* Q1. Variables
* 1) Independent variables:
  + Regions of origin: North vs. South
    - Experiment 1: The South was defined as census division, and participants who spent the past 6 years in the South were considered southern. All other students were considered northern.
    - Experiment 2 & 3: The South was described by its “cultural geography”, and thus was defined as states having a score of 25 or more on Gastil’s Southernness Index.
  + Condition: insulted vs. not insulted
    - Experiment 1: Participants were considered insulted if they got bumped into by a confederate and were called an “asshole”. They were considered not insulted if they did not get bumped at all.
    - Experiment 2 & 3: The insult is the same as Experiment 1. However, participants were either bumped publicly, privately or not at all.
* 2) Dependent variables
  + Experiment 1
    - Emotional reactions:
      * Reactions of anger and amusement; 7-point scales.
      * Reactions of aroused, flustered, resigned or wary.
    - Projective hostility:
      * Word completion: whether participants completed words in a hostile way.
      * Face rating: participants guess what emotion the face was expressing.
      * Scenario completions: beginning or ending filled by participants for a scenario that was either neutral or had affront involved.
  + Experiment 2
    - Emotional reactions—the same as Experiment 1.
    - Physiological measure of stress: difference between the cortisol level before and after the bump.
    - Physiological measure of preparedness for future aggression: testosterone levels before and after the bump.
    - Desire of the participant to demonstrate toughness: public minus private discrepancy in shock taking.
    - Interpretation of ambiguous stimuli: hostility projected on scenarios that only possibly had affront or challenge involved.
  + Experiment 3
    - Perceived effect of the insult on one’s masculine status: the participant who was bumped publicly was asked to guess what that observer really thought of him.
    - Aggressive behavior in a challenge situation: the distance at which the participant decided to “chicken out” or give way to the confederate when they came across in the narrow hallway.
    - Dominance behavior in subsequent encounters after the insult: the firmness of the participant’s handshake and the degree of eye contact were rated by another confederate and rating of the participant’s overall degree of submission was given. Also, the difference between answers of the same participants to two traditional “macho” behavior questionnaires, one of which they perceived as “public” and the other as “private”.
* 3) Other variables
  + - Demographic variables: religion, SES, siblings, academic performance, height and whether participants were athlete or not.
    - Effectiveness of bump: measured through observer-report to make sure the bump worked.
* Q2: Core hypothesis
  + This study advances the causal inference that the southern culture of honor, which prevails in the South more than the North, causes a person to show more aggressive response if the person is insulted.
* Q3: Causal inference
  + 3 criteria for inferring causation:
    - The cause preceded the effect.
    - The cause was related to the effect.
    - We can find no plausible alternative explanation for the effect other than the cause.
  + How the study’s authors attempted to address the 3 criteria
    - “The cause preceded the effect”:
      * Participants got bumped and called an “asshole” (insult) happened before their aggressive response. In other words, participants’ aggressive response was manifested after they got bumped and called an “asshole” (the insult).
      * The culture of honor is rooted in the history of the South, and deeply impacting southerners, whereas the North does not have such a culture. Therefore, participants from two different regions had long possessed different dispositions, which had happened before the experiment, during which aggressive response took place.
    - “The cause was related to the effect”:
      * In experiment 1, the authors used a 2x2 design, showing that northerners where unaffected by insult, whereas southerners became primed for aggression after being insulted. However, when insult was not present, response of two groups of participants was the same.
      * In experiment 2, the authors used a 2x3 design, by measuring physiological response, showing that northerners were hardly affected by the insult, while southerners became more upset and stressed, and prepared for aggression on the physiological level.
      * In experiment 3, the authors used a 2x3 design, showing that after being insulted, southerners manifested increased aggressive and dominance behavior, whereas southerners displayed no such change.
    - “We can find no plausible alternative explanation for the effect other than the cause”:
      * The participants were all nonHispanic white male undergraduate students of U of M, and they all came from families that were well of financially. The similarities of participants in many aspects that are unrelated to the study helped to rule out the possibility that factors other than culture of honor and insult contributed to the different levels of aggressive response.
      * Observers who were evaluating participants’ emotional reactions had not idea if a participant was from the North or the South, which eliminated observer bias’s contribution to results of study.
      * Participants had been told a fake purpose of each study before it started, so that participants would not try to figure out the study purpose, and thus avoided demand characteristics.
      * In experiment 2, participants were told not to talk while having gums in mouths, so that they would not have talked to the observer, which could have affected their emotional reactions.
    - Limitation and imperfection:
      * Criterion 1: Since the baseline of participants emotions was not measured before the insult, it was possible that participants had been feeling aggressive already before the study started, and such emotion lasted after the insult.
      * Criterion 2: Since the culture of honor was no directly measured, but instead was assumed to have existed on southern participants, we cannot be 100% certain that it was the culture of honor, rather than other factors relating to regions, that was associated with different levels of aggressive behavior.
      * Criterion 3: third variables
        + Personality traits: Perhaps the differences of aggressive response between southerners and northerners after being insulted was resulted from different personality traits of people from two regions. Since the authors did not measure participants personality traits, this possibility can not be ruled out.
        + Affectivity: It is possible that the outcomes were affected by participants emotions that day, whose baseline was not obtained before the study started. Southern students of M of U, a northern university, were possibly more anxious or upset, because they were far away from home, dealing alone with all sorts of pressure for the first time.
        + Diversity: Generally, the South is of less diversity than the North, where people of a variety of races from all over the world live together. The lack of diversity might have led to southern participants to be less open toward interpersonal contacts that were not so pleasant.
        + The fake test purpose: the fake study purpose “limited response time conditions on certain facets of human judgment” might have made participants nervous or confused, which contributed to the study results.
* Q4: Internal Validity
* 1) Threats to internal validity in the study:
  + Testing: in each of the 3 experiments, participants were asked to finish a series of tasks after the insult; tasks that were completed earlier might have impacted the performance on tasks later. For example, in experiment 3, each participant were asked to finish two “macho” questionnaires, and it is possible that the answers to the first questionnaire affected how participants answer the second one. Also, in experiment 1, it is not certain if the previous projective hostility task did not affect participants affect, and thus affected their answers on the next task.
  + Selection: Southern participants and northern ones might have had systematic differences in characteristics, such as personality traits, which might have contributed to the study results, and therefore became a threat to internal validity.
* 2) Attempts to remove the threats
  + Random assignment: participants were randomly assigned to control group and insult group, so that characteristics existing on participants before the study were equally assigned to different groups. Randomization helped to eliminate threats such as history and maturation, and to some extent selection bias.
  + Double blind: observers who rated participants’ emotional reactions were not aware of where the participant was from. Also, participants had no idea what the study was really about. Therefore, observers were free from biases, and participants were not likely to display demand characteristics.
  + Control group with pretest and posttest:

Q5: Discriminant and convergent validity

1) Evidence for discriminant validity:

* + No correlation between measures of different constructs were provided in the paper, so there is no direct evidence for discriminant validity. However, there was indirect evidence.
  + Aggressive response: in experiment 1, observer ratings of participants anger and amusement after getting bumped showed very different patterns, with southerners showing more anger than amusement, whereas northerners showing the opposite (Table 1). The measuring of different constructs showed different patterns, which indicated discriminant validity.
* 2) Evidence for convergent validity
  + Again, no correlation was provided among different measures of the same or similar constructs, so I don’t think there is direct evidence for convergent validity. However, there was some indirect evidence.
  + Aggressive response: in the 3 experiment, the authors used various methods to measure participants aggressive response, for example, other report emotional reactions, projective hostility tasks, insult priming scenario, self-report macho questionnaires etc., and it turned out that after being insulted, compared with northerners, southerners showed more aggressive response in all of the ways of measurement, which indicated a consistent pattern.
  + The southern culture of honor: the authors assumed that places of origin are indications of culture of honor. So similarities of the two ways of deciding the South and the North could be regarded as convergent validity of the culture of honor. According to the authors, in experiment 1, whether a state was in the South or the North was defined by census divisions. As it was articulated in the footnote, in experiment 2 and 3, the division of the South and the North in the study was based on the score of Gastil’s Southernness Index. It turned out that compared with experiment 1, only Maryland and Delaware were excluded from the South; the similarity of divisions is an indication of convergent validity of the Southern culture of honor.
  + Insult:
* 3) Study measures:

The physiological measure for stress (cortisol levels and testosterone levels) was based on previous research, and all other measures were homemade, because no citations were found. Homemade measures might create the following problems:

* + - Construct validity is suspicious: we are not certain if the homemade measures were actually measuring aggressive response. For example, in experiment 3, the degree of eye contact and firmness of handshake could have measured social skills or social styles or even personality traits rather than aggression.
    - Reliability is suspicious: homemade measures could have had reliability issues. For example, whether answers that participants gave in the projective hostility measures were mostly out of random decisions, or how much the answers were affected by random decisions was not known.
* Q6. Common Method Bias

Dov’s paper is subject to common method bias.

* + - Common rater effect:
      * In experiment 1, all emotional reactions were rated by observers, which may have created common rater effect, especially the consistency motif. It was possible that, to maintain consistency, observers kept rating according to a certain pattern, such as kept rating people higher on the scales as being angry than being amused, or the other way round. This pattern might have contributed to the result that southerners were rated more angry than amused, and more angry than northerners.
      * In experiment 2, stress was measured only using participants’ saliva samples. This created the problem that it was hard to rule out the impact of other factors on the amount of chemical materials in saliva samples. For example, affectivity that was not related to the insult could have contributed to the result as well.
      * In experiment 3, the “macho” questionnaires were completed through participants’ self report, which could have been affected by the consistency motif. Participants trying to maintain consistent responses might have been the reason why there was no difference between the “public” and the “private” conditions.
    - Recommendations:
      * For experiment 1, add self-report as a second source of rating.
      * For experiment 2, use stress scale as another way of objective measure; also, self-report can be added as a way of subjective measure.
      * For experiment 3, acquaintance-report could be added as a source of other-report, in order to eliminate common method bias.
* Q7: Reliability
* 1) Definition
  + Reliability refers to the measurement of consistency, or extent to which test scores are free from random error.
* 2) Reliability indices
  + Test-retest reliability: the stability of a measure over time.
  + Alternate forms reliability: consistency of a measure over two different forms of the test.
  + Inter-rater reliability: the extent to which different raters agree with each other.
  + Internal consistency reliability:
    - Split-half reliability: consistency of a measure from one half of a test to the other half.
    - Item-total correlation: correlation between a single item and an entire scale (often with that item removed from the scale score).
    - Cronbach’s alpha reliability: because there are many ways to split a test into two halves, some researchers calculate all possible split-half correlations and then average them. This average is called Cronbach’s alpha reliability.
  + Scale length, reliability and observed correlation
    - Spearman-Brown prophecy formula: rnew, new = krxx/[1+(k-1)rxx]
    - In the formula, rnew, new is the new reliability after the length of the scale is changed; k is the factor by which the test length increases; rxx is the reliability of the “old” scale, whose length is about to be changed. The formula assumes that the items added to or subtracted from a measure are parallel.
    - As scale length increases, the reliability of the new scale increases.
    - observed correlation = true correlation \* sqrt (rxx)\*sqrt (ryy)
    - As the number of scale items increases, the reliability of Y (ryy) increase, whereas true correlation and rxx remain the same, and thus the observed correlation increases.
* Q8: Mediator and moderator
  + Moderator:
    - The study hypothesized a moderator, which is “insult”. In fact, either the culture of honor or insult could have functioned as a moderator. However, I decide to take insult as a moderator, which is consistent with the core hypothesis I mentioned in Q2.
    - Hypothesis: Insult moderates the relationship between the southern culture of honor and aggressive response, such that southerners manifest more aggressive response than northerners when insult is present.

The southern culture of honor

Aggressive response

Insult

* + Paraphrase: When insult is presented, southerners are more likely to show more aggressive response than northerners, because southerners subscribe to the southern culture of honor. However, when insult is absent, southerners and northerners show no difference in aggressive response.
  + Mediator
    - As far as I am concerned, the authors did not explicitly claimed any mediator, but there are two variables that I think are possible mediators.
    - Mediator 1: Physiological changes. in experiment 2 (p.49), the authors mentioned that “… are accompanied by physiological changes of a sort that might mediate genuine behavioral aggression”. However, besides this, the authors did not mention anything about mediators in the study.
    - My hypothesis for mediator 1: Physiological changes mediates the effect of the southern culture of honor on aggressive response, such that the effect of the culture of honor on aggressive response operates through physiological changes.

Aggressive response

Physiological changes

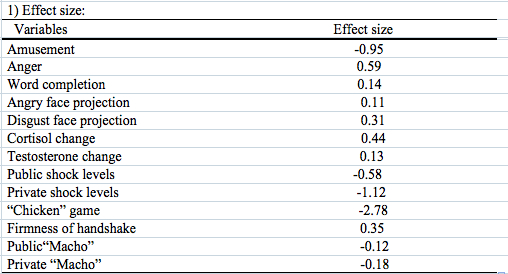
The southern culture of honor

* + Mediator 2: Perceived effect of the insult on masculine status. In experiment 3, the authors measured participants’ perceptions of how being insulted in public affected their masculine status in other people’s eyes.
  + My hypothesis for mediator 2: perceived effect of the insult on one’s masculine status mediates the effect of the southern culture of honor on aggressive response, such that the effect of the culture of honor on aggressive response operates through perceived effect of the insult on masculine status.
  + Paraphrase: the southern culture of honor of southerners does not lead directly to aggressive response. When southerners are insulted in public, they consider it to be a more serious harm to their masculine status in other people’s eyes than notherners do. Therefore, southerners are more likely to response aggressively when getting insulted in public.

Aggressive response

Perceived effect of the insult on one’s masculine status

The southern culture of honor

* Q9 HARKing
* 1) Any evidence of HARKing?
  + Experiment 1: I don’t think there is any evidence of HARKing. The authors reported significant difference between southerners and northerners on amusement and anger, but also reported nonsignificant ones such as how aroused, flustered, resigned or wary participants seemed. Nonsignificant results of projective hostility were also honestly reported.
  + Experiment 2: (also in experiment 3) the authors admitted that they got weak and inconsistent results about the emotional reaction to the bump. Also, the authors explicitly pointed out that the explanation for this result was post hoc. So no HARKing.
  + Experiment 3: Again, the authors did not hide nonsignificant result, which is that there was no difference between participants’ answer to “macho” questionnaire when it was perceived public and private. In addition, the authors admitted that this result was actually not consistent with their expectation.
  + In conclusion, I do not find any evidence for HARKing. However, if the authors were trying hard and left some results on purpose, there is no way I would know. According to the article, I cannot prove the existence of HARKing.
* 2) Six disclosure requirements:
  + 1. Authors must decide the rule for terminating data collection before data collection begins and report this rule in the article
  + -- The authors did not report the rule for terminating data collection.
  + 2. Authors must collect at least 20 observations per cell or else provide a compelling cost-of-data-collection justification.
  + -- All 3 experiments had more than 20 observations.
  + 3. Authors must list all variables collected in a study.
  + -- According to the study, the authors did list quite a few variables they collected. I am not 100% certain that the variables in the article are “all” variables. However, since some nonsignificant results were reported, I choose to believe that the authors listed all variables they collected.
  + 4. Authors must report all experimental conditions, including failed manipulation.
  + -- The authors did included failed manipulation, which did not yield significant results and results that were not consistent with authors’ expectations ( as mentioned in “Any evidence of HARKing?”)
  + 5. If observations are eliminated, authors must also report what the statistical results are if those observations are included.
  + -- The authors did not eliminate any results.
  + 6. If an analysis includes a covariate, authors must report the statistical results of the analysis without the covariate.
  + -- The authors thought being an athlete could have affected participants’ aggressive response, so they ran data analysis both with and without the athletes, and presented both in the article, although the two results were not very different.
* Q10:
* 1) 4 forms of control
  + Manipulation:
    - insulted vs. not insulted
    - publicly insulted vs. privately insulted vs. not insulted
    - tell participants “macho” survey is public vs. tell participants “macho” survey is private
  + Elimination:
    - Race，gender，SES of participants were holding constant ( white nonHispanic nonJewish males coming from financially well families);
    - Settings were holding constant (in the same lab and hallway)
    - Observers were the same who rated emotional reaction in every experiment
    - The confederate was the same person in each study
  + Statistical controls:
    - No ANCOVA or partial correlation was presented, so there was no statistical controls in the study.
  + Randomization:
    - Although it was not explicitly mentioned in the study, I think the authors used randomization when assigning participants to control and insulted groups. This is a very basic and important form of control, so I don’t think it was very likely that the authors did not do it.
* 2) Experiment, quasi-experiment or nonexperiment
  + Although places of origin did not adapt to manipulation, the other independent variable, insult, was manipulated in all of the 3 experiment; also, since I believe that the authors randomly assigned participants into control and insulted groups, I think all 3 studies were true experiments.
* Q11
* 1) Effect size:
* 

According to Cohen (1988), *d=*0.2 be considered a 'small' effect size, 0.5 represents a 'medium' effect size and 0.8 a 'large' effect size. Therefore, amusement, “chicken” game had large effect sizes, whereas word completion, angry face projection, testosterone change, public and private “macho” had small effect sizes, and all the others had medium effect sizes.

* 2) Type I error:
  + The probability of rejecting the null hypothesis when there actually is no effect, and it is a false positive.
    - Experiment 1:
      * Emotional reactions
        + Observers rated northern participants as significantly amused by the bump than southern participants, p < .01.
      * Projective hostility
        + Both control and insulted southerners were less likely to project happiness onto the faces than northerners were, p < .05.
      * Insult Prime Scenario
        + More percent of insulted southerners than control southerners completed the scenario with events in which the man injured or threatened to injure his challenger, p < .001.
        + Southerners were much more likely than northerners to complete the “affront” script with violence if they had been insulted than if they had not, p < 0.005.
    - Experiment 2:
      * Physiological levels:
        + Cortisol levels: after provocation, southerners show cortisol increases over the level of northerners and over the level of control groups, p < .03.
        + Testosterone levels: after getting insulted, southerners show testosterone increases over the level of northerners and over the level of control groups, p < .03.
      * Shock levels:
        + Southerners, whether insulted or not, chose to receive more shock in public than they did in private, p < .03.
    - Experiment 3
      * Damage to reputation:
        + Publicly bumped southerners were more likely to believe that their status was hurt in the eyes of the person who saw the insult, whereas notherners were hardly affected, p < .01.
* 3) p-value:
  + Given that the null hypothesis is true, what is the probability of getting a result as extreme as the one we get.
* 4) Statistical power:
  + The probability of finding the effect that is actually there.
  + Ways to improve power:
    - Get a larger sample: the standard error decreases as the sample size becomes bigger.
    - Increase the strength of treatment: give participants bigger insult.
* Q12: External validity
* 1) Sampling procedure:
  + Participants were recruited by telephone and paid for their time ($5 for experiment 1, $10 for experiment 2, and $15 for experiment 3).
* 2) Who is the sample?
  + In experiment 1 and 2, participants were nonHispanic nonJewish white male undergraduate students of U of I, and none of them were Michigan resident. In experiment 3, the composition of the samples were the same as Experiment 1 and 2, except that Michigan residents were excluded.
* 3) Response rates
  + Experiment 1: 83 in total, 42 northern, 41 southern;
  + Experiment 2: 173 in total, 111 northern, 62 southern;
  + Experiment 3: 148 in total, 88 northern, 60 southern.
* 4) Because the sample was limited to students at U of Michigan, the results of the study cannot generalize to all white male northerners and southerners. I think it is more appropriate to say that the sample is representative of white males American students in the U of Michigan, or perhaps universities in the North.
* 5) This study’s inferences might be vulnerable to the following threats:
  + Interaction of causal relationship with units
    - Because the sample was limited to white males, it is possible that results wouldn’t hold constant among other races or among women. Also, since participants were all undergraduate students, it is possible that results wouldn’t hold among people with lower education levels.
    - To partially rule out this threat, we could conduct a study with participants of a different race or gender. We could also get sample of different levels of education.
  + Interaction of causal relationship with settings
    - All 3 experiments in this study were carried out in the lab, where variables except for independent ones were under strict control. Therefore, it is reasonable to think that when the setting is changed from lab to, say, real life, where none variable is controlled, the results of this study will not likely to replicate.
    - To partially rule out this threat, we could replicate the study in a real life setting or natural setting.
  + Interaction of causal relationship with outcomes
    - For example, according to observers’ report, after getting insulted, a higher percentage of southerners showed anger than amusement, whereas the opposite was true among notherners. If the outcome observation is changed from observer-report to self-report, or test scores, it is possible that the same result would not hold.
    - To partially rule out this threat, we could replicate the study with a different way of obtaining measures of dependent variable (e.g., emotional reactions).
  + Interaction of causal relationship over treatment variations
    - Bumping into a participants and calling him an “asshole” is only one of the many forms of insult. Adopting other forms of insult, might induce different reactions.
    - To partially rule out this threat, I would adopt other types of insult, for example, hitting on one’s girlfriend or wife.
* Q13: Confidentiality and informed consent
  + I would like to believe that the authors used confidentiality and informed consent.
  + The authors did not say if the had used confidentiality and informed consent in the study. However, throughout the article, it is not hard to see that the authors did care about participants’ well-being. For example, after every experiment, the authors told participants about the real purpose of the study, and made sure that participants were not upset about the insult.
  + Also, the authors must have taken the IRB training, in which confidentiality and informed consent is emphasized.
  + If the authors cared enough to do work to make sure participants were not emotionally hurt by the study, it was very probable that they use confidentiality and informed consent.
  + Confidentiality and informed consent shows psychologists respect to the dignity and worth of all the people, and the rights of individuals to privacy, confidentiality and self-determination. It also helps participants to build the sense of trust to researchers and the study. Failing to use confidentiality and informed consent will be a violation of participants rights, as well as the IRB.
* Q14
* 1) Strategy:
  + The present study falls into experimental simulation, because the study was carried out in the deliberately contrived settings, and the authors were attempting to retain some realism of content (e.g., participants got bumped in the very same hallway by the same confederate; observers who rated participants’ emotional reactions were the same etc.) ; but at the same time, there were also trying to make setting as natural as possible, so that participants were not aware that the insult was the treatment (e.g., participants had no idea that getting bumped was part of the experiment, and they did know they were being observed etc.)
* 2) Dilemma
  + The experimental simulation octant lies in between the B maxima, which is concern with precision of measurement of behavior, and the C maxima, which is concern with system character of context.
  + B: the study was carried out in lab, where many irrelevant variables were controlled, and participants’ behavior was measured under contrived setting; since many other variables were controlled, the accuracy of measure of behavior would not be very poor. However, because the control was not perfect, plus the problem of common method bias, the accuracy is far from impeachable, too.
  + C: the study was carried out in lab, but was different from lab experiment in that there was some realism in the setting; participants were made to believe that when they got bumped in the hallway can called an “asshole”, they were dealing with real life insult. Therefore, this study is more like a combination of lab experiment with a little bit “real life” experience, though the latter was more dependent on participants’ perceptions.
  + A: the study is at a very low point on concern with generalizability with respect to populations. The sample of the study is one major cause of it. As discussed earlier, participants were all nonHispanic white male undergraduate students from U of Michigan. This has seriously constrained the results’ ability to targeted population.
* 3) Research proposal
  + The research I propose is aiming at improving the generalizability of the results. In order to maximize A (population generalizability), I am planning to adopt sample survey strategy.
  + Step 1: We randomly pick some white males from both the South and the North. To do this, we can randomly pick several northern and southern cities, and randomly pick streets in these cities, and then randomly pick buildings and houses on the streets, and randomly pick apartments in the buildings to decide whom to include in the study.
  + Step 2: We ask participants to report their background and life experience, including levels of education, incomes, how long they have been staying in the South (or the North),and their attitudes to the Southern culture of honor by asking about their stands on items related to the culture of honor.
  + Step 3: Measure participants’ aggressive response by providing stories about an insulting or noninsulting plot or situation (insulting ones for insulted groups while noninsulting ones for control groups), and then ask about participants reactions towards theses stories (e.g., What would you do if you were facing the situation?)
  + Step 4: Compare the differences across insult/noninsult and northern/southern groups.
* **Reference**
* Cohen, J. (1988). *Statistical power analysis for the behavioral sciencies*. Routledge.