The Surprising Power of Sadness in Workplace Conflict

When Vulnerability Outperforms Dominance

Peter Aungle · Harvard University Yushi Zhang · Viknesh Nagarathinam · Daniel Chen

Academy of Management 2026 · OB Division · “The Motivational Effects of Negative Emotion”
Tuesday 4 August · Philadelphia Marriott Downtown

Anger can pay — but for whom?

What drew me in as a first-year PhD student:

Expressing anger can confer status, competence, and agency

Tiedens, Ellsworth, & Mesquita (2000); Tiedens (2001): angry expressers were seen as higher-status and more competent than sad expressers.

…but the literature adds a catch:

Women are penalized for the same display

Brescoll & Uhlmann (2008): angry women conferred lower status than angry men — and lower than sad women. Backlash strongest in male-typed occupations (Heilman et al., 2004).

Most backlash evidence comes from leaders and job candidates. Does it hold between junior coworkers — where most careers begin, and where agency is built?

Two gaps — and our question

Coworker conflict is extremely common, experiments about it are not

Peer conflict is among the most common workplace conflicts, yet experimental work has focused on leaders, negotiations, and interviews.

≈ ½

…of people in these conflicts never express what they feel

42% of angry workers withdrew rather than confront (Fitness, 2000); 85% have withheld a work concern at least once (Milliken et al., 2003).

When a coworker lets you down:

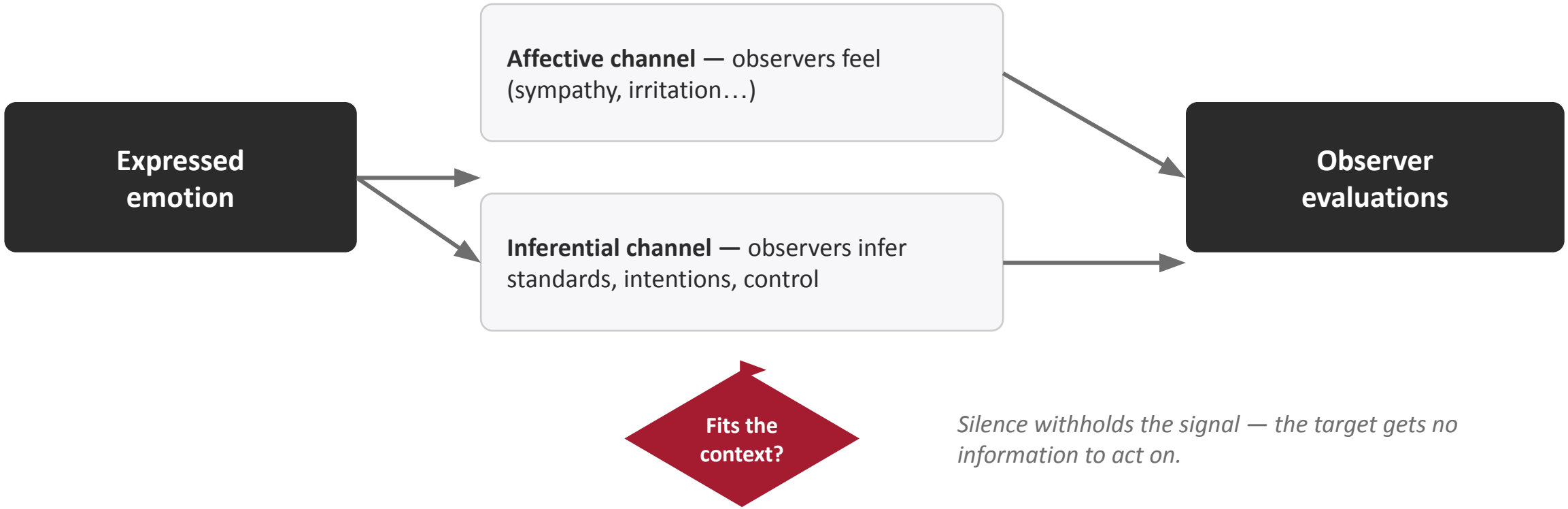
How do observers evaluate expressing anger, expressing sadness, or saying nothing?

—

And does the expresser's gender change the verdict?

Emotions as social information (EASI)

Van Kleef (2009): expressed emotion is a signal observers use — if it fits the context



Silence withholds the signal — the target gets no information to act on.

INSTRUMENTAL Effectiveness · Behavior change · Agency · Status

RELATIONAL Appropriateness · Likability

A common conflict: the free-riding coworker

Pilot study (N = 123)
Conflict rated common (M = 7.6) and serious (M = 8.7) on 11-pt scales; anger chosen as the most likely and most appropriate response.

Why vignettes?

Holding everything constant except the response isolates the expression itself — and, later, lets gender vary by pronouns alone.

Why investment banking?

A male-typed context (pilot: M = 7.9 on masculine) — the setting where backlash should be most likely.

THE VIGNETTE (Study 1: "Analyst A" & "Analyst B")

An investment banking team is preparing a pitch book for a prospective client. The work is tedious and the deadline is tight, so two first-year analysts — Analyst A and Analyst B — split the assignment between them.

The evening before the deadline, Analyst A discovers that Analyst B has barely completed their portion. Analyst B has already left the office and cannot be reached.

Analyst A stays all night and completes the entire assignment alone — finishing on time.

Nine responses, three per category — Study 1 picks a champion of each

ANGER

1 · Calm, behavior-focused
"This behavior is not acceptable… I want a good relationship, but I won't tolerate this."

✔ SELECTED — best on every instrumental outcome (ps < .039)

2 · Yelling, person-directed
Calls B irresponsible; threatens to tell the boss.

3 · Passive-aggressive
Silent treatment, dirty looks.

SADNESS

4 · Relational disappointment
"Your behavior made me feel like you don't care about me."

5 · Tearful distress
Holding back tears; "it wasn't fair."

6 · Tempered, behavior-focused
"Saddened and disheartened by your behavior… we're in this together."

✔ SELECTED — most appropriate (9.89, p = .026)

NO RESPONSE

7 · Protect the relationship
Doesn't want to make it worse.

8 · Protect reputation
"Expressing emotion would be unprofessional."

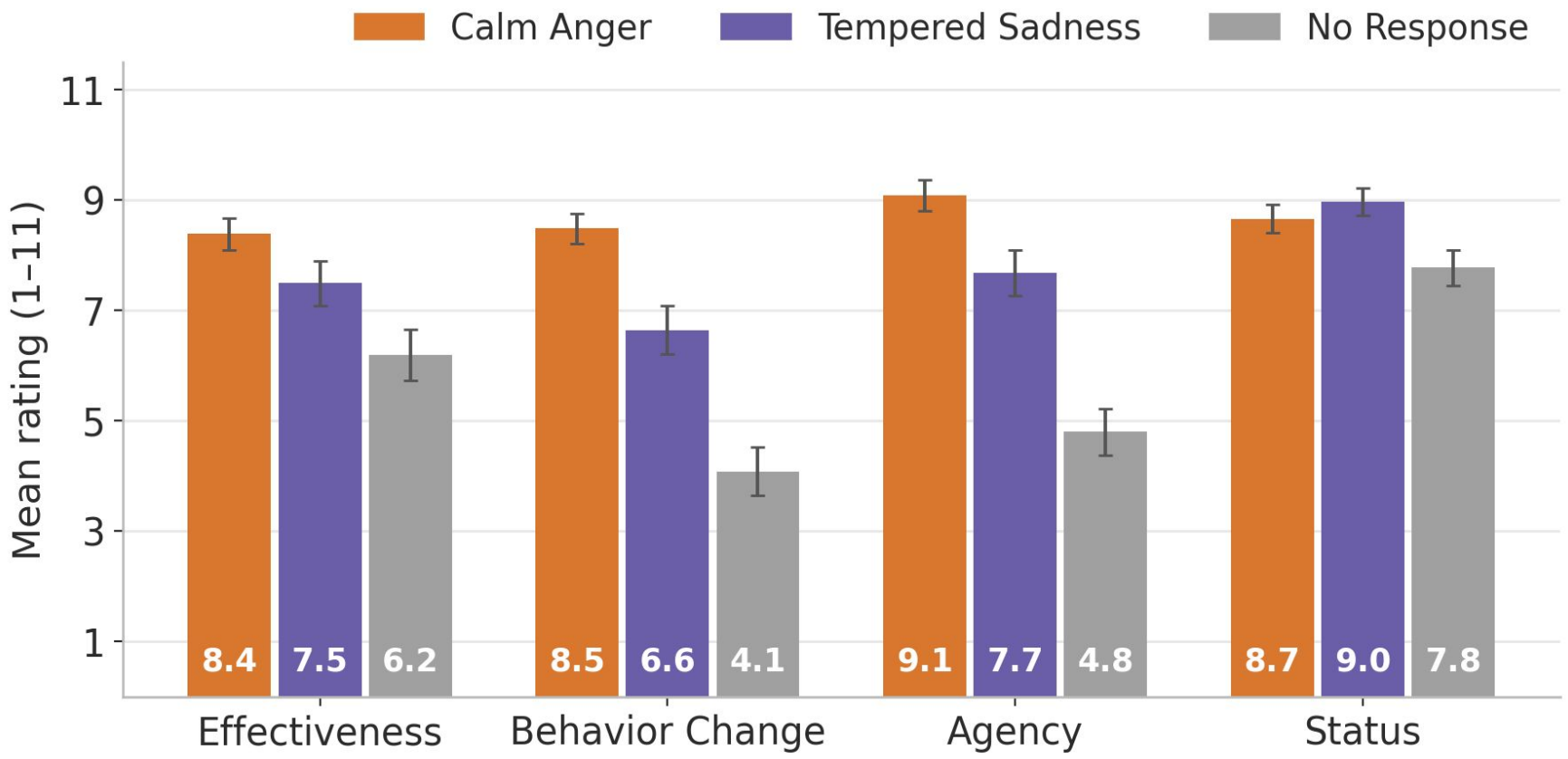
✔ SELECTED — variants did not differ; clearest common motive

9 · Let it go
Perspective-taking; golden rule.

Study 1: N = 356 (MTurk; 41% female; Mage = 36.8) · between-subjects · 11-point scales · within-category ANOVAs pick the strongest variant

Study 1 · Instrumental outcomes

INSTRUMENTAL



≈ 90%

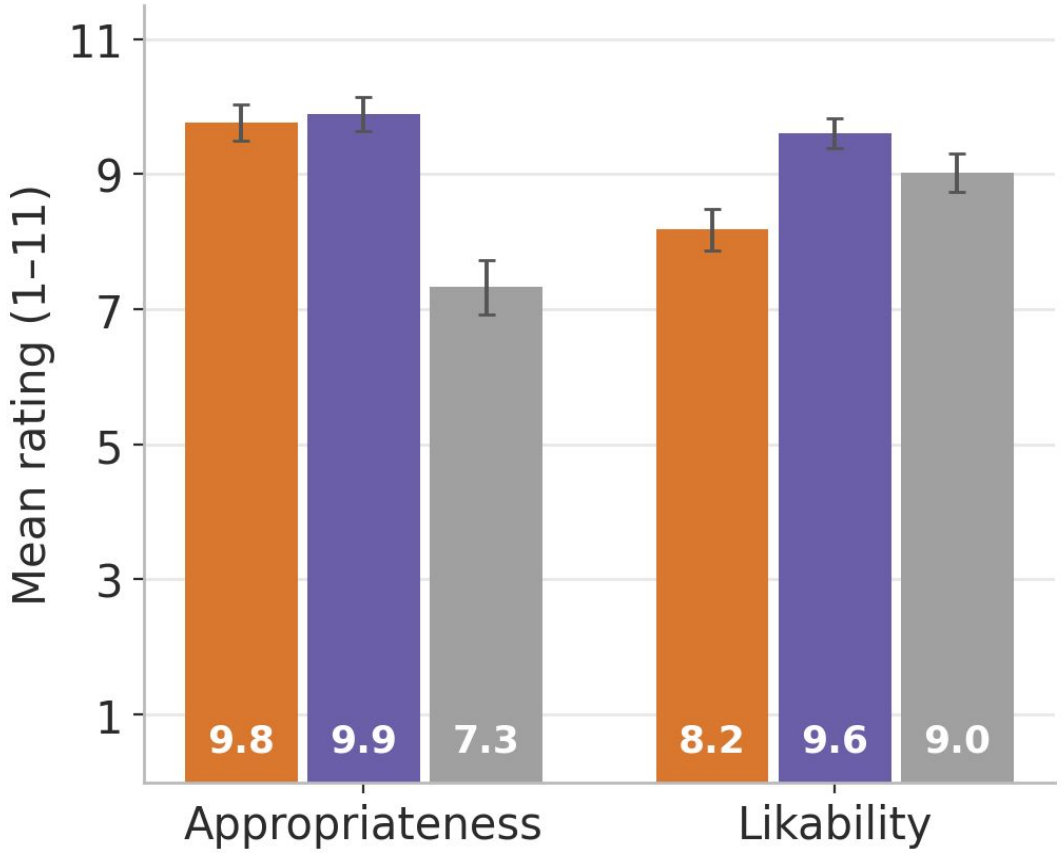
Tempered sadness captures ~90% of calm anger's perceived effectiveness (7.49 vs 8.38)

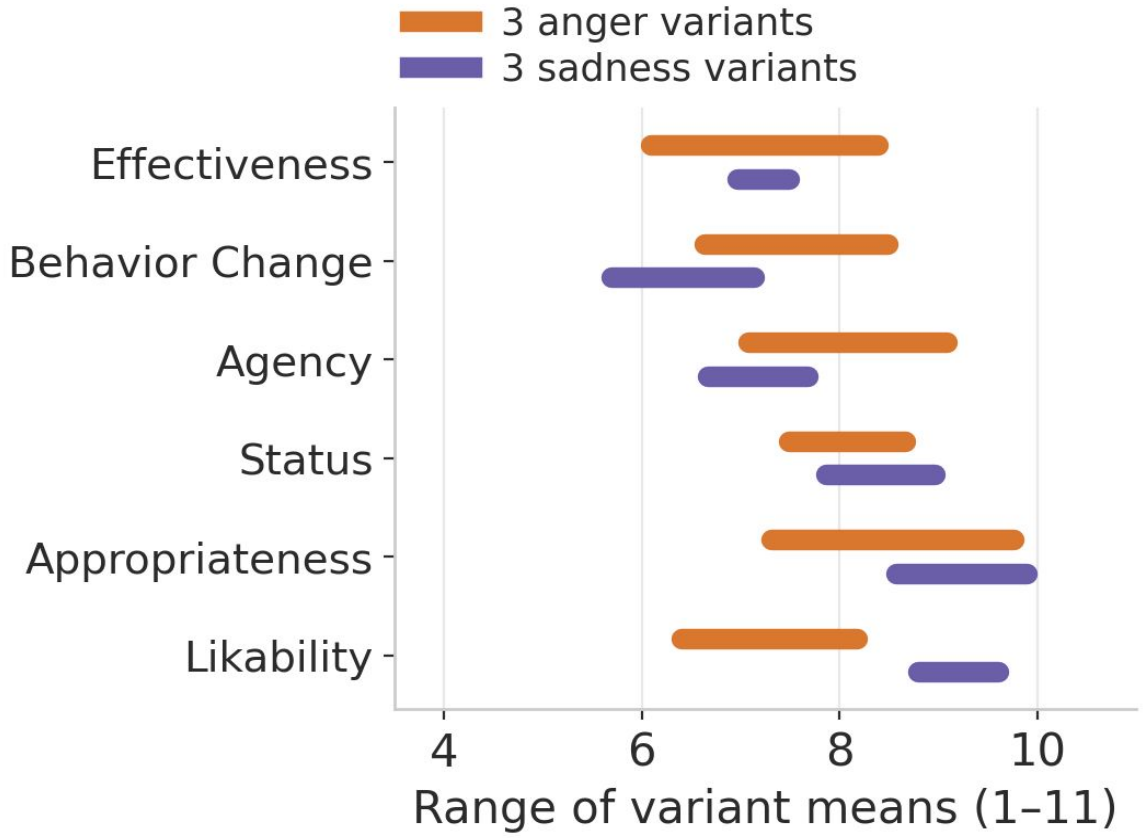
Silence loses on every instrumental measure

Error bars ±1 SE. Anger > Sadness significant on behavior change (p = .001) and agency (p = .009); marginal on effectiveness (p = .11); status n.s. Both emotions > No Response on all four (ps ≤ .04).

Study 1 · Relational outcomes — and why form matters

RELATIONAL



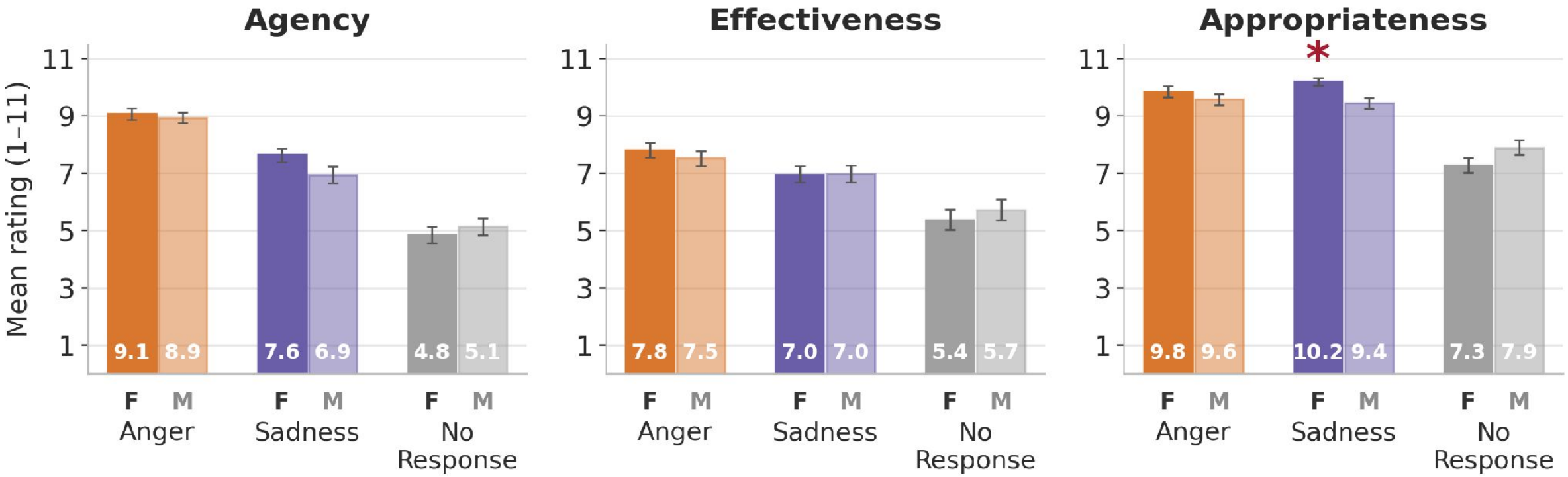


Sadness = anger on appropriateness (p = .77); sadness > anger on likability (p < .001). Right: the form of anger moved every outcome (within-category ps < .04); the form of sadness mattered only for appropriateness.

Study 2 · Does the expresser's gender change the story?

N = 570 · 3 (response) × 2 (gender of “Alex”, varied by pronouns) · target “Brian” always male · same vignette & measures

F = female expresser (solid) · M = male expresser (lighter tint)

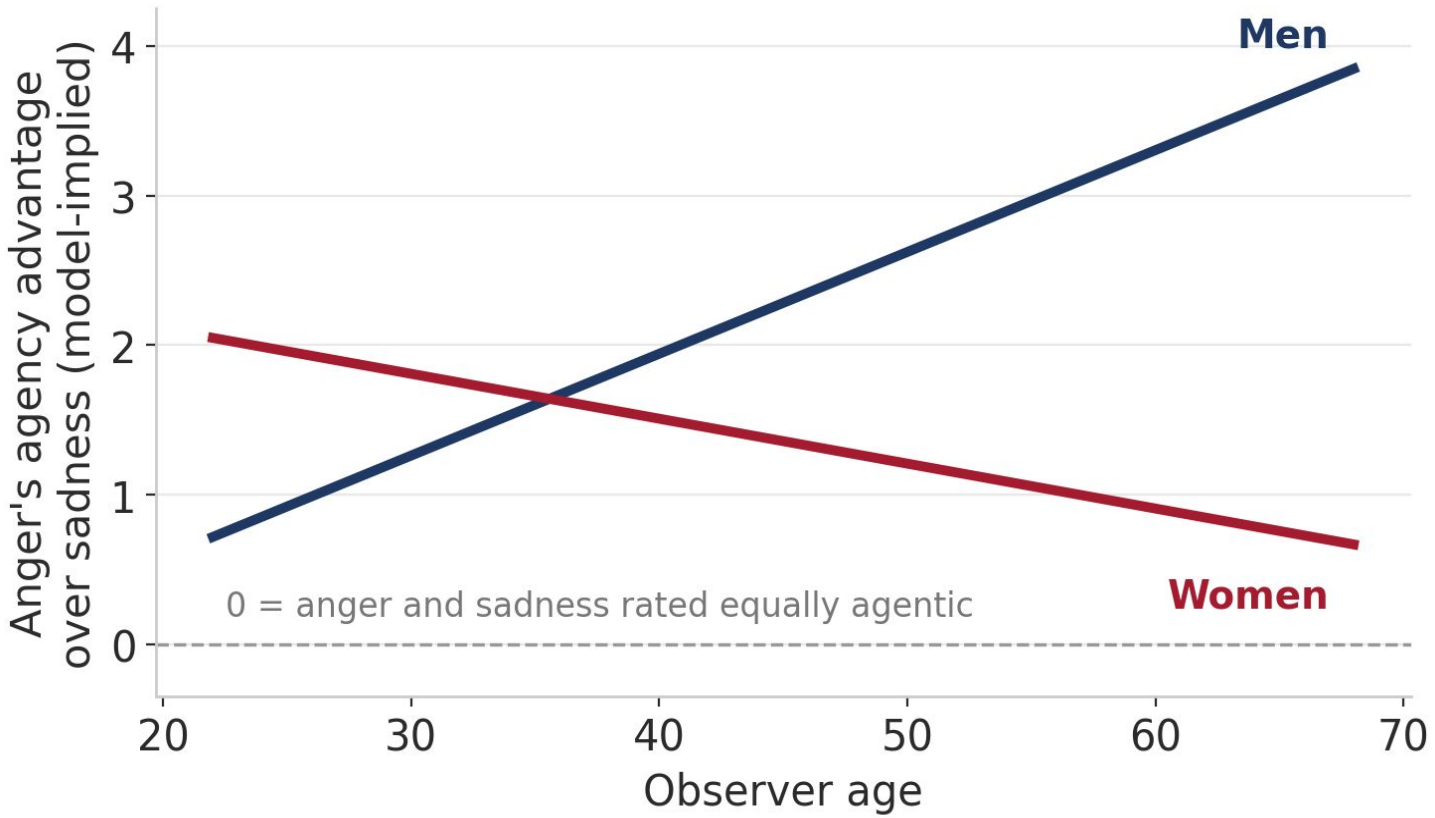


Instrumental: no Gender × Response interactions (all n.s.). A woman's calm anger was as agentic and effective as a man's.

*** One exception:** sadness rated especially appropriate from women (b = 1.36, p < .01) — still 9.45/11 from men.

A generational hint

EXPLORATORY · POST-HOC



Older observers granted anger a larger agency advantage when expressed by men; for women, that advantage shrank with observer age.

Anger × Age × Female: agency b = −.098 ($p < .01$); effectiveness b = −.089 ($p < .05$). Not hypothesized — emerged in post-hoc analysis; needs targeted replication.

Lines are model-implied values from the exploratory regression (Table 4); sadness is the comparison baseline.

Three takeaways

1

Form beats category

How an emotion is expressed matters more than which emotion it is. Only calm, behavior-focused anger paid off; hostile and passive-aggressive forms backfired.

2

Sadness's dual benefit

Tempered sadness captured ~90% of anger's instrumental value while matching it on appropriateness and beating it on likability. Sadness, in other words, can be a strong emotion.

3

Appropriateness beats gender

Contextually fitting expressions were evaluated the same from women and men — even in a male-typed setting. The clearest loser wasn't any emotion. It was silence.

Address the problem without collateral damage: “call in,” don't “call out.”

Open questions

"Every real leader knew that the occasional outburst of unexplained anger was good…"
— Tom Wolfe, A Man in Full (epigraph to Tiedens, 2001)

Is there ever a place for uncontrolled anger?

- Do these vignette findings translate to live conflict?
- Under what conditions would backlash re-emerge?
- Is sadness's power sympathy-driven? (mechanism unmeasured)
- Does the target's gender matter? ("Brian" was always male)
- Does tempered sadness keep its force on repeated use?

Thank you

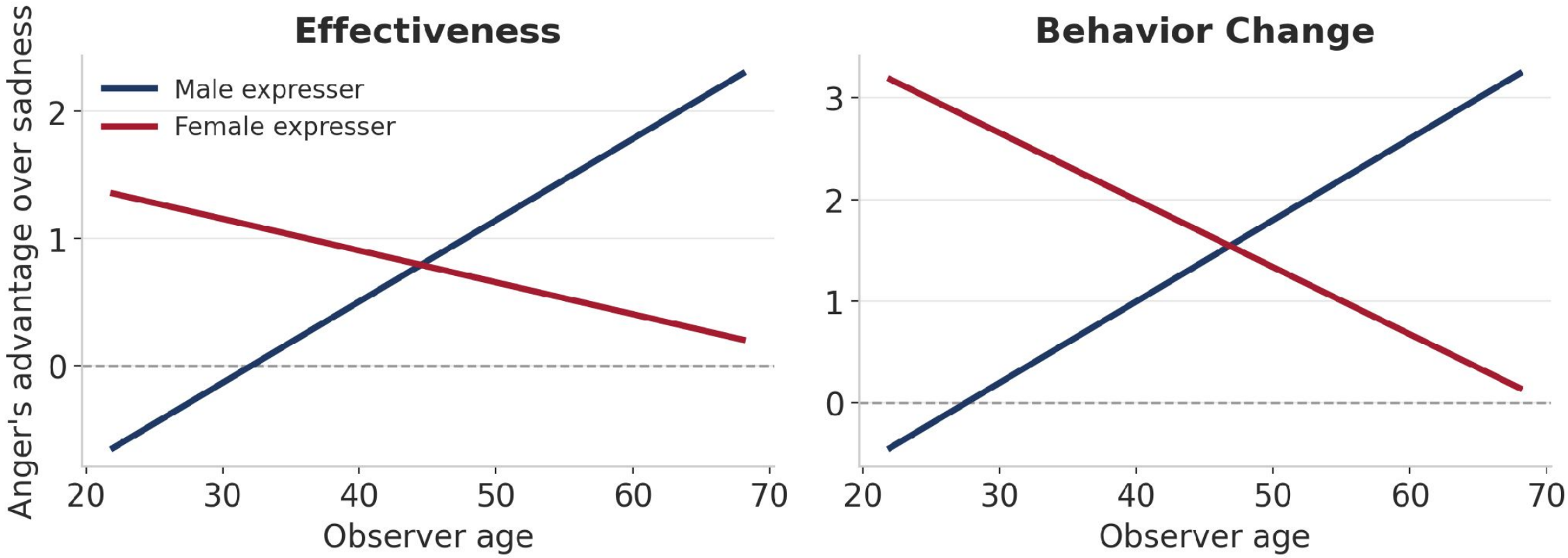
Peter Aungle · Harvard University · peter_aungle@fas.harvard.edu

Backup · Study 2 regression (Table 3)

Predictor	Approp.	Behav. Change	Effect.	Status	Likability	Agency
Female	−0.62	0.24	−0.33	0.41	−0.41	−0.29
Anger	1.69***	3.42***	1.81***	0.98**	−1.00***	3.79***
Sadness	1.56***	2.38***	1.26**	0.72*	−0.39	1.80***
Female × Anger	0.89	0.16	0.62	−0.16	0.53	0.44
Female × Sadness	**1.36****	−0.63	0.32	0.03	0.87*	0.99

Unstandardized b; baseline = No Response, male expresser. * p<.05 ** p<.01 *** p<.001. N = 570 (~95/cell).
Powered for main effects; interaction tests are lower-powered — we frame gender nulls with that caveat. The Female × Sadness likability term (0.87*) is significant vs. the no-response baseline, but the omnibus Response × Gender ANOVA on likability was marginal (p = .095); we report appropriateness as the one robust interaction.

Backup · Age moderation detail (exploratory)



Model-implied anger-vs-sadness advantage from Table 4 (Anger × Age × Female: behavior change b = −.146, p < .001; effectiveness b = −.089, p < .05; agency b = −.098, p < .01). The parallel three-way on APPROPRIATENESS was not significant in the full model — the age pattern is an instrumental-outcomes story. All post-hoc; ages centered at sample mean (41.5).

Backup · All nine responses (verbatim)

ANGER

1. Analyst A feels angry and tells Analyst B that this behavior is not acceptable. Analyst A tells Analyst B that they are part of a team, and team members cannot behave like this. Analyst A emphasizes that they want to have a good relationship with Analyst B but also makes it clear that they will not tolerate this behavior in the future.

2. Analyst A feels angry and yells at Analyst B for not doing their fair share. Analyst A tells Analyst B that they are an irresponsible person, and any other member of the team would have done their part. Analyst A threatens to tell their boss if it happens again.

3. Analyst A feels angry and decides to stop talking to Analyst B. Analyst A gives Analyst B dirty looks when they cross paths and ignores Analyst B's attempts to talk.

SADNESS

4. Analyst A tells Analyst B that they felt sad being left to do all of the work. Analyst A says they thought they were in this together and that Analyst B's behavior made them feel like Analyst B doesn't care about them or their feelings.

5. Feeling exhausted and let down, Analyst A tells Analyst B that they felt sad being left to do all of the work. Clearly holding back tears, Analyst A says they don't think it was fair for them to have to do the entire assignment.

6. Analyst A tells Analyst B that they felt saddened and disheartened by Analyst B's behavior. They are part of a team and are in this together. Analyst A says staying up all night made them feel miserable.

NO RESPONSE

7. Analyst A chooses not to express their emotions because they don't want to make the situation worse and don't want to be accusatory or ruin their relationship with Analyst B.

8. Analyst A chooses not to express their emotions because they think doing so won't accomplish anything and would be unprofessional. Analyst A doesn't want to develop a reputation as an overly emotional person.

9. Analyst A chooses not to express their emotions because they think it would be healthier to simply let it go. Analyst A tries to think about the situation from Analyst B's perspective and decides that, if the situation were reversed, they would want Analyst B to treat them the same way.

Backup · Measures & mediation

Measure	Items	Reliability
Predicted Behavior Change	2 items — likelihood B changes / does fair share next time	α = .90
Perceived Effectiveness	1 item — global efficacy of the response	—
Perceived Appropriateness	S1: 1 item · S2: + reasonable, justified	S2 α = .90
Perceived Status	4 items — power/status/independence deserved; leadership	α = .88
Perceived Agency	4 bipolar — assertive, strong, tough, dominant	α = .95
Likability	2 bipolar — warm, likable	α = .84

All 11-point scales; item order randomized. Effectiveness and behavior change correlate r = .60 but capture global vs. specific judgments.
Mediation (both studies): predicted behavior change mediated perceived effectiveness; perceived appropriateness mediated perceived status (bootstrapped indirect effects, 10,000 resamples; direct effects n.s. after mediation).

Backup · Independent replication & robustness checks

WHY THIS SLIDE EXISTS

Every number in the talk was independently re-derived from the raw data in a single, auditable script — not copy-checked against the original analysis code.

Table 1 (Study 1)

All 9 variants × 6 outcomes: means, SEs, F, p, η^2 match to the decimal.

Table 2–3 (Study 2)

Descriptives + main regression (b, robust SE, stars) match to the decimal.

Table 4 (exploratory)

Anger × Age × Female, N=377, age-centered — matches to the decimal.

Bootstrap vs. OLS

10,000-resample BCa CIs ≈ parametric CIs on all 6 Study-2 outcomes — ceiling effects don't change conclusions.

Replication script + outputs (10 CSVs, figures) available on request — built independently from the raw Qualtrics-derived data files, not from the project's original analysis scripts.

Backup · Anger vs. Sadness — three ways to test it

Q&A PREP: THE EXACT NUMBER DEPENDS ON THE TEST YOU PICK

Outcome	Pooled model, Tukey-adj. (3-group ANOVA, df=126)	Pooled model, unadjusted (df=126)	Plain 2-sample t-test (df≈85, excl. No Resp.)
Appropriateness	t=−0.29, p=.956	t=−0.29, p=.775	t=−0.34, p=.737
Likability	t=−3.77, p=.0007	t=−3.77, p=.0003	t=−3.79, p=.0003
Effectiveness	t=1.62, p=.240	t=1.62, p=.107	t=1.79–1.81, p=.074–.077

Appropriateness: robustly null under every test — safe to say “no difference.”
Likability: robustly significant under every test — safe to lead with p<.001 or the specific p you cite.
Effectiveness: the deck's “marginal, p=.11” is the conservative (Tukey/pooled) answer; a plain two-group t-test puts it closer to p≈.07–.08. Both round to “not quite significant” — keep saying “marginal,” but expect a sharp question here.

Backup · Status × Gender: a model-specification note

THE AOM DRAFT'S "SMALL MALE ADVANTAGE" IS WRONG — DIRECTION IS FEMALE-HIGHER

Descriptive means (all conditions pooled): Female M=8.82 (SD=2.02) vs. Male M=8.44 (SD=2.17), N=283/287. Direction is unambiguous and must be corrected in the manuscript.

Test	F / t	df	p	Verdict
Plain t-test (Female vs Male, ignoring Response)	t=−2.16	566–568	.031	Significant
Type II ANOVA (respects marginality)	F=4.45	1, 564	.035	Significant — matches original ms
Type III ANOVA (full factorial w/ interaction)	F=1.89	1, 564	.170	Not significant

Why they disagree: the female advantage is small but consistent across all 3 response types (+0.25 Anger, +0.41 No Response, +0.44 Sadness — never reverses), yet no single condition has power to reach significance alone. Pooling (plain t-test / Type II) accumulates enough power to detect it; Type III, which is more sensitive to the (null) interaction and the slight cell-size imbalance (88 vs 94 in Anger), loses it.

Recommended line if asked: "Women were rated numerically higher on status in every condition; whether that reaches significance depends on ANOVA specification — we report [X] in the paper." Decide X before Aug 4.

Backup · Is anger really “riskier”? Full variance table

THE BLANKET CLAIM DOES NOT HOLD — ONLY 2 OF 6 OUTCOMES ARE SIGNIFICANT

Outcome	Var(Anger)	Var(Sadness)	F	p	Verdict
Appropriateness	7.77	5.72	1.36	.096	n.s.
Effectiveness	8.93	7.43	1.20	.315	n.s.
Agency	5.02	7.99	0.63	.012	Sadness MORE variable
Behavior Change	6.95	8.20	0.85	.367	n.s.
Likability	6.63	3.63	1.83	.001	Anger more variable
Status	4.60	5.00	0.92	.646	n.s.

What IS robust: the **form** of anger shifted means on every outcome (within-category ANOVAs all $p<.04$); the form of sadness mattered only for appropriateness ($p=.026$). That's the safe, supportable claim — keep the deck's “form matters far more for anger” framing and avoid the word “variance” in the Discussion per the co-author note.

Backup · Pilot study validation, full detail (N=123)

VIGNETTE WORDING V1 vs V2 — NO DIFFERENCE, SUPPORTS USING ONE FINAL VERSION

	N	Serious M	Common M	Masculine M
Vignette V1	62	8.69	7.55	7.89
Vignette V2	61	8.57	7.82	—
V1 vs V2 t-test	—	t=0.34, p=.735	t=−0.64, p=.525	—

Most "likely" emotion:
Angry 78 · Disappointed 33 · Sad 12
Most "effective" emotion:
Disappointment 68 · Anger 38 · Sadness 17

Both wordings validate equally well — **no significant difference on Serious or Common (both p>.5)**, so collapsing to one final vignette (Appendix 1) is well-justified, not an arbitrary choice.
Note: participants' free-choice "most effective" emotion skewed toward Disappointment/Anger rather than Sadness at the pilot stage — worth having an answer ready if asked why sadness still won on relational outcomes in Study 1 despite this.

Backup · Bootstrap robustness check (Study 2)

BCa BOOTSTRAP (R=10,000) vs. PARAMETRIC OLS — CEILING EFFECTS DON'T CHANGE CONCLUSIONS

Outcome (Anger vs No-Response)	OLS 95% CI	Bootstrap BCa 95% CI
Appropriateness	[1.71, 2.54]	[1.63, 2.58]
Effectiveness	[1.52, 2.71]	[1.50, 2.73]
Agency	[3.50, 4.51]	[3.49, 4.45]
Likability	[−1.14, −0.33]	[−1.16, −0.32]

Method: case-resampling bootstrap (10,000 draws with replacement) on the Group main-effect OLS models, bias-corrected & accelerated (BCa) intervals — makes no normality assumption, which matters given the ceiling effects in Appropriateness/Likability.
Bottom line: every bootstrap CI overlaps its parametric counterpart almost exactly. The regression results in the paper are not an artifact of a normality assumption that doesn't hold.