnicity, transformation direction, or the volition on the
categorization and similarity ratings; rather, these factors serve as exploratory variables. To test the infrahumanization effects, the design was separate 2 (participant ethnicity: Asian vs. White) by 2 (transformation direction: Asian-to-White vs. White-to-Asian) by 2 (volition: voluntary vs. involuntary) by 2 (valance: positive vs. negative) mixed factorial ANOVAs with the last factor serving as a repeated measure on the uniquely human traits and on the human nature traits. Support for the infrahumanization hypothesis would be overall lower ratings of both positive and negative traits in the voluntary transformation condition.

Method

Participants. Five hundred and fifty-eight self-identified Asian and White participants were recruited from Amazon’s Mechanical Turk. Participants were paid $0.50 for taking part in the study. Seventy-two participants were excluded from analyses for incorrectly answering manipulation check questions (described below) and/or for missing data. An additional five participants did not allow their data to be used. While the manipulation checks resulted in a nearly 13% drop-out rate, Oppenheimer, Meyvis, and Davidenko (2009) demonstrated that the use of such manipulation checks can substantially improved the validity of the data collected. Of the remaining 481 participants, 183 self-identified as female, 296 identified as male, and two did not specify. Additionally, 292 self-identified as White and 189 identified as Asian or Asian American.
Materials. Faces from the pilot study were used in Study 1. They were created using FACES software (IQ Biometrix, Inc.) and were identical with the exception of the eyes. One image was intended to be an Asian woman and the other was intended to be a White woman. The pilot study confirmed that participants from the population we sampled from agreed on the categorization and physical similarity of the images.

Vignettes were created that depicted the woman’s transformation. In all conditions, participants were told that the woman had contracted an eye disease and surgery was required to preserve her vision. She was given the choice of changing the shape of her eyes or keeping them the same. After the surgery, her eyes had been changed. All vignettes were identical with the exception of the direction of the transformation and whether the change was voluntary or involuntary. The labels used in the vignette explicitly state that the nature of the transformation (exact vignettes are available in Appendix B).

Categorization and similarity ratings were obtained by using measures adapted from Rips (1989). Categorization was measured by asking participants “How would you categorize this woman’s race?” Participants replied using a four-point scale (1 = Asian; 2 = More Asian than White; 3 = More White than Asian; 4 = White). Similarity was measure by asking participants “What race would you physically describe this woman as being most similar to?” Participants
replied using the same four-point scale (1 = Asian; 2 = More Asian than White; 3 = More White than Asian; 4 = White).

Infrahumanization was measured by participants rating the extent to which they believed the woman possessed the five positive uniquely human traits (broadminded, conscientious, humble, polite, thorough; $\alpha = .848$), five negative uniquely human traits (disorganized, hard-hearted, ignorant, rude, stingy; $\alpha = .854$), five positive human nature traits (active, curious, friendly, helpful, fun-loving; $\alpha = .838$), and five negative human nature traits (impatient, impulsive, jealous, nervous, shy; $\alpha = .726$) on a seven-point Likert scale (1 = not at all; 7 = very much so).

Two manipulation check questions were included to ensure that they understood and agreed with the intention of the vignette that described the transformation. One question asked about the volition of the transformation (i.e., whether the woman chose to have her appearance changed or not). The other question asked about the direction of the transformation (i.e., after the surgery, indicate shape of the woman’s eyes).

**Procedure.** Upon completing an implied consent, participants were randomly assigned to one of the four experimental conditions (Asian-to-White, Voluntary; Asian-to-White, Involuntary; White-to-Asian, Voluntary; White-to-Asian, Involuntary). Participants were presented with an image depicting the woman’s pre-transformation state. Participants categorized her on the four-point scale
adapted from Rips (1989). Participants also rated which race she was most similar
to (based on her physical appearance) on the four-point scale adapted from Rips
(1989). After rating the image, participants read the vignette, which described the
transformation that corresponded to their condition. Participants then viewed the
second image depicting the post-transformation state and completed a second set
of categorization and similarity ratings based on that image. Order of the
categorization and similarity ratings, as well as the order of the scale points on
those measures, was counterbalanced across participants.

After completing the categorization task, participants rated the extent to
which they believed the woman possessed the ten uniquely human and ten human
nature traits.

Participants then completed the two manipulation check questions and
basic demographic information (i.e., self-identified gender, age, and self-
identified ethnicity). After completing these measures, participants were
debriefed, asked for permission to use their data, and thanked.

**Data Transformation.** Categorization and similarity ratings were recoded in the
same manner as in the pilot study. That is, the ratings were recoded such that
lower values indicate ratings that correspond to the target’s intended initial race.
For example, the White target rated as “White” would be a value of “1” as would
the Asian target being rated as “Asian.” Furthermore, difference scores were
computed to assess how much participants’ ratings moved from the original category to the final category from pre-transformation to post-transformation.

As another form of manipulation check, only participants who agreed on the pre-transformation categorization and similarity of the target were included in subsequent analyses. Specifically, participants were included if they categorized and rated the physical similarity of the initial target in accordance with the pilot ratings. For example, participants in the Asian-to-White condition were included if on the initial categorization and similarity ratings, they stated that the woman was either “Asian” or “More Asian than White.” This was to ensure that participants included in the analyses had similar perceptions of the pre-transformation phase.

The five items from each of the types of trait ratings (positive uniquely human, negative uniquely human, positive human nature, and negative human nature) were averaged to create a composite score for each of the four measures. The four composite scores for each participant were used in subsequent analyses.

Results

Categorization and Similarity Ratings. With categorization and similarity difference scores as the dependent measure, a 2 (transformation direction: Asian-to-White vs. White-to-Asian) by 2 (volition: voluntary vs. involuntary) by 2 (participant ethnicity: Asian vs. White) by 2 (rating type: similarity vs. categorization) mixed-factorial ANOVA, with the last variable measured within-
subjects, yielded a main effect of rating type, \( F(1, 408) = 40.002, p < .001, \eta^2_p = .089 \), indicating that people rate the woman’s similarity \((M = 1.212, SE = .064)\) as having changed more from time-one to time-two than categorization \((M = 0.976, SE = .050)\), supporting the *origin essentialism hypothesis* (see Figure 1). There was also a significant main effect of transformation direction, \( F(1, 408) = 3.935, p = .048, \eta^2_p = .010 \), suggesting that participants allow for more movement in the Asian-to-White condition \((M = 1.141, SE = .055)\) than in the White-to-Asian condition \((M = 0.981, SE = .059)\). The analysis also showed a significant main effect of participant ethnicity, \( F(1, 408) = 4.423, p = .036, \eta^2_p = .011 \), indicating that Asian participants \((M = 1.146, SE = .064)\) move the target more than White participants \((M = 0.976, SE = .064)\). Additionally, there was a significant rating type by transformation direction interaction, \( F(1, 408) = 4.022, p = .046, \eta^2_p = .010 \). Bonferroni adjusted simple effects analyses showed that, on the categorization ratings, there was a significant difference between the Asian-to-White condition \((M = 1.038, SE = .068)\) and the White-to-Asian condition \((M = 0.782, SE = .073)\), \( t(414) = 3.958, p = .011, d = .388 \). However, on the similarity ratings, there was no significant difference between the Asian-to-White condition \((M = 1.244, SE = .059)\) and the White-to-Asian condition \((M = 1.179, SE = .064)\), \( t(414) = 0.989, p = .461, d = .097 \). No other main effects or interactions were statistically significant, all \(Fs \leq 3.616, ps \geq .058\).
**Uniquely Human Traits.** To analyze subtle dehumanization of the woman, first a 2 (transformation direction: Asian-to-White vs. White-to-Asian) by 2 (volition: voluntary vs. involuntary) by 2 (participant ethnicity: Asian vs. White) by 2 (valance: positive vs. negative) mixed-factorial ANOVA, with the last variable measured within-subjects, was conducted on the uniquely human traits. The analysis revealed a significant main effect of valance, $F(1, 408) = 412.340, p < .001, \eta^2_p = .503$, demonstrating that, similar to the pilot study, participants were more likely to attribute positive traits ($M = 4.640, SE = .046$) than negative traits ($M = 2.948, SE = .052$). Additionally, there was a significant valance by volition interaction, $F(1, 408) = 11.253, p = .001, \eta^2_p = .027$. Bonferroni adjusted simple effects analyses showed that in the voluntary transformation condition, participants attributed positive traits ($M = 4.509, SE = .061$) to a lesser extent than in the involuntary condition ($M = 4.771, SE = .069$), $t(414) = -2.311, p = .005, d = .228$. Furthermore, participants were more likely to attributed negative traits to the woman in the voluntary condition ($M = 3.096, SE = .069$) compared to the involuntary condition ($M = 2.800, SE = .078$), $t(414) = 2.602, p = .005, d = .257$ (see Figure 2). There was also a significant valance by participant ethnicity interaction, $F(1, 408) = 7.898, p = .005, \eta^2_p = .019$. Bonferroni adjusted simple effects analyses revealed that Asian participants rated the woman as possessing negative traits ($M = 3.105, SE = .082$) to a greater extent than White participants ($M = 2.791, SE = .064$), $t(414) = 2.701, p = .003, d = .272$. However, when rating
negative traits, there was no significant difference between Asian participants \( (M = 4.561, SE = .073) \) compared to White participants \( (M = 4.718, SE = .057) \), \( t(414) = 1.350, p = .090, d = .136 \). No other main effects or interactions were statistically significant, all \( Fs \leq 2.240, ps \geq .135 \).

**Human Nature Traits.** Similar to the uniquely human traits, a 2 (transformation direction: Asian-to-White vs. White-to-Asian) by 2 (volition: voluntary vs. involuntary) by 2 (participant ethnicity: Asian vs. White) by 2 (valance: positive vs. negative) mixed-factorial ANOVA, with the last variable measured within-subjects, was conducted on the human nature traits. Like the uniquely human traits, the analysis revealed a significant main effect of valance, \( F(1, 408) = 203.208, p < .001, \eta^2_p = .332 \), indicating that participants were more likely to attribute positive traits \( (M = 4.630, SE = .044) \) than negative traits \( (M = 3.626, SE = .046) \). Additionally, the analysis yielded a significant main effect of volition, \( F(1, 408) = 6.166, p = .013, \eta^2_p = .015 \), indicating that people in the voluntary condition \( (M = 4.197, SE = .037) \) rated the woman as possessing human nature traits to a greater extent than those in the involuntary condition \( (M = 4.059, SE = .042) \). However, this main effect was qualified by a valance by volition interaction, \( F(1, 408) = 10.441, p = .001, \eta^2_p = .025 \). Bonferroni adjusted simple effects analyses revealed that participants in the voluntary condition were more likely to attribute negative traits to the woman \( (M = 3.809, SE = .060) \) compared to the involuntary condition \( (M = 3.443, SE = .069) \), \( t(414) = 3.803, p < .001, d = \)
.375, yet there was no significant difference for the attribution of positive traits in
the voluntary condition ($M = 4.586, SE = .058$) compared to the involuntary
condition ($M = 4.675, SE = .066$), $t(414) = .935, p = .307, d = .092$ (see Figure
3). These findings suggest that the main effect of volition is primarily driven by
the greater attribution of negative human nature traits to the woman who
undergoes the transformation voluntarily. A significant valance by participant
ethnicity interaction, $F(1, 408) = 14.825, p < .001, \eta^2_p = .035$, was also observed
for the human nature traits. Bonferroni simple effects analyses showed that, in the
voluntary condition, Asian participants ($M = 3.786, SE = .072$) rated the woman
as possessing negative human nature traits to a greater extent than White
participants ($M = 3.466, SE = .056$), $t(414) = 3.253, p = .001, d = .328$.
Additionally, Asian participants ($M = 4.519, SE = .069$) rated the woman as
possessing positive traits to a lesser extent compared to White participants ($M =
4.742, SE = .054$), $t(414) = -2.257, p = .012, d = .228$. No other main effects or
interactions were statistically significant, all $Fs \leq 2.001, ps \geq .158$.

**Discussion**

When comparing the amount of movement from pre- to post-
transformation ratings, the observed dissociation between categorization and
similarity supports the *origin essentialism hypothesis*. This initial, experimental
evidence suggests that people base social categorization not on surface features
alone, but on a deep, underlying essence. In addition to the overall tendency to
allow for more movement on the similarity ratings compared to the categorization ratings, there were some interesting nuances in the ratings. Specifically, the tendency for participants to allow for more movement in the Asian-to-White condition may suggest that people view this type of transformation as more natural and/or more common than a White-to-Asian transformation. Similarly, Asian participants allowing more movement overall compared to White participants might be indicative of similar beliefs. That is, these findings may suggest that Asian participants view these types of transformations as more natural and/or more common. However, because the participant ethnicity did not interact with rating type or transformation direction, it becomes difficult to understand what aspects of the categorization process are driving this effect.

Study 1 did not provide direct support for the *infrahumanization hypothesis*, as this necessitated rating the woman as possessing uniquely human and/or human nature traits to a lesser extent in the voluntary condition. However, there were interesting results for both types of ratings. The primary finding for these ratings was the overall negative evaluations of the woman when she undergoes the racial cosmetic surgery voluntarily. Again, while this is not a classic infrahumanization effect, it does highlight the importance people place on discrete category boundaries. Specifically, by choosing to cross a racial category boundary, there appears to be negative social repercussions directed toward the individual who undergoes this transformation. Additionally, participant ethnicity
interacted with these ratings, however, ethnicity did not interact with volition or transformation direction. While this is an interesting observation, the data suggest that Asian participants are rating these traits to a greater extent overall.

While the negative repercussions observed in Study 1 are very compelling, there are some alternative explanations that need to be addressed. First, despite being matched on ratings in the pilot study, it is possible the combination of the images and the voluntary transformation vignettes may have impacted the ratings in Study 1. Additionally, it is possible people may simply stigmatize cosmetic surgery in general. Therefore, any negative attitudes observed in Study 1 would also be present for any type of physical transformation, race-related or otherwise.

**Study 2**

The purpose of Study 2 was to address some of the limitations of Study 1. Specifically, we hoped to rule out any effects that may have been caused by the stimulus images. Additionally, we tested the alternative explanation that people may hold negative toward an individual who has voluntarily undergone any type of cosmetic surgery. Overall, the structure of Study 2 was very similar to Study 1, with the exception that there were no categorization or similarity ratings and no images were presented to participants. Participants read a vignette that described a woman who underwent a surgical procedure that altered either race phenotypic feature (i.e., her eyes) or a non-race phenotypic feature (i.e., her chin). Like Study 1, these transformations were either voluntary or involuntary. While we did not
observe any effects of transformation direction in Study 1, we wanted to ensure that the woman's race would not influence the results of Study 2, so target race was also included as a factor. Participants then rated the extent to which they believed the woman possessed the same uniquely human and human nature traits from Study 1. Thus, the design was two separate 2 (transformation type: race phenotypic vs. non-race phenotypic) by 2 (volition: voluntary vs. involuntary) by 2 (target race: Asian vs. White) by 2 (valance: positive vs. negative) mixed factorial ANOVAs for both types of traits, with the valance factor measured within-subjects.

We predict that the traits ratings will replicate the findings from Study 1 such that people will rate the woman as possessing negative traits to a greater extent and positive traits to a lesser extent after a voluntary transformation. However, we hypothesize that people will view a voluntary race phenotypic transformation as violating the natural order of social categories and, therefore, the repercussions incurred in this condition will be more negative overall. An alternative hypothesis is that is possible people will view a race phenotypic transformation as calling the human essence into question, regardless of volition. If this is the case, then the non-race phenotypic transformation may serve as a control condition. Therefore, if the *infrahumanization hypothesis* is supported in Study 2, we predict lower ratings overall for the uniquely human and/or human
nature traits for the race phenotypic condition compared to the non-race phenotypic condition.

**Method**

**Participants.** Two hundred and seventeen participants were recruited from Amazon’s Mechanical Turk and paid $.50 for taking part in the study. Thirty-three participants were excluded from analyses for incorrectly answering manipulation check questions (described below) and/or for missing data. An additional three participants did not allow their data to be used. Of the remaining 181 participants, 94 self-identified as female and 87 identified as male. As participant ethnicity did not interact with the main findings in Study 1, we recruited from all ethnicities in Study 2 and did not include participant ethnicity as a factor in analyses. One hundred and forty-three participants self-identified as White, 11 identified as Asian or Asian American, and 27 identified as other.

**Materials.** Vignettes were created that depicted the woman’s transformation. The vignettes for the race phenotypic transformation conditions were identical to those in Study 1. In the non-race phenotypic conditions, participants were told that the woman had an infected gash on her chin. Upon surgery, she was given the choice of changing the shape of her chin or keeping it the same. After the surgery, her chin had been changed to appear narrower. All non-race phenotypic vignettes were identical with the exception of the woman’s initial race and whether the change was voluntary or involuntary. The labels used in the vignette explicitly
state the nature of the transformation (exact vignettes are available in Appendix B).

Infrahumanization was measured by participants rating the extent to which they believed the woman possessed the five positive uniquely human traits (broadminded, conscientious, humble, polite, thorough; α = .810), five negative uniquely human traits (disorganized, hard-hearted, ignorant, rude, stingy; α = .914), five positive human nature traits (active, curious, friendly, helpful, fun-loving; α = .770), and five negative human nature traits (impatient, impulsive, jealous, nervous, shy; α = .768) on a seven-point Likert scale (1 = not at all; 7 = very much so).

Two manipulation check questions were included to ensure that they understood and agreed with the intention of the vignette that described the transformation. One question asked about the volition and type of the transformation (i.e., whether the woman chose or not to have either her chin or eyes changed). The other question asked about the direction of the transformation (e.g., after the surgery, what was the shape of the woman’s chin).

Procedure. Upon completing an implied consent, participants were randomly assigned to one of the eight experimental conditions. Participants read the vignette that described the transformation corresponding to their condition. Participants rated the extent to which they believed the woman possessed the uniquely human and human nature traits.
Participants then completed the two manipulation check questions and basic demographic information (i.e., self-identified gender, age, and self-identified ethnicity). After completing these measures, participants were debriefed, asked for permission to use their data, and thanked.

**Data Transformation.** The five items from each of the types of trait ratings (positive uniquely human, negative uniquely human, positive human nature, and negative human nature) were averaged to create a composite score for each of the four measures. The four composite scores were used in subsequent analyses.

**Results**

**Uniquely Human Traits.** To examine whether there are more negative repercussions for voluntarily undergoing a race phenotypic compared to a non-race phenotypic transformation, we submitted the uniquely human trait ratings to a 2 (transformation type: race phenotypic vs. non-race phenotypic) by 2 (volition: voluntary vs. involuntary) by 2 (target race: Asian vs. White) by 2 (valance: positive vs. negative) mixed factorial ANOVA, with the last factor measured within-subjects. Like Study 1, we observed a significant main effect of valance, \( F(1, 173) = 101.415, p < .001, \eta^2_p = .370 \), indicating that participants rate targets as possessing positive traits (\( M = 4.233, SE = .073 \)) to a greater extent than negative traits (\( M = 2.825, SE = .092 \)). Also replicating Study 1, the analysis revealed a significant volition by valance interaction, \( F(1, 173) = 29.532, p < .001, \eta^2_p = .146 \) (see Figure 4). While there was no significant volition by valance
by transformation type interaction, a planned contrast was conducted to test the hypothesis that a voluntary, race phenotypic transformation would result in more negative attitudes held toward the target. Specifically, for the voluntary conditions, we compared the trait ratings for the race phenotypic to the non-race phenotypic transformations. Bonferroni adjusted comparisons revealed that, for the voluntary conditions, people rated the woman as possessing positive traits to a lesser extent when the surgery was race phenotypic ($M = 3.591, SE = .143$) compared to when the surgery was non-race phenotypic ($M = 4.140, SE = .142$), $t(96) = -2.015, p = .007, d = .415$ (see Figure 5). However, there was no significant difference in the attribution of negative traits for voluntary race phenotypic transformations ($M = 3.331, SE = .180$) compared to voluntary non-race phenotypic transformations ($M = 3.104, SE = .180$), $t(96) = .837, p = .372, d = .172$. The analysis also revealed a marginally significant transformation type by valance interaction, $F(1, 173) = 3.605, p = .059, \eta^2_p = .020$, however, this effect is driven by the difference reported in the planned contrast above. No other main effects or interactions were significant, all $Fs \leq 2.938, ps \geq .088$.

**Human Nature Traits.** Attribution of human nature traits was examined by submitting ratings to a 2 (transformation type: race phenotypic vs. non-race phenotypic) by 2 (volition: voluntary vs. involuntary) by 2 (target race: Asian vs. White) by 2 (valance: positive vs. negative) mixed factorial ANOVA, with the last factor measured within-subjects. The analysis revealed a significant main
effect of valance, $F(1, 173) = 36.725, p < .001, \eta^2_p = .175$, demonstrating that people rated the woman as possessing positive traits ($M = 4.212, SE = .066$) to a greater extent than negative traits ($M = 3.488, SE = .082$). There was also a significant main effect of volition, $F(1, 173) = 21.497, p < .001, \eta^2_p = .111$, indicating that people attribute human nature traits to a greater extent after reading about a voluntary surgery ($M = 4.058, SE = .062$) compared to an involuntary surgery ($M = 3.642, SE = .065$). However, this main effect was qualified by a volition by valance interaction, $F(1, 173) = 12.733, p < .001, \eta^2_p = .069$, replicating Study 1 (see Figure 6). Bonferroni adjusted simple effects analyses suggest that the main effect of volition is driven primarily by greater attribution of negative traits in voluntary surgery conditions ($M = 3.909, SE = .114$) compared to involuntary conditions ($M = 3.067, SE = .119$), $t(173) = 5.024, p < .001, d = .747$. For the attribution of positive traits, there was no significant difference between the voluntary surgery conditions ($M = 4.217, SE = .096$) and the involuntary conditions ($M = 4.207, SE = .092$), $t(173) = .059, p = .942, d = .008$.

The planned contrasts carried out for the uniquely human traits revealed no significant differences ($ts \leq .812, ps \geq .304$), indicating that there is no differential attribution of human nature traits between voluntary race phenotypic and non-race phenotypic surgeries. No other main effects or interactions were significant, all $F$s $\leq 3.289, ps \geq .071$. 
Discussion

The goal of Study 2 was to address some of the limitations from Study 1. Specifically, we presented participants with vignettes without the addition of images to ensure that the negative evaluations observed in Study 1 were not stimulus driven. In this regard, Study 2 replicated the main findings from Study 1 regarding negative social repercussions after voluntary cosmetic surgery.

In addition to eliminating any stimulus specific effects from Study 1, Study 2 sought to compare a voluntary surgery involving a race phenotypic feature to a voluntary surgery involving a non-race phenotypic feature. Analysis of the attribution of uniquely human traits supported our hypothesis that after voluntarily altering a race phenotypic feature, the negative repercussions incurred were more extreme than after voluntarily altering a non-race phenotypic feature. These findings suggest that people place great importance on social category boundaries and violation of these boundaries has social ramifications.

General Discussion

The findings of the current research suggest that, when encountering a woman who has changed her appearance to look like another race, people rate her physical appearance closer to her new racial category, yet state she is still more likely a member of her original category. This finding provides the first experimental evidence that people base social categorization not simply on physical appearance, but on a deep, unchanging essence. Not only does this
finding have valuable theoretical importance for the formation and activation of social categories, but has practical importance as well.

Having a more complete understanding of social categorization processes lends support to much of the prejudice and stereotyping literature. Not only are essences thought to be natural and immutable, people also tend to believe they are informative (Haslam, et al., 2000). Therefore, by categorizing in an essentialist manner, people may also believe that they have knowledge about the traits and behaviors of the person they are categorizing. Basing knowledge about an individual based on a perceived essence may lead to a more homogenous view of the category in question and, thus, result in stereotyping and prejudice. In fact, Bastian and Haslam (2006) found that people who hold more essentialist beliefs about human categories and attributes are more likely to endorse stereotypes. Additionally, in a review of the essentialism literature, Prentice and Miller (2007) highlight many of the issues that arise when essentializing social categories.

In addition to finding support that people base social categories on essence, Studies 1 and 2 found that when someone chooses to cross a racial category boundary, the individual is faced with negative social repercussions. While the infrahumanization hypothesis was not directly supported in the current research, there was overall negative evaluations of a woman who voluntarily changed her appearance to look like another race (Study 1). Additionally, this finding was more extreme, as indexed by the denial of positive, uniquely human
traits, when compared to a voluntary transformation of a non-race phenotypic feature (Study 2). Negative social repercussions observed in the current research, suggest that by crossing a racial category boundary, the individual is violating what is perceived as a highly natural system and, thus, subject to backlash. That is, when encountering an individual who is violating the natural order of social categories, people may feel the need to penalize the individual in order to maintain the status quo (Jost & Banaji, 1994).

Although the current findings are impactful, there are some issues that should be addressed in future studies. First, the current studies focused solely on a woman altering either a race phenotypic or non-race phenotypic feature. Additionally, only one type of race phenotypic feature was changed across studies. One possible explanation for why the causal homeostasis hypothesis was not supported is that people may not believe that there is a causal link between the specific features changed in the current research and the deeper properties. This may not be the case for other racially diagnostic features or for physical attributes associated with gender categories. The present findings may be specific to female targets and/or our specific manipulation. Therefore, future studies should address these concerns to enhance the generalizability of the results. Also, failure to detect an overall infrahumanization effect may be due to the lack of a true baseline. While Study 2 addressed an alternative explanation for the negative attitudes observed in Study 1, neither study compares the attribution of uniquely human
and human nature traits to an individual who has not undergone any cosmetic surgery. One possibility is that all transformation conditions would suffer from infrahumanization, yet when the transformation is voluntary, the individual incurs harsher backlash in the form of negative evaluations, and a voluntary, race phenotypic transformation results in repercussions that are harsher still. Finally, if the negative evaluations stem from a desire to maintain the status quo, then future studies should measure individual differences in the extent to which people hold system justifying beliefs.

The findings of the current research suggest that people view social categories as highly essentialized and place great weight on the category boundaries. In addition to addressing the methodological issues mentioned above, future studies should examine what can be done to decrease the importance of category boundaries in order to reduce essentialist categorization processes and, in turn, stereotyping and prejudice as well as the negative attitudes held toward those who blur the lines of social categories.
References


doi:10.1017/CBO9780511529863.009


doi:10.1111/j.1467-8721.2007.00504.x


doi:10.1017/CBO9780511529863.004


doi:10.1016/0010-0285(75)90024-9

incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology, 48*(1), 165-179.
Figure 1

*Results of categorization and similarity ratings from Study 1.*

Note. Lower values indicate participants allowed for little movement between Time 1 and Time 2 on category and similarity ratings. Higher values indicate that participants allowed for greater movement between Time 1 and Time 2 on category and similarity ratings.
Results of uniquely human trait ratings from Study 1.

Figure 2

*Involuntary*  *Voluntary*

Note. Higher values indicate that the target possesses the traits to a greater extent.
Figure 3

Results of human nature trait ratings from Study 1.

Note. Higher values indicate that the target possesses the traits to a greater extent.
Figure 4

Results of uniquely human trait ratings from Study 2.

Note. Higher values indicate that the target possesses the traits to a greater extent.
Results of the planned contrasts between voluntary, race phenotypic and voluntary non-race phenotypic surgery conditions on uniquely human trait ratings from Study 2.

Figure 5

Note. Higher values indicate that the target possesses the traits to a greater extent.
Figure 6

Results of human nature trait ratings from Study 2.

![Graph showing trait ratings for negative and positive valence](image)

Note. Higher values indicate that the target possesses the traits to a greater extent.
Appendix A

Pre-transformation Asian-to-White Stimulus/

Post-transformation White-to-Asian Stimulus

Pre-transformation White-to-Asian Stimulus/

Post-transformation Asian-to-White Stimulus
Appendix B

Asian-to-White voluntary (Studies 1 and 2)

An Asian woman contracted an eye infection and surgery was required to preserve her vision. She underwent surgery by an ophthalmologist who was also a plastic surgeon. The surgeon gave the woman the option of keeping her eyes the same or changing them. The woman requested that her eyes be changed to look like a White person’s. The surgeon complied with her request and her eyes ended up looking like a White person’s.

Asian-to-White involuntary (Studies 1 and 2)

An Asian woman contracted an eye infection and surgery was required to preserve her vision. She underwent surgery by an ophthalmologist who was also a plastic surgeon. The surgeon gave the woman the option of keeping her eyes the same or changing them. The woman requested that her eyes stay the same. The surgeon tried to comply with her request, but due to a surgical error her eyes ended up looking like a White person’s.

White-to-Asian voluntary (Studies 1 and 2)

A White woman contracted an eye infection and surgery was required to preserve her vision. She underwent surgery by an ophthalmologist who was also a plastic surgeon. The surgeon gave the woman the option of keeping her eyes the same or changing them. The woman requested that her eyes be changed to look
like an Asian person’s. The surgeon complied with her request and her eyes ended up looking like an Asian person’s.

**White-to-Asian involuntary (Studies 1 and 2)**

A White woman contracted an eye infection and surgery was required to preserve her vision. She underwent surgery by an ophthalmologist who was also a plastic surgeon. The surgeon gave the woman the option of keeping her eyes the same or changing them. The woman requested that her eyes stay the same. The surgeon tried to comply with her request, but due to a surgical error her eyes ended up looking like an Asian person’s.

**Asian Target non-race phenotypic voluntary (Study 2)**

An Asian woman was in an accident that left large, infected gash on her chin. She underwent surgery by an oral and maxillofacial doctor who was also a plastic surgeon. The surgeon gave the woman the option of keeping her chin the same or changing it. The woman requested that her chin be changed to appear narrower. The surgeon complied with her request and her chin ended up looking narrower.

**Asian Target non-race phenotypic involuntary (Study 2)**

An Asian woman was in an accident that left large, infected gash on her chin. She underwent surgery by an oral and maxillofacial doctor who was also a plastic surgeon. The surgeon gave the woman the option of keeping her chin the same or changing it. The woman requested that her chin stay the same. The
surgeon tried to comply with her request, but due to a surgical error her chin ended up looking narrower.

**White Target non-race phenotypic voluntary (Study 2)**

A White woman was in an accident that left large, infected gash on her chin. She underwent surgery by an oral and maxillofacial doctor who was also a plastic surgeon. The surgeon gave the woman the option of keeping her chin the same or changing it. The woman requested that her chin be changed to appear narrower. The surgeon complied with her request and her chin ended up looking narrower.

**White Target non-race phenotypic involuntary (Study 2)**

A White woman was in an accident that left large, infected gash on her chin. She underwent surgery by an oral and maxillofacial doctor who was also a plastic surgeon. The surgeon gave the woman the option of keeping her chin the same or changing it. The woman requested that her chin stay the same. The surgeon tried to comply with her request, but due to a surgical error her chin ended up looking narrower.