Beyond Stimulus Sampling: A Harness for Designing and Auditing Experimental Stimuli

David Hagmann[1], George Loewenstein[2], and Zachary Wojtowicz[3]

[1]Department of Management, The Hong Kong University of Science and Technology

[2]Department of Social and Decision Sciences, Carnegie Mellon University

[3]Massachusetts Institute of Technology

Author Note

David Hagmann https://orcid.org/0000-0002-2080-997X

Correspondence concerning this article should be addressed to David Hagmann, Department of Management, The Hong Kong University of Science and Technology, Hong Kong. Email: hagmann@ust.hk

Abstract

Experiments using hypothetical, researcher-generated scenarios and messages are common in organizational research. But writing stimuli that meaningfully manipulate a construct is challenging, and successful manipulations may unintentionally change other relevant attributes. We introduce a software harness for large language models (LLMs) to construct and audit such materials in a transparent and reproducible way. Our harness generates message families that introduce variation in a focal attribute of interest, evaluates intended contrasts and construct spillover, and revises materials that fail prespecified criteria. The workflow and resulting documentation can be adapted to different constructs and organizational settings and is agnostic to the LLM used. We illustrate the application with a design examining whether expressing uncertainty increases persuasiveness, showing the problem of construct spillover and how the harness supports the development of more selective manipulations.

Keywords: generative artificial intelligence, stimulus sampling, construct validity, experimental methods

Beyond Stimulus Sampling: A Harness for Designing and Auditing Experimental Stimuli

Organizational scientists design experiments with hypothetical scenarios, descriptions of people, and open-ended messages to study a wide range of topics. Participants in such research may judge the fairness of a manager who provides feedback to an employee, report whether they would trust a coworker, or assess whether an organization acted fairly (Aguinis & Bradley, 2014). Related designs are widely used in psychology, management, political science, and other fields (Chester & Lasko, 2021; Fong & Grimmer, 2023). Their value depends on whether the materials convey the intended difference strongly enough to support a meaningful test. A researcher asking whether certainty affects persuasion, for example, must write messages that readers perceive as differing meaningfully in certainty. To test whether a message influenced the dimension of interest, researchers can rely on pretesting and manipulation checks.

Even when that contrast is strong, however, other relevant features may also change. A more certain author may present stronger evidence, dismiss disagreement, or seem less willing to compromise. We term this *construct spillover*: the intended construct changes along with additional features the experimenter did not intend to vary. When readers respond differently, the experiment cannot establish which attribute caused the difference: the intended focal attribute or an unintended spillover feature. Manipulation strength and construct isolation therefore require separate assessments.

A single stimulus per treatment creates further vulnerability because a finding may be explained not by the variation of interest but by incidental wording that may be difficult to detect even when materials are public. Stimulus sampling offers one way to reduce this concern. In this approach, researchers generate multiple stimuli for each experimental treatment to examine relationships across multiple messages (Judd et al., 2012; Wells & Windschitl, 1999). A stimulus plot can then show whether an effect is driven by a single outlier stimulus, for example. However, this does not guarantee a meaningful target contrast or the absence of systematic differences in other constructs that could persist across multiple stimuli.

Generative AI can help address both challenges by producing and revising many versions

of a stimulus. We introduce a software *harness* that coordinates LLM calls and researcher decisions from construct definition through generation, evaluation, and revision.[1] We focus here on text-based stimuli and provide a complete recipe: prompts, assessment criteria, revision rules, and an audit report. Researchers adapt it by interactively specifying their question, context, and criteria as part of a chat, without rewriting software. Building on AI-assisted stimulus development (Lechuga & Karle, 2026; van Berlo et al., 2024), this reusable research infrastructure applies across constructs and language models. We illustrate it with two experiments of certainty and persuasion on politically charged topics.

Generating and Evaluating Stimuli

Many constructs that are instantiated via written text involve a package of linguistic cues and are therefore difficult to define precisely. Consider a hypothetical design in which we want to vary a manager's certainty when advising an employee to apply for the promotion. Panel A of Table 1 provides a high-certainty reference message. Panel B hedges both feedback and advice. Panel C preserves specific accomplishments but qualifies the praise and sounds less supportive. Panel D preserves the positive performance assessment but expresses uncertainty in the advice to apply for the promotion. Which message is appropriate requires researcher judgment: weaker support may belong to one definition of an uncertain recommendation and constitute spillover under another. Researchers must specify intended contrasts and which attributes should be held fixed, a laborious task for one stimulus and difficult to scale when sampling dozens.

These alternatives show why one-shot prompting cannot establish stimulus validity: messages B-D all vary uncertainty but in different ways. Structured prompts can help generate stimuli and identify confounds (Simonsohn et al., 2025), but variation and construct spillover still need independent evaluation. To illustrate this challenge, we present unpublished research on certainty and persuasion. ChatGPT generated five supporting points for each position on whether gun sales without background checks are common, then wrote essays consisting of five

[1] The harness is available via Github, along with the data and code to reproduce the analyses in this paper: https://github.com/dhagmann/beyond-stimulus-sampling.

Table 1

Alternative expressions of uncertainty in advice about applying for the promotion. Panel A provides the high-certainty reference message. Panel B hedges both feedback and advice. Panel C qualifies the praise and sounds less supportive. Panel D keeps the praise but expresses uncertainty in the advice to apply for the promotion.

A. High certainty (Reference). On the last project, you kept the team on schedule even when unexpected problems arose. Your regular updates helped everyone understand what needed to happen next. Those results show that you can take on greater responsibility. Your performance is exceptional and you are ready for a promotion. You should put yourself forward.	**B. Hedging feedback and recommendation.** On the last project, you appear to have kept the team on schedule even when unexpected problems arose. Your regular updates seem to have helped everyone understand what needed to happen next. Those results make it seem like you can take on greater responsibility. Your performance seems to be exceptional and you may be ready for a promotion. Maybe you could put yourself forward.
C. Reduced praise and support. On the last project, you kept the team on schedule even when unexpected problems arose. Your regular updates helped everyone understand what needed to happen next. Those results may be a sign that you can take on greater responsibility. I suspect that your performance is exceptional and perhaps you are ready for a promotion. I am unsure, but maybe you could think about putting yourself forward if you think it is a good idea.	**D. Hedging recommendation.** On the last project, you kept the team on schedule even when unexpected problems arose. Your regular updates helped everyone understand what needed to happen next. Those results suggest you could take on greater responsibility. Your performance is exceptional and I think that you may be ready for a promotion. You may want to consider putting yourself forward.

paragraphs using those points. We used four messages, one per certainty condition (high and low certainty) and argument direction (a high share or a low share).

We then recruited 600 participants from Prolific who estimated the share of guns sold without a background check before and after reading either the high or low certainty message, arguing in a direction against their existing view. Readers rated the author of the high certainty message as more certain than that of the low certainty message (4.49 vs. 4.02 on a five-point scale, $t(590.96) = 8.36$, $p < .001$), showing successful manipulation of the construct. The key finding is that those who read the less-certain message were more persuaded (12.3 vs. 9.4 percentage points of belief change, $t(596.11) = 2.08$, $p = .038$. See Figure 1, Panel A).

A closer reading of the messages, however, reveals a broader contrast beyond certainty. The high-certainty messages also dismissed opposing views, referring, for example, to "stubborn myths circulating among the misinformed."[2] We describe this interpersonal stance as *brashness*. Random assignment identified an effect of the messages, but could not separate certainty from their treatment of those with disagreeing views.

The harness makes these distinctions explicit before generation: researchers define the construct and context, then classify related dimensions proposed by the model as intended features or spillovers. For receptiveness, for example, hedging and acknowledging agreement might be intended features (Yeomans et al., 2020), while concessions and praise might not. For our persuasion application, the harness builds pools of arguments and argumentation styles. Styles describe how an argument is organized, such as opening with an example. The harness then randomly samples from both pools. Each draw is used for every condition, forming a *stimulus family*: messages that share a position, arguments, and style while varying the intended constructs. Repeated draws introduce variation across families without systematically tying content or style to a treatment.

The process then draws on separate writer and judge subagents that use separate contexts to limit copying between writers and anchoring between judges. A coordinating agent supplies

[2] See Appendix A for one high- and low-certainty message pair.

writers with shared inputs and assigned conditions, then passes their stimuli to judges without treatment labels or other judges' ratings. It applies the criteria to returned ratings and concerns, requests revisions if necessary, and has the complete family reassessed prior to approval (see Figure 2). Instructions, drafts, ratings, and decisions are saved to an audit report before participant data collection (see Appendix C).

Returning to our experiment on uncertainty and persuasion, we can now easily and independently vary certainty and brashness across two topics: gun sales and regret following an abortion. Within each topic and each argument direction (high and low estimate), we generated 15 arguments and 8 argumentation styles. Each essay family consisted of four messages (high/low certainty, high/low brashness) that used the same five arguments and argumentation style from this set. We generated 20 families for each combination of topic and argument direction: 80 families with four messages each, for a total of 320 messages (see Table 2 for excerpts from one such family, with the full essays shown in Appendix B). We then recruited 3,185 participants from Prolific who again made percentage estimates, read one counter-attitudinal message, and provided a second estimate. Now, authors expressing high certainty were more persuasive than uncertain authors (12.5 vs. 10.2 percentage points, $t(3157.81) = -3.32$, $p < .001$), and brashness led to less persuasion (10.6 vs. 12.0 points, $t(3182.90) = -1.96$, $p = .051$. See Figure 1, Panel B).

Table 2

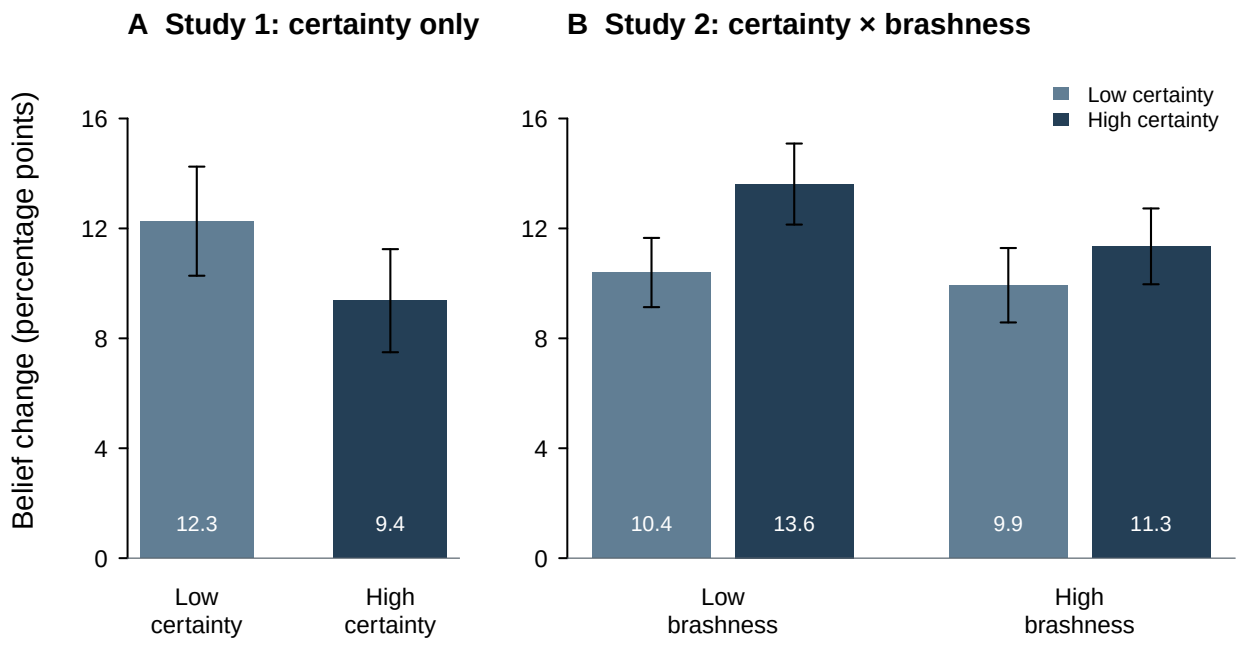
A Stimulus Family Independently Varying Certainty and Brashness

Certainty	Low brashness	High brashness
Low	One reason for this, it appears, is a gap in federal law that arguably leaves a wide category of transactions unscreened. Under current federal statute, private sales between individuals — meaning sales not conducted through a licensed dealer — are evidently not required to involve a background check. […] Reasonable people may weigh this evidence differently, and I respect that.	Only a limited number of states have stepped in to enact universal background check laws of their own. That means residents of the majority of states can, it appears, legally purchase firearms from private sellers with no screening whatsoever. If you think this is a minor footnote, you're kidding yourself—perhaps willfully so.
High	The most significant gap is, without question, the federal law governing private sales. Federal law does not require background checks when a firearm is sold between two private individuals rather than through a licensed dealer. […] People who believe otherwise are not foolish; they are simply working with an incomplete picture. The complete picture, however, leaves no room for doubt.	The reality is that federal law does not require background checks for private sales between individuals. This is not a matter of interpretation or debate; it is established legal fact. […] If you still think few guns escape background checks, you are clinging to an illusion that ignores what the law plainly allows.

Note. Excerpts from one four-message family used in Study 2. The full messages are shown in Appendix B.

Figure 1

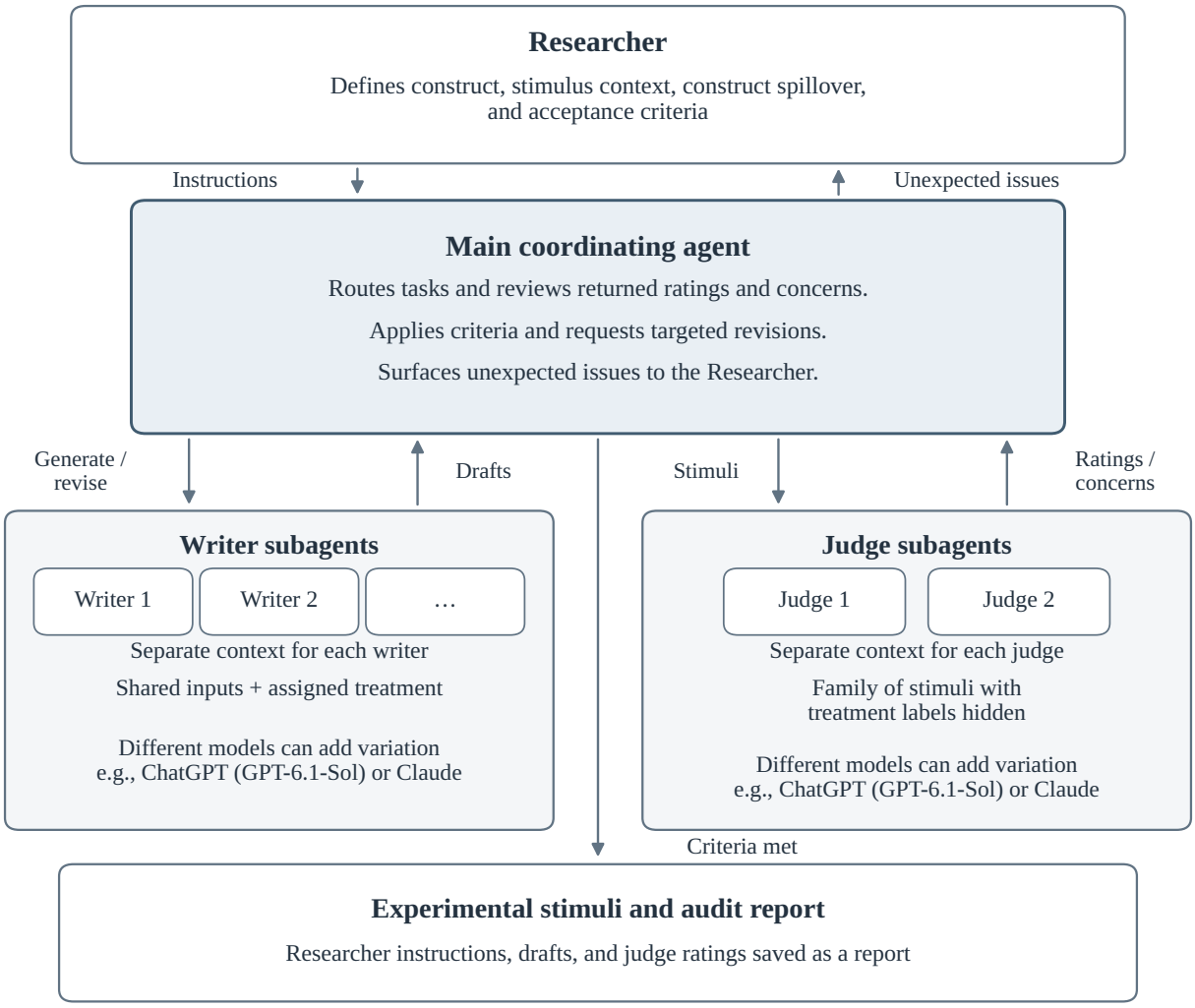
Certainty and Persuasion Before and After Separating Brashness



Note. Bars show mean belief change toward the message's position, in percentage points. Error bars show 95% confidence intervals. Panel A: Study 1, N = 600. Panel B: Study 2, N = 3,185, pooled across topics.

Figure 2

Human and Agents in the Harness



Note. The coordinator routes work between independent subagents, requests revisions, and raises unexpected issues for researcher review.

Discussion

Designing text stimuli for experiments is an art with substantial challenges: messages have to meaningfully shift perceptions on a focal construct while holding fixed potential confounds from spillover. Stimulus sampling can reduce dependence on particular wording, but it does not ensure that the intended construct is isolated. Our persuasion studies illustrate this distinction: the initial finding that low certainty messages are more persuasive did not hold once we orthogonally varied brashness. The harness we introduce here generates controlled stimulus families that closely follow one another within one set but offer meaningful variation across families, then audits and revises them using prespecified criteria for intended contrasts and construct spillover.

The next development phase will evaluate agreement between LLM judges' assessments and independent human judgments. The central risk is that judges may accept contrasts readers do not perceive or miss relevant spillover. We will also examine performance across organizational settings and stimulus types beyond persuasive essays. For example, the workflow could separate candor from disrespect in performance feedback, or empathy from apology in responses to employee concerns, with trust, coordination, or willingness to speak up as outcomes.

LLM responses and ratings can vary across runs, and models change over time. As with researchers writing stimuli or assistants coding them, individual outputs may not be exactly reproducible. However, documenting the workflow supports conceptual replication by letting other researchers and models follow the same process without having to try and reconstruct the original researcher's reasoning that may not have been documented.

The harness makes stimulus development easier by guiding key decisions and automating generation, screening, revision, and documentation. Researchers can generate many varied families and inspect a short audit report (see Appendix C for an example). As reusable research infrastructure, the package lets others inspect these decisions and adapt or rerun the workflow, requiring less manual drafting and recordkeeping and facilitating conceptual replications.

References

Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, *17*(4), 351–371. https://doi.org/10.1177/1094428114547952

Chester, D. S., & Lasko, E. N. (2021). Construct validation of experimental manipulations in social psychology: Current practices and recommendations for the future. *Perspectives on Psychological Science*, *16*(2), 377–395. https://doi.org/10.1177/1745691620950684

Fong, C., & Grimmer, J. (2023). Causal inference with latent treatments. *American Journal of Political Science*, *67*(2), 374–389. https://doi.org/10.1111/ajps.12649

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*(1), 54–69. https://doi.org/10.1037/a0028347

Lechuga, J., & Karle, N. N. (2026). Generating experimental text stimuli for psychological research using ChatGPT. *Advances in Methods and Practices in Psychological Science*, *9*(1). https://doi.org/10.1177/25152459251407631

Simonsohn, U., Montealegre, A., & Evangelidis, I. (2025). Stimulus sampling reimagined: Designing experiments with Mix-and-Match, analyzing results with stimulus plots. *Journal of Personality and Social Psychology*, *129*(1), 71–90. https://doi.org/10.1037/pspa0000449

van Berlo, Z. M. C., Campbell, C., & Voorveld, H. A. M. (2024). The MADE framework: Best practices for creating effective experimental stimuli using generative AI. *Journal of Advertising*, *53*(5), 732–753. https://doi.org/10.1080/00913367.2024.2397777

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin*, *25*(9), 1115–1125. https://doi.org/10.1177/01461672992512005

Yeomans, M., Minson, J., Collins, H., Chen, F., & Gino, F. (2020). Conversational receptiveness: Improving engagement with opposing views. *Organizational Behavior and Human Decision Processes*, *160*, 131–148. https://doi.org/10.1016/j.obhdp.2020.03.011

Appendix A

Study 1 Message Pair

These two messages were used in Study 1 and argue that few gun sales occur without a background check. The study also included the opposite argument direction.

High certainty

In the heated debate about gun control, it's time to lay down the facts and dispel the stubborn myths circulating among the misinformed. Let's get one thing straight: the idea that a vast majority of gun purchases are made without background checks is not just exaggerated—it's outright false.

Federal law is clear and uncompromising in its mandate that all federally licensed firearms dealers must conduct rigorous background checks on potential buyers before any gun sale can be completed. This isn't a suggestion; it's a strict requirement enforced across the nation, ensuring that criminals and those deemed unfit to own firearms are systematically weeded out.

But let's not stop there. Many states, dissatisfied with merely following federal guidelines, have taken matters into their own hands to fortify these regulations. These states extend the reach of background checks to include private sales and transfers, effectively closing gaps that might be exploited by those looking to circumvent the law. Critics often paint a picture of lax regulations, but the reality is that state laws have built a formidable barrier against unauthorized gun ownership, demonstrating a commitment to safety and responsibility that extends far beyond federal directives.

Opponents often hinge their arguments on outdated notions of gun owners and sellers as renegades who flout the law, but the current landscape of gun ownership is characterized by high levels of compliance and awareness. Responsible gun owners and dealers are not the caricatures portrayed by gun control zealots; they are, in fact, some of the most stringent adherents to legal stipulations, often going above and beyond to ensure the integrity of every transaction.

The myth of the "gun show loophole" is another favorite scare tactic used to inflame public opinion. Yes, private sellers at gun shows aren't required by federal law to conduct

background checks, but here's what the fearmongers won't tell you: the majority of sellers at these events are licensed dealers who are, indeed, bound by law to run background checks. The narrative of a lawless marketplace is a gross distortion of reality, peddled by those who have either failed to understand the facts or choose to ignore them for the sake of political expediency.

Finally, let's ground this discussion in hard evidence. Numerous surveys and studies have consistently shown that the vast majority of gun transactions are conducted with full compliance to background check requirements. This isn't conjecture; it's concrete data debunking the alarmist rhetoric that too often dominates the conversation around gun sales. Those clinging to the myth of rampant unchecked gun purchases are either woefully uninformed or deliberately deceptive. It's high time they are confronted with the truth and called out for their contributions to a baseless, fear-driven dialogue that distorts public understanding and policy.

Low certainty

When discussing the topic of firearms and background checks in the United States, it's essential to recognize the genuine concerns that arise from the possibility of guns being purchased without proper checks. However, it's also crucial to consider the substantial measures already in place that ensure the safety and legality of gun transactions.

Under federal law, all firearms dealers holding a Federal Firearms License (FFL) are required to perform background checks on potential buyers before completing a sale. This mandate covers the vast majority of gun sales in the country, ensuring that individuals who are prohibited from owning firearms—such as those with certain criminal convictions or mental health issues—are unable to legally purchase guns. In addition to federal requirements, many states have taken further steps to tighten regulations. These state laws often extend the necessity for background checks to private sales and transfers, including those conducted at gun shows and through private parties. This layer of regulation helps to close gaps that might otherwise exist, significantly reducing the number of firearms transactions that occur without a background check.

The effectiveness of these laws is reflected in the high level of compliance among both gun sellers and buyers, most of whom are committed to legal and responsible gun ownership.

This responsible majority understands the importance of background checks in maintaining public safety and preventing crime, and they adhere rigorously to these laws.

It’s important to address concerns regarding gun shows and online sales, often pointed to as loopholes in the system. While it’s true that these have been avenues for unchecked sales in the past, changes in both public awareness and state regulations have greatly diminished these risks. Many sellers at gun shows are indeed licensed dealers who conduct mandatory background checks, and online gun sales typically require the buyer to complete the transaction through a licensed dealer, thus undergoing a background check.

Surveys and research studies consistently show that the majority of gun transactions in the U.S. involve background checks. Estimates suggest that the percentage of sales made without background checks is quite small when compared to the overall volume of gun sales. This data is encouraging because it demonstrates a broad adherence to laws designed to keep our communities safe.

In conclusion, while the concerns about guns being purchased without background checks are valid, the reality is that a comprehensive system of laws and regulations is effectively keeping such occurrences to a minimum. Through continuous public education and legal enforcement, we can further ensure that the sale and ownership of firearms are conducted safely and legally.

Appendix B

Full Messages From the Reference Family

This appendix reproduces the four complete messages excerpted in Table 2. The family takes the position that many guns sold in the United States are sold without a background check. The versions share the same argument inputs and vary in assigned certainty and brashness.

Low certainty, low brashness

It is entirely reasonable to believe that the background check system in the United States covers most gun sales, and thoughtful people hold this view for understandable reasons. The National Instant Criminal Background Check System is, after all, a real and functioning institution, and licensed dealers do run checks on millions of transactions every year. That is genuinely worth recognizing. However, when one looks more closely at the full landscape of how firearms change hands in this country, it seems possible that the picture is more complex than it first appears, and that many sales may fall outside the reach of this system entirely.

One reason for this, it appears, is a gap in federal law that arguably leaves a wide category of transactions unscreened. Under current federal statute, private sales between individuals — meaning sales not conducted through a licensed dealer — are evidently not required to involve a background check. This is not a fringe loophole or a contested legal theory; it seems to be a well-documented feature of the existing regulatory framework. It is possible that many Americans are simply unaware of this distinction, and one can hardly blame anyone for assuming that all gun sales go through the same process. The reality, however, may be that a significant number of firearms are transferred person to person with no screening whatsoever.

This gap appears to be compounded by the fact that only a limited number of states have enacted universal background check laws on their own. Residents of the majority of states, it seems, can legally purchase firearms through private channels without any screening. The patchwork nature of state-level regulation perhaps means that the experience of buying a gun varies dramatically depending on where one lives, and for many Americans, the background check requirement may simply not apply to a private purchase.

Gun shows arguably represent one of the more visible settings where this dynamic plays out. Many sellers at these events appear to operate without Federal Firearms Licenses, which would seem to mean they can complete transactions without conducting background checks. The ATF has itself acknowledged, it seems, the difficulty of enforcing background check requirements on private sales, since these transactions leave little to no paper trail and are nearly impossible to monitor. This is not a criticism of the hardworking people at these agencies — they face a genuinely difficult task — but it does suggest that enforcement may be structurally limited in ways that matter.

Beyond the legal private market, there also appear to be black market and underground channels through which firearms are transferred entirely outside any regulatory framework. These sales, by their very nature, involve no screening of any kind. When one considers all of these pathways together — the federal private sale exemption, limited state-level universal check laws, gun show transactions, enforcement challenges, and underground markets — it seems plausible that many guns sold in the United States do change hands without a background check. Reasonable people may weigh this evidence differently, and I respect that. But the full picture, tentatively considered, appears to point toward a system with more gaps than many of us might initially assume.

High certainty, low brashness

It is completely understandable to believe that the background check system in the United States effectively covers most gun sales, and thoughtful people hold this view for good reasons. After all, licensed firearms dealers are undeniably required to conduct background checks on every transaction, and the NICS system processes millions of these checks each year. That is a real and meaningful part of the picture, and anyone who points to it is drawing on solid ground. However, the full picture clearly reveals that many firearms change hands in this country without any background check whatsoever, through multiple well-documented channels that operate legally and illegally outside the reach of that system.

The most significant gap is, without question, the federal law governing private sales.

Federal law does not require background checks when a firearm is sold between two private individuals rather than through a licensed dealer. This is not a contested claim or a matter of interpretation — it is an established legal reality. Only a limited number of states have enacted universal background check laws to close this gap, which means that residents of the majority of states can legally purchase firearms from private sellers with no screening at all. Reasonable people may assume that federal law covers all sales uniformly, and it is easy to see why that assumption feels natural, but the evidence proves otherwise.

Gun shows serve as a clear illustration of how this gap functions in practice. At these events, many sellers operate without Federal Firearms Licenses, meaning they are legally classified as private sellers. Transactions at their tables can be completed without any background check requirement. These are not small or hidden marketplaces — they are well-attended, widely advertised, and represent a major channel through which firearms are transferred outside the screening system. The reality of what happens at these events is thoroughly documented.

The Bureau of Alcohol, Tobacco, Firearms and Explosives has itself acknowledged the difficulty of enforcing background check requirements on private sales. These transactions leave little to no paper trail and are nearly impossible to monitor. This is not a speculative concern; it is an operational reality that the agency charged with enforcement has openly recognized. Without documentation, without records, and without oversight, there is no mechanism to ensure that private buyers have been screened. The enforcement challenge is undeniable.

Beyond these legal private sales, black market and underground gun sales operate entirely outside any regulatory framework. These transactions represent an additional and substantial channel through which firearms are transferred without any screening whatsoever. No background check, no record, no oversight of any kind touches these sales. When we consider the full landscape — the federal private sale exemption, the limited adoption of universal background check laws across states, the vibrant marketplace of gun shows, the enforcement challenges the ATF faces, and the entirely unregulated underground market — the conclusion is obvious. Many guns sold in the United States are sold without a background check. People who believe

otherwise are not foolish; they are simply working with an incomplete picture. The complete picture, however, leaves no room for doubt.

Low certainty, high brashness

It seems reasonable, at first glance, to believe that the background check system in the United States catches most firearm sales in its net—and if you hold that view, you're not entirely wrong about the mechanics of licensed dealer transactions. Federal Firearms Licensees are indeed required to run checks, and that part of the system appears to function more or less as intended. But if you think that accounts for the full picture, you've frankly been fooling yourself. The evidence suggests that the reality is far messier, and anyone still clinging to the comforting notion that "few" guns slip through without screening probably hasn't bothered to look at how firearms actually change hands in this country.

Consider, for a moment, what federal law actually requires—and more importantly, what it does not. Private sales between individuals are not, under federal statute, subject to background check requirements. This is not some fringe loophole that critics have invented; it is a well-documented gap in the regulatory framework. Only a limited number of states have stepped in to enact universal background check laws of their own. That means residents of the majority of states can, it appears, legally purchase firearms from private sellers with no screening whatsoever. If you think this is a minor footnote, you're kidding yourself—perhaps willfully so.

Gun shows offer a particularly telling illustration of this gap, and anyone dismissing their significance hasn't thought it through. These events serve as major marketplaces where many private sellers operate without Federal Firearms Licenses. Transactions at these venues can seemingly be completed with nothing more than a handshake and cash. No form, no phone call, no check. It is possible that some defenders of the current system will wave this away as insignificant, but the sheer volume of firearms that arguably move through these channels makes that dismissal look naïve at best.

The enforcement picture appears equally sobering. The ATF has itself acknowledged the difficulty of policing private sales, since these transactions leave little to no paper trail and seem

nearly impossible to monitor. Without records, without oversight, the system may well be blind to a substantial portion of transfers. And that is before one even considers black market and underground sales, which operate entirely outside any regulatory framework. These channels represent an additional avenue through which firearms are apparently transferred with no screening whatsoever. Pretending this shadow market is negligible is the kind of head-in-the-sand thinking that deserves no patience.

So where does this leave the confident assumption that few guns are sold without background checks? In tatters, it would seem. Between the federal private-sale exemption, the patchwork of state laws, the realities of gun show commerce, and the underground market, there appear to be many avenues through which firearms change hands unscreened. The evidence, taken together, suggests that the background check system—however well it functions for licensed dealers—arguably misses a significant share of gun sales. If you've been telling yourself otherwise, it is perhaps time to stop repeating comfortable fictions and start reckoning with what the regulatory landscape actually looks like.

High certainty, high brashness

It is an undeniable fact that many guns sold in the United States change hands without any background check whatsoever, and if you believe otherwise, you simply haven't been paying attention. Now, let's be fair for a moment: there is something the opposing view gets right. Licensed firearms dealers are indeed required to run background checks through the National Instant Criminal Background Check System, and millions of transactions every year do go through this process. That much is true, and no one disputes it. But if you stop there and conclude that few guns are sold without screening, you have made a foolish mistake — one that ignores the gaping holes in the system that are clearly documented and well understood by anyone who has bothered to look.

The reality is that federal law does not require background checks for private sales between individuals. This is not a matter of interpretation or debate; it is established legal fact. When one private citizen sells a firearm to another, no federal statute compels either party to

conduct a screening. Only a limited number of states have enacted universal background check laws to close this gap, which means residents of the majority of states can legally purchase firearms from private sellers with zero screening. If you still think few guns escape background checks, you are clinging to an illusion that ignores what the law plainly allows.

Gun shows serve as major marketplaces where this private sale gap plays out on a massive scale, and anyone who dismisses this hasn't thought the issue through. Many sellers at these events operate without Federal Firearms Licenses, which means they are under no obligation to run a background check before completing a transaction. These are not rare or marginal occurrences. They happen at events across the country, week after week. The ATF itself has acknowledged the difficulty of enforcing background check requirements on private sales, since these transactions leave little to no paper trail and are nearly impossible to monitor. There is no doubt about this — the agency charged with enforcement has openly admitted the challenge.

Beyond what happens in the legal marketplace, black market and underground gun sales operate entirely outside any regulatory framework. These transactions represent an additional and substantial channel through which firearms are transferred without any screening whatsoever. Stolen guns, straw purchases funneled into illegal networks, and person-to-person deals conducted in back alleys and parking lots — none of these involve a background check, and it is obvious to anyone thinking clearly that pretending these channels are negligible is deeply irresponsible.

The evidence proves, without question, that many firearms sold in this country bypass the background check system entirely. Legal private sales, gun show transactions, and underground markets collectively ensure that a significant number of guns reach buyers with no screening at all. If you walked into this believing that few guns are sold without a background check, it is time to abandon that comfortable fiction. The facts are staring you in the face, and continuing to ignore them does not make them go away. Anyone still holding onto the notion that background checks capture nearly all gun sales ought to know better — because the truth is right there, plainly visible, for anyone willing to see it.

Appendix C

Example Stimulus Audit

Construct and spillover audit

Topic: AI data centers' share of U.S. office construction spending; higher estimate.

Certainty/brashness checks: PASS **Naturalness: REVIEW**

Manipulated: certainty × brashness. **Fixed:** arguments and factual content.

Judge ratings (1–7)

J1: GPT-6.1-Sol; J2: Claude Opus 5.5.

Assigned certainty / brashness	Certainty	Brashness	Naturalness
Low / Low	2 / 2	1 / 1	5 / 4
Low / High	2 / 2	6 / 5	6 / 5
High / Low	7 / 7	1 / 2	6 / 5
High / High	7 / 7	6 / 6	6 / 5

Additional construct ratings: J1 (1–7)

Assigned certainty / brashness	Warmth	Competence	Trust
Low / Low	5	5	6
Low / High	3	5	5
High / Low	5	6	5
High / High	2	5	4

Intended contrasts and spillover

Change (other factor fixed)	Target mean Δ ≥ 1.5	Spillover |mean Δ| ≤ 1.5
Certainty (brashness low)	5.0	0.5
Certainty (brashness high)	5.0	0.5
Brashness (certainty low)	4.5	0.0
Brashness (certainty high)	4.5	0.0

Other checks

Check	Observed	Rule
Content fidelity	7.0	≥ 5.5
Content equivalence (family)	7.0	≥ 5.0
Fluency	6.5	≥ 5.0
Coherence	6.5	≥ 5.0
Naturalness	4.5	≥ 5.0
Words per message	294–305	270–350
Word-count spread	11	≤ 30