

Inglorious Bliss

The Team Mentality

Part Two

I wrote Part One of Inglorious Bliss: The Team Mentality before I ever knowingly used AI. I also wrote the original Part Two. Later, I used AI to polish Part One's language without adding new material.

I published Part One after noticing that some ophthalmology practices use the word “team” on their websites while others do not. That difference brought me back to these articles and the questions behind them.

If I can draw one informed conclusion from my experience, it is that a workplace is never static. As people come and go, its relationships and working dynamics change. Any account of a workplace captures a particular combination of people, relationships, and circumstances at one point in that continuing change.

What follows, beginning with the Reverse Cliché Index, was written by two AI models, Claude and ChatGPT, based on the context I provided in Part One and my original Part Two. I did not write it. The two models developed it through an exchange guided by my instructions, and I brought it together through my review of their responses. It remains a work in progress and may contain numerous errors. I am comfortable sharing unfinished work, as long as it remains open to examination and correction.

What fascinates me is how AI can help an original thought develop, expose weaknesses in an argument, and suggest possibilities I might otherwise overlook. In my view, if you are not exploring your own ideas with more than one AI model, you are missing an important part of what these tools make possible. With thorough research and careful, AI-assisted revision, I can bring what follows closer to the standard I envision. It may take thirty or forty rounds of editing, but I now have the means to pursue that degree of refinement.

Reverse Cliché Index

From the word to a test

The question starts with a common assumption: employees called a team cooperate. This proposal separates two questions. Does the word lead readers to estimate more information sharing than a group's recorded conduct shows? Does it change how the people named behave? The first study measures readers' estimates of information sharing. The second tests sharing during a group assignment. Neither assumes that team helps or harms cooperation. No results are reported.

Expecting more cooperation from a group called a team is not, by itself, an error; the label may carry real information. The stronger test is whether readers still overestimate sharing after they have seen what the group actually did. Such a result would apply to this task and these materials only, and it would not establish anyone's motives.

Prior research supplies a reason to investigate, rather than an expected answer. Liberman and colleagues found more cooperation in a prisoner's dilemma framed as the Community Game than as the Wall Street Game. [1] Ellingsen and colleagues found results consistent with a beliefs explanation in their experiments. [2] Related studies include a Teamwork Game label, while later research reports

mixed framing effects and a role for both preferences and beliefs. [3, 4] These findings do not establish the direction or size of a team versus staff effect. This proposal makes no claim that the comparison is new.

Study	Random assignment	Primary question
Readers	Individuals to label and conduct record	Does team increase factual overestimation under identical evidence of limited sharing?
Participants	Groups of four to label and sharing instruction	Does team change accurate information sharing compared with staff under task-only instructions?

Study one: what readers do with evidence

Use the same task format, information packs, and scoring key planned for study two. Prepare clearly identified simulated chat records showing higher or lower sharing of essential facts. Code the actual fraction shared in each record in advance. Include several matched record variants, balanced across conditions, so one unusual exchange does not determine the result. The claim remains limited to the tested materials.

Randomly assign adult readers to team, staff, or a description listing the four participants' names without a collective label. Independently assign a higher-sharing or lower-sharing record, creating six conditions. The team and staff descriptions differ by that word alone. Every condition displays the same names. Within each record variant, conduct, task facts, reading access, and all other content are identical across labels. Participants do not see editorial descriptions such as cooperative or uncooperative.

First obtain a label-only estimate of the percentage of essential facts the group shared, together with confidence in that estimate. Then ask whether the reader would like to see how the group actually worked, and record the answer. Give every reader the assigned conduct record and the essential-fact list regardless of that answer. Ask for a final estimate of the percentage accurately shared, followed by confidence and a broader cooperation rating. Put recall and comprehension questions after the final outcomes. This sequence deliberately examines revision after evidence; a later independent sample could test whether making an initial estimate itself changes the result.

The primary outcome is final signed error: the estimated percentage shared minus the percentage established by the scoring key. Positive values mean overestimation. The primary contrast is team versus staff among readers given lower-sharing records. Names-only comparisons, initial expectations, confidence, the decision to view the record, higher-sharing records, and changes between estimates are secondary. A higher team estimate is not necessarily an error: interpretation must establish overestimation relative to the known record, rather than a correction of underestimation. Declining to view the record is reported descriptively; on its own, it does not show that the label shaped the final estimate. Because every reader receives the record, that decision does not change the evidence behind the final estimate.

Study two: what the wording changes

Recruit a separate sample of adults who have not taken study one. Form groups of four strangers and complete a neutral eligibility and technical check before randomization. Randomly assign whole groups to three labels crossed with two instruction conditions. Use equal allocation, with task variants balanced across the six conditions. Prevent repeat participation and specify permitted tools before recruitment.

Introduce the label with “Our team will complete this assignment,” “Our staff will complete this assignment,” or the four display names followed by “will complete this assignment.” Use the assigned form exactly twice, at the same points in the procedure. All groups see the same display names elsewhere. The names-only condition is an additional comparison, with different phrasing; it is not an inert placebo.

All groups receive the task goal and know that members hold different private information needed for an accurate decision. The task-only condition adds no reporting procedure. The explicit-sharing condition adds “Before submitting the group decision, each person must accurately report every fact in their private information pack.” This factor tests a reporting instruction. It does not by itself test formal role clarity, accountability, or realistic workload. Keep other instructions, resources, time, supervision, and payment rules identical across conditions.

The task should require information from every member without specialist knowledge. Pilot the reading burden, time limit, and difficulty to avoid near-perfect or near-zero sharing. Use the same fixed payment and any shared accuracy incentive in every condition, with no individual ranking reward. This supplies no explicit individual competitive incentive; it does not make competitive conduct impossible. Do not ask about cooperation or competition before the main task. Such questions may change the behavior being studied; examine expectations in a separate pilot or a separately randomized measurement extension. [5]

Behavior to examine

The primary outcome is the proportion of predetermined essential facts accurately communicated in the group chat before submission or the fixed deadline, whichever comes first. Count each fact once, accepting the paraphrases specified by a meaning-based scoring key. A correctly communicated fact counts whether or not the final answer uses it. Record false statements separately. The primary contrast is team versus staff in the task-only condition. The explicit-sharing contrasts, names-only comparisons, and label-by-instruction interaction are secondary unless separately powered and registered as confirmatory.

Two coders score each transcript independently. Report agreement before resolving disagreements under a written rule. Code interpersonal conduct separately, including help offered and requested, acknowledgment of contributions, dismissive remarks, and responses to correction. For exploratory rivalry indicators, predefine explicit personal-ranking or credit-dispute statements; do not treat all disagreement as competition. Use categories observable in text chat; rapid messages alone do not establish interruption. Record decision accuracy, noncompletion, and time. Omitted information does not by itself demonstrate intentional withholding or competition.

After outcomes are complete, ask participants which wording they recall, then what it meant, beginning with an open question. Ask about cooperation, rivalry, and the perceived purpose of the

study afterward. Retain groups in the primary analysis regardless of recall or hypothesis guesses. Any recall-based subgroup comparison is descriptive and does not preserve the randomized comparison. [6]

Controls and blinding

Deliver instructions automatically using neutral recruitment and consent materials. Automation standardizes delivery but does not guarantee that researchers are unaware of condition. Record which staff can access the assignment key. Participants know their assigned wording; the hypothesis and other conditions may be withheld only under the consent and debriefing arrangements approved for the study.

Have a data manager prepare masked scoring copies, replacing label references with a common token where this does not remove conduct needed for scoring. Keep original records for checking those redactions. Code any label-specific remarks requiring their original wording separately. Assessors may still infer condition from context or from the reporting procedure; record disclosures and assess the success of masking. Analysts use coded conditions until scoring, exclusions, and the registered analysis are complete. Describe masking by role rather than calling the study double blind.

Sample size and analysis

Choose a practically meaningful difference in percentage points for each primary outcome before collecting the main data. Ten percentage points is an illustration, not a justified threshold already selected for these tasks. Set the final margin from the task's consequences and measurement properties. Pilot both studies to estimate variability and completion, then simulate power for the actual primary contrasts and allocation. Use cautious ranges for nuisance parameters rather than treating a noisy pilot effect as the expected truth.

Standardized difference	Units per condition	Power target
0.5	64	80 percent
0.3	176	80 percent
0.2	394	80 percent

These verified figures illustrate one independent, equal-allocation, equal-variance two-sample t test at a two-sided 0.05 level. A standardized difference is the mean difference divided by the common standard deviation. The figures are not totals for either six-condition design and do not supply power for interactions, equivalence, multiple confirmatory comparisons, or missing data. Because each primary contrast uses two of six equally allocated conditions, the totals are six times the per-condition figure. For a standardized difference of 0.3, that is about 1,056 readers in study one and 1,056 groups of four (about 4,224 people) in study two. For 0.2, it is about 2,364 readers and 2,364 groups (about 9,456 people).

Analyze individuals in study one and groups in study two. Register models containing the assigned factors and task variant, with the stated primary contrasts and suitable uncertainty estimates. Use a two-sided 0.05 test for the one primary contrast in each study. Do not treat four members or multiple

facts from one group as independent randomized units. Report effect estimates and 95 percent confidence intervals. Keep other outcomes explicitly exploratory or control their declared comparison family.

Plan equivalence separately against the prespecified practical margin. For one contrast, two one-sided tests at 0.05 correspond to a 90 percent confidence interval entirely within the equivalence bounds. [7] A nonsignificant difference is insufficient. Design-specific interaction power must also be calculated: Gelman's sixteenfold illustration assumes a balanced design, comparable variances, and an interaction half the size of the main effect on the stated scale. It is not a general multiplier. [8]

Retain the original assignment of groups that lose a member. Do not insert a replacement into an ongoing group. Score sharing from the retained chat through the cutoff, using all essential facts assigned at randomization as the denominator, and report dropout and noncompletion by condition. A missing or corrupted record remains missing, rather than being assigned a zero. Register sensitivity analyses for missing outcomes and a fixed recruitment and stopping rule. Automatically discarding and replacing incomplete groups could bias the comparison. [6]

Working conditions to examine next

Later experiments could distinguish formal role clarity, accountability, workload, supervisory consistency, repeated labels, and competition for desirable assignments or recognition. Test "work as a team" separately from "our team." A label-by-individual-incentive comparison could examine whether the word changes conduct when members have something explicit to compete for. This would be an extension, rather than a reason to define spontaneous rivalry out of the initial study.

Group size, age, ethnicity, culture, religion, language familiarity, and gender composition remain separate questions, including the question about all-female groups. Related gender-framing research already exists and must inform those hypotheses; it does not directly test all-female workplace groups. [9] Collect characteristics only for a stated purpose. Measure relevant language proficiency and work experience without assuming either explains the effect. Randomize wording within planned composition categories. Distinguish the causal effect of assigned wording from an unsupported causal claim about demographic identity. Power each interaction rather than searching a long list of subgroups for a favorable result.

What the findings could support

If labels change initial expectations but readers accurately judge the later record, the label affects first impressions but not estimates made with the record in hand. If team produces greater factual overestimation under identical lower-sharing records, that would show the label biasing estimates even when the evidence is available, in this task. Check comprehension, confidence, and record-specific results before giving the finding a broader interpretation.

Study two could find more sharing, less sharing, practical equivalence, or an inconclusive comparison. A precise absence of a label effect would concern the chosen margin and conditions. A label can fail to increase sharing even when the group shares a great deal, so no added benefit does not mean no cooperation. Likewise, a reporting instruction's benefit does not establish that labels caused harm. The two studies answer related questions; results from separate samples do not establish a causal psychological mechanism between them.

Assign no numerical Cliché Index. Numerical research outcomes are measurements of specified judgments or conduct, not an index score. Claims about actual workplace conflict, turnover, productivity, or discovery and innovation require further research in workplaces with their history, hierarchy, and repeated relationships. A full-scale behavioral study should depend on feasibility, importance, and a credible precision plan, rather than solely on a significant result in study one.

Before recruitment

Finalize the task and evidence records, scoring manual, consent for chat recording, compensation and withdrawal arrangements, debriefing, data handling, and the appropriate ethics determination. Register the hypotheses, materials, primary comparisons, practical margins, sample sizes, exclusions, missing-data analyses, and stopping rules before the main studies. Archive the dated protocol and amendments so an independent researcher can distinguish the plan from later exploratory work.

References

- [1] Liberman, V., Samuels, S. M., & Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. *Personality and Social Psychology Bulletin*, 30(9), 1175–1185.
<https://doi.org/10.1177/0146167204264004>
- [2] Ellingsen, T., Johannesson, M., Mollerstrom, J., & Munkhammar, S. (2012). Social framing effects: Preferences or beliefs? *Games and Economic Behavior*, 76(1), 117–130.
<https://doi.org/10.1016/j.geb.2012.05.007>
- [3] Eriksson, K., & Strimling, P. (2014). Spontaneous associations and label framing have similar effects in the public goods game. *Judgment and Decision Making*, 9(5), 360–372.
<https://doi.org/10.1017/S1930297500006756>
- [4] Bernold, E., Gsottbauer, E., Ackermann, K. A., & Murphy, R. O. (2023). Accounting for preferences and beliefs in social framing effects. *Frontiers in Behavioral Economics*, 2, Article 1147492.
<https://doi.org/10.3389/frbhe.2023.1147492>
- [5] Hauser, D. J., Ellsworth, P. C., & Gonzalez, R. (2018). Are manipulation checks necessary? *Frontiers in Psychology*, 9, Article 998.
<https://doi.org/10.3389/fpsyg.2018.00998>
- [6] Montgomery, J. M., Nyhan, B., & Torres, M. (2018). How conditioning on posttreatment variables can ruin your experiment and what to do about it. *American Journal of Political Science*, 62(3), 760–775.
<https://doi.org/10.1111/ajps.12357>
- [7] Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4), 355–362.
<https://doi.org/10.1177/1948550617697177>
- [8] Gelman, A. (2018, March 15). You need 16 times the sample size to estimate an interaction than to estimate a main effect. *Statistical Modeling, Causal Inference, and Social Science*.
<https://statmodeling.stat.columbia.edu/2018/03/15/need16/>
- [9] Ellingsen, T., Johannesson, M., Mollerstrom, J., & Munkhammar, S. (2013). Gender differences in social framing effects. *Economics Letters*, 118(3), 470–472.
<https://doi.org/10.1016/j.econlet.2012.12.010>