

Running Head: 'Cheating' vs. 'Being a Cheater'

When Cheating Would Make You a Cheater:

Noun Wording Prevents Unethical Behavior

Christopher J. Bryan

Gabrielle S. Adams

Stanford University

London Business School

Benoît Monin

Stanford University

Christopher J. Bryan, Department of Psychology, Stanford University; Gabrielle S. Adams, Organisational Behaviour Area, London Business School; Benoît Monin, Department of Psychology, Stanford University

The authors thank Elizabeth Mullen, Gregory Walton and the Monin-Mullen Morality Lab for input, and Hattie Bluestone and Ashli Carter for assistance.

Correspondence should be addressed to Christopher J. Bryan, Department of Psychology, Stanford University, 450 Serra Mall, Bldg. 420, Stanford, CA 94305-2130. E-mail: cbryan@stanford.edu.

Word count (incl. abstract and main text): 2987

PAPER UNDER REVIEW – PLEASE CHECK WITH AUTHORS FOR CORRECT CITATION

Abstract

In three experiments, using two different paradigms, people were less likely to cheat for personal gain when a subtle change in phrasing framed such behavior as diagnostic of an undesirable identity. Participants were given the opportunity to claim money they were not entitled to at the experimenters’ expense; instructions referred to cheating with either a verb (e.g., “cheating”) or a predicate noun (e.g., “being a cheater”). Participants in the verb condition claimed significantly more money than participants in the noun condition, who showed no evidence of having cheated at all. This difference occurred both in a face-to-face interaction (Experiment 1) and in the private context of an online experiment (Experiments 2 and 3). These results demonstrate the power of a subtle linguistic difference to prevent even private unethical behavior by invoking people’s desire to maintain a self-image as good and honest.

When Cheating Would Make You a Cheater:

Noun Wording Prevents Unethical Behavior

Think of a number from 1 to 10. Imagine that, before you reveal it, we tell you we are studying the prevalence of cheating and will give you \$5 if your number is even. If you thought of an odd number (as most people do), would you tell us? Would you be more honest if, instead of studying cheating, we told you we were studying the prevalence of “cheaters?” In this paper, we propose that such subtle linguistic cues can influence ethical decisions by invoking identity concerns.

Specifically, we focus on the implications of framing behavior as reflecting one's identity for ethical decision-making. A long tradition of research in moral psychology demonstrates that individuals who are motivated to engage in unethical behavior deploy strategies that weaken the behavior-identity link (e.g., Bandura, 1999; Mills, 1958). To reconcile their unethical behavior with their desire to see themselves as good and ethical (Blasi, 1980; Dunning, 2005; Monin & Jordan, 2009; Steele, 1988), people use biased reasoning to sever the link between acting unethically and being unethical (Ayal & Gino, 2011; Mazar, Amir & Ariely, 2008). That is, people downplay the seriousness of their ethical lapses and tell themselves that occasional instances of cheating do not make one a dishonest person. In doing so, people can engage in dishonest behavior while avoiding the correspondent inference (Jones & Nisbett, 1971; Ross, 1977) that they are the *kind* of person who behaves dishonestly, allowing them to have their cake (reap the benefits of unethical behavior) and eat it too (preserve a positive self-image).

It follows from this analysis that one way to decrease the incidence of unethical behaviors might be to strengthen the link between such behaviors and their associated

undesirable identities. In this paper, we test whether highlighting the identity implications of cheating could prevent people from cheating, using a subtle manipulation of phrasing. We referred to the act of cheating with either a self-relevant noun (e.g., “please don’t be a cheater”) or a verb (e.g., “please don’t cheat”). Noun wording frames cheating as the enactment of an identity—a reflection of the kind of person one is. We expected that framing cheating in this way would make it difficult for people to ignore the implications of this behavior for their desired view of themselves as honest.

Effects of Noun vs. Verb Wording

This noun vs. verb manipulation is inspired by previous research investigating how subtle differences in language can affect people’s perceptions of themselves and of others (Gelman & Heyman, 1999; Walton & Banaji, 2004). Noun wording signals that an attribute is representative of a person’s essential identity (Gelman, Hollanger, Star & Heyman, 2000). In one study, people rated their own preferences as stronger and more stable when instructed to describe them with nouns (e.g., “I am a Shakespeare-reader”) than with verbs (e.g., “I read Shakespeare a lot”) (Walton & Banaji, 2004). Moreover, exposure to a survey that referred to voting in an upcoming election with a noun (e.g., “How important is it to you to be a voter [vs. “to vote”] in tomorrow’s election?”) caused more people to vote the next day (Bryan, Walton, Rogers & Dweck, 2011). This suggests that noun wording signaled to participants that, by voting, they could claim a desirable identity (“voter”), motivating them to vote. Thus, noun wording ascribes symbolic significance to a behavior, suggesting it has implications for the kind of person one would be by performing it.

So far, noun wording has only ever been shown to cause approach effects (e.g., motivating voting). This leaves open the possibility that the effect is caused not by the motivation to assume an identity, but rather by a more purely cognitive “self-perception” process—that the noun wording caused participants simply to see themselves as voters and they behaved accordingly (Bem, 1972). Such an account would suggest that noun wording should always increase people’s tendency to engage in the relevant behavior. But our theory suggests the opposite prediction in the case of undesirable identities; noun wording should cause people to *avoid* the behavior.

Overview of Research and Theoretical Contributions

We tested this hypothesis in three experiments. In each, participants engaged in a game with real financial stakes where the majority of participants who did not win as much as they could have, had the opportunity to claim money they were not entitled to, knowing their individual cheating could not be discovered. We manipulated the specific wording used to refer to cheating, and predicted that participants would claim more money in the verb than in the noun conditions.

Showing this would make important theoretical contributions in several areas of psychology. First, by showing that noun wording not only increases the appeal of positive behavior (Bryan et al., 2011) but also decreases the appeal of negative behavior, it would provide support for our emerging theory that noun wording influences behavior by emphasizing its implications for identity. Second, it would provide direct empirical support for recent theoretical models asserting the importance of the self in regulating ethical behavior (Mazar, Amir & Ariely, 2008; Monin & Jordan, 2009; Zhong et al., 2010). Third, it would build on the rich tradition of research on causal attribution and

correspondent inference (Jones & Nisbett, 1971; Ross, 1977) by showing that manipulating the availability of internal (or “person”) attributions for people’s own actions—before they even happen—can affect their behavior. Finally, it would contribute to the growing literature examining ways in which small and seemingly incidental features of language have profound effects on the way we think and behave (Dils & Boroditsky, 2010; Fausey & Boroditsky, 2010; Thibodeau & Boroditsky, 2011).

Experiment 1

Participants

Participants were approached on the campus of Stanford University by a student experimenter and asked if they would be willing to participate in a 3-minute study for a chance to win \$5. Fifty-four people agreed, however participation was limited *a priori* to people who were native English speakers, which left a final sample of 53 (25 women; $M_{age}=23.09$).

Procedure

Participants were randomly assigned to either the noun or the verb condition. The manipulation was embedded in the study instructions. The content of the instructions was identical in the two conditions; the only difference was whether cheating was referred to with a verb or a noun:

“We’re interested in how common [cheating is/cheaters are] on college campuses. We’re going to play a game in which we will be able to determine the approximate [rate of cheating/number of cheaters] in the group as a whole but it will be impossible for us to know whether you’re [cheating/a cheater].”

Next, participants were asked to think of a number between 1 and 10 without revealing the number to the experimenter. Once they had thought of the number, they were told they would receive \$5 if their number was even but nothing if it was odd (Williams, Pizarro & Ariely, 2009). Participants were then asked to reveal their number, and paid (or not) as promised. We intentionally paid for even numbers because previous research has found that participants instructed to generate a random number typically show a strong bias towards odd numbers (Kubovy & Psotka, 1976) so we expected that this payment scheme would make more participants lose and face the temptation to cheat.

Results

As predicted, only a small proportion of participants in the noun condition reported having thought of an even number (6 of 26, or 23.1%) whereas this proportion more than doubled (14 of 27, or 51.9%) in the verb condition, $\chi^2(1)=4.67$, $p=0.031$, $w=.30$, (Fig. 1).

To confirm the previously documented bias toward odd numbers (Kubovy & Psotka, 1976), we approached 26 additional participants on Stanford's campus and simply asked them to think of a number between 1 and 10 with no promise of reward. Few (5 of 26, or 19.2%) thought of an odd number, a rate nearly identical to that in the noun condition, $p>.95$. Thus, it appears many participants in the verb condition misreported their number and collected money they did not deserve, but there is no evidence that anyone in the noun condition did so.

Experiment 2

In Experiment 2, we sought to rule out the possibility that the effect observed in Experiment 1 relies on the presence of another person, which may have triggered self-

presentation concerns. Experiment 2 was conducted in the more private and impersonal setting of an online study in which participants never met or expected to meet the experimenters.

Experiment 2 also used a new task (coin flipping) in which the expected outcome in the absence of cheating was more straightforward. This allowed us to interpret more directly the difference between wording conditions relative to what would be expected by chance.

Finally, whereas Experiment 1 simply evoked the verb/noun framing in the absence of any direct admonition to the participant (e.g., “It will be impossible for us to know whether you are [cheating/a cheater]”), Experiment 2 tested the bolder prediction that, even when participants in both conditions were directly asked not to cheat, noun wording (“Please don’t be a cheater”) would curb dishonesty more effectively than verb wording (“Please don’t cheat”).

Participants

Participants were members of a university-administered online participant pool who volunteered for a study advertised as being about “psychokinesis.” Eighty-eight people volunteered, however 4 did not meet the *a priori* criterion that they be native English speakers. Five additional people were excluded for having completed the experiment faster than pilot testing suggested was reasonable for a person participating in good faith (see online supplement). Thus, the final sample included 79 participants (62 women; $M_{\text{age}}=39.87$).

Procedure

Online instructions explained that a recent controversial article claimed to report the first scientific evidence for “paranormal phenomena” (a vague reference to a paper by Bem, 2011, which had received considerable media attention). Participants were told they would flip a coin 10 times and should try to influence the outcome of each toss with their minds, making the coin land “heads” as often as possible. They were told that, to ensure that they were “properly motivated,” they would receive \$1 for every toss landing heads. To forestall any perception of experimental demand to cheat, the instructions signaled that the present experimenters were skeptical that psychokinesis is real. Participants were randomly assigned to either the noun or the verb condition. The manipulation was embedded in the instructions that followed:

NOTE: Please don't [cheat/be a cheater] and report that one or more of your coin flips landed heads when it really landed tails! **Even a small [amount of cheating/number of cheaters]** would undermine the study, making it appear that psychokinesis is real.

The manipulation was also embedded in the instructions on the next page, where participants logged the outcomes of their 10 coin flips. At the top of the page, in large, red capital letters, was the message: “PLEASE DON'T [CHEAT/BE A CHEATER].”

We used the average number of heads participants claimed to have obtained to estimate cheating rates.

Results

As predicted, participants in the verb condition claimed to have obtained significantly more heads ($M=5.49$, $SD=1.25$) than those in the noun condition ($M=4.88$, $SD=1.38$), $t(77)=2.06$, $p=0.043$, $d=0.46$. Moreover, the number of heads reported, on

average, in the verb condition was significantly higher than the 5.00 that would be expected by chance, $t(38)=2.43$, $p=0.020$, $d=0.39$, suggesting that cheating occurred. The average number of heads reported in the noun condition was not different from chance, $t(39)=0.570$, $p>0.50$ (Fig. 2A).

Although we observed cheating in the verb condition, noun wording apparently eliminated cheating. Furthermore, by replicating the essential finding from Experiment 1 in a relatively anonymous setting, Experiment 2 demonstrates that noun wording can reduce cheating even when it merely raises the private specter of taking on an undesired identity.

Experiment 3

Experiment 3 replicated the design of Experiment 2, adding a baseline condition with no appeal for honesty or reference to cheating. This allowed us to test whether the verb condition had any effect relative to no message at all, and to ascertain that any difference between the noun and verb conditions resulted from decreased cheating in the noun condition and not from increased cheating in the verb condition. To further rule out impression management, we also ensured that participants would feel anonymous and disconnected from the experimenters by using an *ad hoc* sample with no relationship with the university.

Participants

Participants were users of Facebook.com who lived in the United States and clicked on an advertisement for a “Stanford web study.” One hundred fifty-four people volunteered. Of those, 131 met the *a priori* criterion that they be native English speakers and 99 also met our completion time criterion for good-faith participation (see online

supplement). The final sample comprised 99 participants (54 women; $M_{\text{age}}=22.94$). The higher rate of time-based exclusions in this study was likely due to the *ad hoc* nature of the sample: Whereas participants in Experiment 2 had registered to take part in surveys regularly and had presumably set time aside for the study, this sample consists of casual internet browsers who landed on the survey somewhat unexpectedly, clicking an ad they discovered while browsing Facebook, and who therefore were presumably less likely to have time and motivation to take the study seriously.

Procedure

The procedure was identical to Experiment 2 except that a baseline condition was added in which cheating was not mentioned.

Results

The omnibus effect of condition was significant, $F(2, 96)=4.38, p=0.015$. Using pairwise comparisons with the pooled *MSE*, we determined that participants in the verb condition claimed to have obtained significantly more heads ($M=6.22, SD=1.55$) than participants in the noun condition ($M=5.23, SD=1.18$), $t(96)=2.52, p=0.013, d=0.71$. Participants in the baseline condition also claimed to have obtained significantly more heads ($M=6.31, SD=1.72$) than participants in the noun condition, $t(96)=2.95, p=0.004, d=0.66$. The numbers of heads claimed in the verb and baseline conditions were not significantly different, $t(96)=0.25, p>0.80$.

Further, the numbers of heads claimed in both the verb and baseline conditions were significantly higher than chance, $t(36)=4.79, p<0.0005, d=0.79$ and $t(35)=4.55, p<0.0005, d=0.78$, respectively. Finally, there was no evidence of cheating in the noun

condition; the number of heads claimed in that condition was not different from chance, $t(25)=1.00, p>0.30$ (Fig. 2B).

Discussion

In three studies, we showed that simply using a noun rather than a verb to refer to unethical behavior curbed cheating significantly. In Experiment 1, participants in the noun condition were half as likely to say they had thought of a winning number as those in the verb or baseline conditions. In Experiments 2 and 3, participants in the noun condition reported flipping coins at chance, whereas those in the verb and baseline conditions reported above-chance earnings. These effects obtained in face-to-face interaction (Experiment 1), in a university online subject pool (Experiment 2), and even when we recruited strangers online through ads placed on Facebook.com (Experiment 3). In all three studies our subtle manipulation yielded sizeable effects ($w = 0.30$ in Experiment 1, $d = 0.46$ in Experiment 2, and $d = 0.71$ in Experiment 3).

One intriguing finding is that direct instructions not to cheat that used verb wording were completely ineffective. In Experiment 3, participants cheated to the same degree in the verb condition as they did in the baseline condition, where there was no appeal for honesty. But a simple shift to noun wording appears to have eliminated cheating completely. It is perhaps not surprising that the verb-based appeal had no effect in such an anonymous setting. In this context, the most salient rationale for honesty in the verb condition may have been that someone participants had never met and had no reason to care about was asking them not to cheat. But the noun-based appeal shifted the rationale to something most people care a great deal about: cheating would say something about *them*. It is fascinating to consider that institutions may unwittingly greatly

moderate the effectiveness of such admonitions simply by choosing noun or verb wording (e.g., "Please don't litter" vs. "Please don't be a litterbug"; "Don't drink and drive" vs. "Don't be a drunk driver"). Awareness of the effect documented here holds the promise of increasing the effectiveness of appeals for pro-social behavior at little cost.

While the potential of noun wording to reduce the incidence of unethical behavior in society is exciting, it is important to consider a possible risk our theory suggests might be associated with such techniques. Because noun wording signals that cheating has implications for identity, it is unclear what the effect might be on someone who is exposed to this treatment and then goes on to cheat anyway. Such a person might come to see being a "cheater" as part of his or her identity (Miller, Brickman & Bolen, 1975) and be *more* likely to cheat in the future.

In conclusion, these findings add to an emerging perspective suggesting that the self plays a central role in governing ethical behavior. Further, this effect demonstrates how even subtle linguistic cues can prevent dishonesty by harnessing people's desire to maintain a view of themselves as ethical and honest. This suggests the potential for simple interventions to help curb dishonest behavior in our society.

References

- Ayal, S., & Gino, F. (2011). Honest rationales for dishonest behavior. In M. Mikulincer & P. R. Shaver (Eds.), *The Social Psychology of Morality: Exploring the Causes of Good and Evil*. Washington, DC: American Psychological Association.
- Bandura, A. (1999). Moral disengagement in the preparation of inhumanities. *Personality and Social Psychology Review*, 3, 193-209. doi:10.1207/s15327957pspr0303_3
- Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.). *Advances in experimental social psychology* (Vol. 6, pp. 1–62). New York: Academic Press.
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100, 407-425. doi:10.1037/a0021524
- Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin*, 88, 1-45. doi:10.1037/0033-2909.88.1.1
- Bryan, C. J., Walton, G. M., Rogers, T., & Dweck, C. S. (2011). Motivating voter turnout by invoking the self. *Proceedings of the National Academy of Sciences, USA*, 108, 12653-12656. doi:10.1073/pnas.1103343108
- Dils, A. T., & Boroditsky, L. (2010). Visual motion aftereffect from understanding motion language. *Proceedings of the National Academy of Sciences, USA*, 107, 16396–16400. doi:10.1073/pnas.1009438107
- Dunning, D. (2005). *Self-insight: Roadblocks and detours on the path to knowing thyself*. New York: Psychology Press.

Fausey, C. M., & Boroditsky, L. (2010). Subtle linguistic cues influence perceived blame and financial liability. *Psychonomic Bulletin & Review*, *17*, 644–650.

doi:10.3758/PBR.17.5.644

Gelman, S. A., & Heyman, G. D. (1999). Carrot-eaters and creature-believers: The effects of lexicalization on children's inferences about social categories.

Psychological Science, *10*, 489-493. doi:10.1111/1467-9280.00194

Gelman, S. A., Hollander, M., Star, J., & Heyman, G. D. (2000). The role of language in the construction of kinds. In D. Medin (Ed.), *Psychology of learning and motivation* (Vol. 39, pp. 201–263). New York: Academic Press.

Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E.

Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79 –94). Morristown, NJ: General Learning Press.

Kubovy, M., Psotka, J., (1976). The predominance of seven and the apparent spontaneity of numerical choices. *Journal of Experimental Psychology: Human Perception and Performance*, *2*, 291–294. doi:10.1037//0096-1523.2.2.291

Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance, *Journal of Marketing Research*, *45*, 633-644.

doi:10.1509/jmkr.45.6.633

Miller, R. L., Brickman, P. & Bolen, D. (1975). Attribution versus persuasion as a means for modifying behavior. *Journal of Personality and Social Psychology*, *31*, 430–

441. doi:10.1037/h0076539

Mills, J. (1958). Changes in moral attitudes following temptation. *Journal of Personality*,

26, 517-531.

- Monin, B., & Jordan, A. H. (2008). The dynamic moral self: A social psychological perspective. In D. Narvaez & D. Lapsley (Eds.), *Personality, identity, and character: Explorations in moral psychology* (pp. 341-354). New York, NY: Cambridge University Press.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10, pp. 173–220). New York, NY: Academic Press.
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 21, pp. 261–302). New York, NY: Academic Press.
- Thibodeau, P. H., & Boroditsky, L. (2011). Metaphors we think with: The role of metaphor in reasoning. *PLoS ONE*, 6, e16782. doi:10.1371/journal.pone.0016782
- Walton, G. M., Banaji, M. R. (2004). Being what you say: The effect of essentialist linguistic labels on preferences. *Social Cognition*, 22, 193-213.
doi:10.1521/soco.22.2.193.35463
- Williams, E. F., Pizarro, D., & Ariely, D. (2009, February). *Visceral states influence moral decision making*. Poster presented at the meeting of the Society for Personality and Social Psychology, Tampa, FL.
- Zhong, C. B., Ku, G., Lount, R. B., & Murnighan, J. K. (2010). Compensatory Ethics. *Journal of Business Ethics*, 92, 323-339. doi:10.1007/s10551-009-0161-6

Acknowledgements

We thank Elizabeth Mullen, Gregory Walton and the Monin-Mullen Morality Lab for input, and Hattie Bluestone and Ashli Carter for assistance. The research was funded by Stanford University's Graduate School of Business.

Figure Captions

Figure 1. Percentage of participants in the noun (“cheater”) and verb (“cheating”) conditions claiming to have won the think-of-a-number game in Experiment 1.

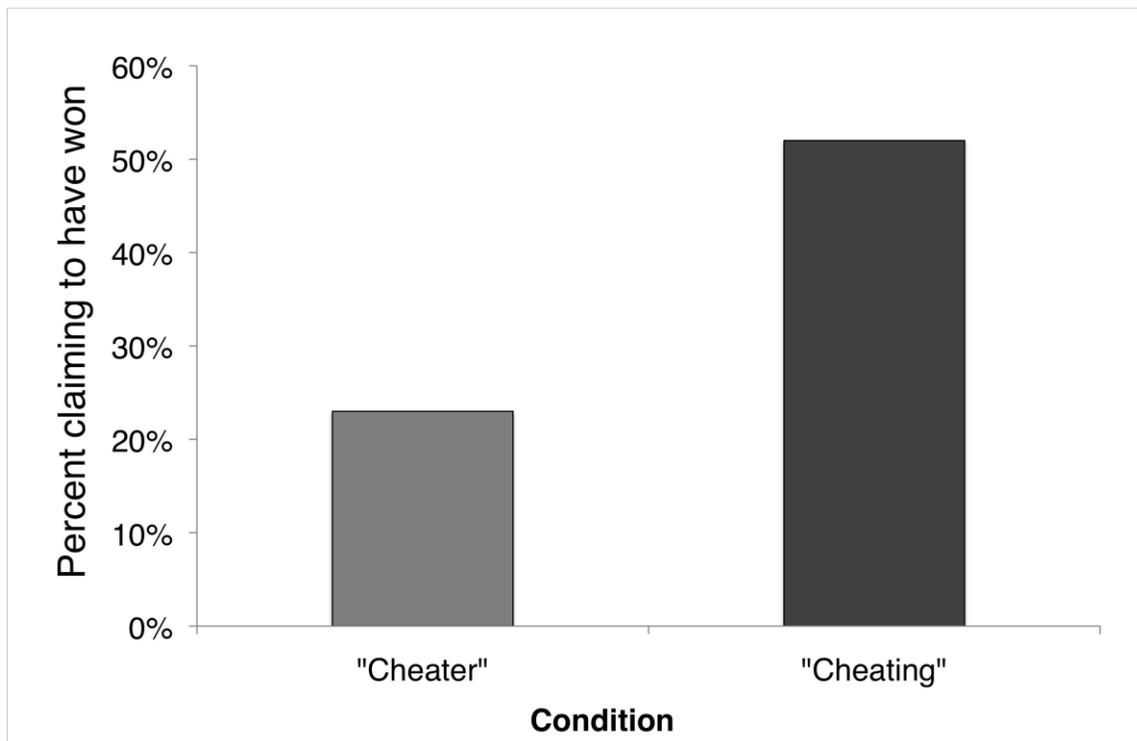
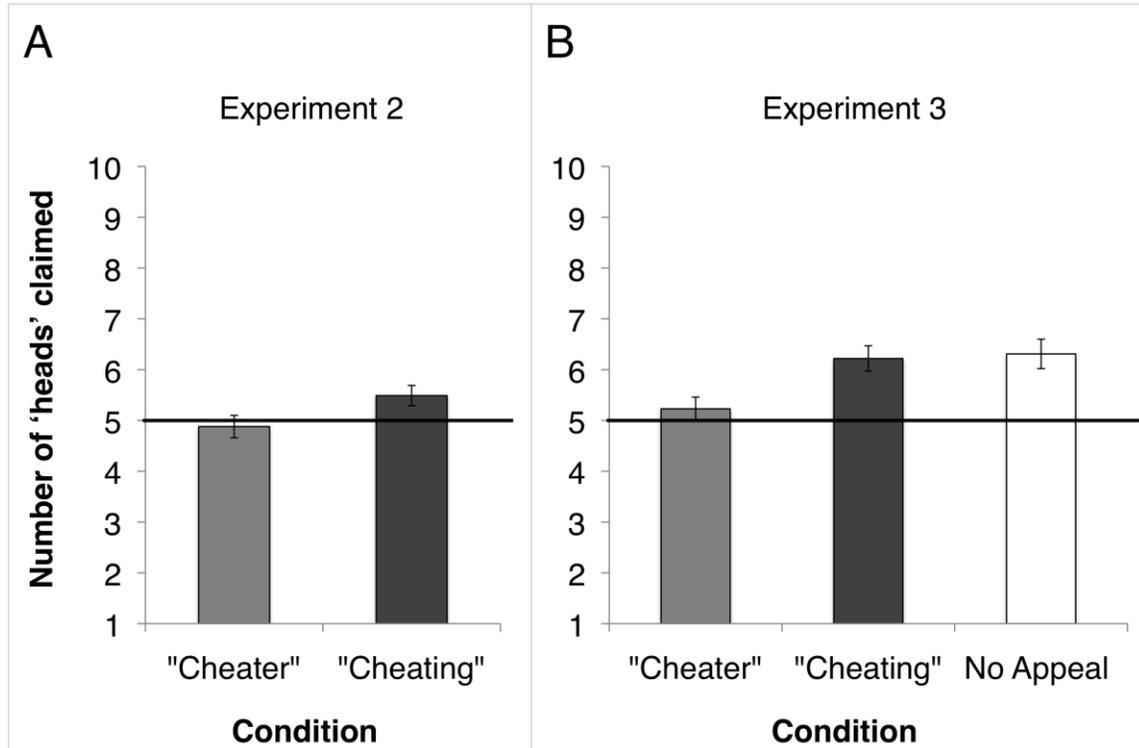


Figure 2. Number of “heads” participants claimed to have flipped in each experimental condition. (A) Experiment 2: Noun (“cheater”) and verb (“cheating”) conditions. (B). Experiment 3: Noun (“cheater”), verb (“cheating”) and no-appeal baseline conditions.



Online Supplemental Material

Pilot time trial

Because online samples often include participants who rush through studies without paying appropriate attention (Johnson, 2005), we established a minimum completion time (MCT) criterion for inclusion in the final sample for our two online experiments (2 and 3). One conventional procedure for eliminating outliers uses distributional criteria (e.g., 2 *SDs* above or below the mean). This method is not suitable here, however, because completion times are limited on the low end but not on the high end—there is a minimum time in which it is possible to complete an experiment in good faith but there is no practical maximum. Instead, we conducted a pilot test to determine the shortest time in which one could reasonably participate in good faith. We asked 5 colleagues who were unfamiliar with the experiment to complete the online experiment as quickly as possible while still reading all essential instructions and questions but skipping the consent form, having a coin immediately accessible for flipping, and ignoring two open-ended questions at the end of the study to ensure they completed the study as quickly as possible.

The mean completion time by our rushed testers was 3.09 minutes (*SD* = 0.55); this was set as our MCT. We emphasize that this is not the mean time in which a good faith participant would be *expected* to complete the experiment. Rather, it is the mean time in which a good faith participant could *possibly* be expected to complete the experiment. An even more conservative criterion (but, we believe, an unrealistic one given the already conservative instructions pilot participants were given) is the time in which our fastest rushed tester completed the experiment: 2.38 minutes. We report results of analyses using the 3.09-minute MCT in the main text. Analyses using the highly conservative 2.38-

minute criterion yield similar results and are reported in the Additional Analyses section below, as are results with no time-based exclusions. In Experiment 2, the 3.09-minute criterion excludes 5 people from the sample and the 2.38-minute criterion excludes 3. In Experiment 3, those two criteria exclude 32 and 19 people, respectively. As we note in the main text, the higher rate of time-based exclusions in Experiment 3 is likely the result of the sample we used in that experiment. Whereas participants in Experiment 2 had registered to take part in surveys regularly, and therefore had an existing relationship with the university and presumably had set time aside for the survey, the Experiment 3 sample consisted of casual internet browsers who landed on the survey somewhat unexpectedly, clicking an ad they discovered while browsing Facebook, and who therefore were presumably less likely to have time and motivation to take the study seriously.

Additional Analyses

Analyses using the 2.38-minute completion-time criterion

In Experiment 2, using the highly conservative 2.38-minute time criterion, participants in the verb condition claimed, on average, to have obtained 5.48 heads ($SD=1.24$) while those in the noun condition claimed to have obtained 4.90 heads ($SD=1.37$), $t(79)=1.97$, $p=0.053$, $d=0.44$. The number of heads reported in the verb condition was still significantly higher than the 5.00 expected by chance, $t(39)=2.42$, $p=0.020$, $d=0.39$, and the number of heads reported in the noun condition was still not different from chance, $t(40)=0.45$, $p>0.6$.

In Experiment 3, using the highly conservative 2.38-minute criterion, the omnibus effect of condition remained significant, $F(2, 109)=4.62$, $p=0.012$, as did the differences between the verb and noun conditions, $t(109)=2.56$, $p=0.012$, $d=0.67$, and between the

baseline and noun conditions, $t(109)=2.81, p=0.006, d=0.71$. The difference between the baseline and verb conditions remained non-significant, $t(109)=0.25, p>0.80$. Further, the number of heads claimed in both the verb and baseline conditions remained significantly higher than chance, $t(39)=4.81, p<0.0005, d=0.76$ and $t(41)=5.15, p<0.0005, d=0.79$, respectively. The number of claimed heads in the noun condition was still not significantly different from chance, $t(29)=1.47, p>0.15$.

Analyses with no time-based exclusions

In Experiment 2, including in the sample the 3 people who completed the experiment more quickly even than our highly conservative 2.38-minute criterion also did not change the results meaningfully. Participants in the verb condition claimed, on average, to have obtained 5.59 heads ($SD=1.41$) while those in the noun condition claimed to have obtained 4.86 head ($SD=1.36$), $t(82)=2.40, p=0.019$. The number of heads reported in the verb condition was still significantly higher than the 5.00 expected by chance, $t(40)=2.65, p=0.011$, and the number of heads reported in the noun condition was still not different from chance, $t(42)=0.68, p>0.5$.

In Experiment 3, including in the sample the 19 people who completed the experiment more quickly even than our highly conservative 2.38-minute criterion did mask our effect, however. Including them, the omnibus effect of condition is no longer significant, $F(2, 128)=0.94, p=0.394$, nor are the differences between the verb and noun conditions, $t(128)=1.25, p=0.215$, or between the baseline and noun conditions, $t(128)=1.16, p=0.248$. The difference between the baseline and verb conditions remained non-significant, $t(128)=0.11, p>0.91$. Finally, the number of heads claimed in the noun ($M=5.82$), verb ($M=6.27$) and baseline ($M=6.23$) conditions were all significantly higher

than chance, $t(38)=5.82$, $p<0.005$, $t(45)=6.27$, $p<0.0005$, and $t(48)=6.23$, $p<0.0005$, respectively, suggesting that the 19 people who completed the experiment so quickly cheated at a much higher rate than our good-faith participants.

Ancillary analyses of condition standard deviations

It might appear, on first glance, that the standard deviation in the noun condition in Experiment 3 ($SD=1.18$) is low compared to the expected value if participants were indeed reporting their coin tosses honestly ($E(s)=\sqrt{Np(1-p)}=\sqrt{10*.5*.5}=1.58$), raising the possibility that participants in that condition opted to report the most honest-sounding number of heads (i.e., 5) without even tossing a coin. However, a computer simulation with 5,000 samples of 27 participants, each flipping a fair coin 10 times yields a 95% confidence interval of {1.16; 2.00}, which includes the observed *SD*. Thus the observed variability is within the range of what can be reasonably expected, and we cannot reject the null to assume a restriction of the range. Participants seem indeed to be reporting their tosses honestly in the noun condition. Moreover, similar simulations show that the other two *SDs* in this study (as well as all *SDs* observed in Study 2) fall well within the expected 95% confidence interval corresponding to those cell sizes.

Reference

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality*, 39, 103-129.

doi:10.1016/j.jrp.2004.09.009