

Detecting Cherry-Picked Evidence in Texts: Challenges for Undergraduate Students

Hiroki Oura, Tokyo University of Science, houra@rs.tus.ac.jp
Toshio Mochizuki, Senshu University, tmochi@mochi-lab.net
Clark Chinn, Rutgers University, clark.chinn@gse.rutgers.edu
Eowyn Winchester, Rutgers University, epw34@scarletmail.rutgers.edu
Etsuji Yamaguchi, Kobe University, etuji@opal.kobe-u.ac.jp

Abstract: Authors of digital documents often seek to mislead readers by presenting *cherry-picked evidence*—e.g., a single study supporting a claim when most studies support a different claim. We report results of two experiments to examine whether undergraduate students adjust their epistemic judgments to account for cherry-picked evidence when they read multiple texts with conflicting claims. In Study 1, students adjusted their epistemic judgments when cherry picking was explicitly indicated with warnings called out in texts. In Study 2, however, with a different topic and four conditions that manipulated different degrees to which cherry picking of evidence was explicit, students did not adjust their epistemic judgments, even when evidence was blatantly cherry picked. In addition, very few students mentioned cherry picked evidence in explaining the grounds for their judgments, even when cherry picking was explicit but without such warnings in Study 1. These suggest that most students attend little to whether evidence is cherry picked, except the condition in which authors call out warnings of cherry-picking in texts.

Introduction

Citizens must often weigh information from multiple sources of information to make judgments on issues such as which medicine they should take for a health condition; much of the available information is unvetted and inaccurate. One way to attempt to determine which information to believe is to scrutinize the evidence presented for claims (Duncan et al., 2018). However, sources may intentionally or inadvertently mislead readers who seek to scrutinize evidence by presenting cherry-picked evidence; that is, they may selectively choose one or two studies that support a position, although the majority of studies support a different position. Accordingly, people need to develop their reasoning skills to judge whether evidence has been cherry picked and to reduce reliance on cherry-picked evidence. They should instead seek out information about what the larger bodies of evidence support. This is an important competence that can improve success in reasoning about conflicting evidence encountered online. However, researchers have not yet systematically investigated people's competence in recognizing cherry-picked evidence and in adjusting their beliefs in line with whether evidence has been cherry picked. In this paper, we report the results of two studies that investigate these competences.

Texts presenting cherry-picked evidence are pervasive in social media, and laypeople need to be able to recognize when it occurs and to adjust their conclusions accordingly (Levitan, 2017). However, sources who cherry pick evidence rarely make this clear: They may not realize it themselves, or they may conceal the fact that they have cherry picked the evidence that they present. Still, there may be signs that the evidence has been cherry picked. Authors may hint that they searched for evidence before finding any supporting evidence. Authors may betray cherry picking by not referring to what any larger body of evidence demonstrates, or by not giving any reason for choosing the evidence that they have highlighted. Further, authors of other sources of information may point out that an author has cherry picked evidence. They may also explicitly point out what positions are supported by larger bodies of evidence, highlighting possible cherry picking by those who tout unrepresentative studies. To detect such nuances in texts, readers need to be attentive to trying to find out what larger bodies of evidence say, not just what one or two studies say (Duncan et al., 2018). The purpose of the experiments reported in this paper is to explore whether undergraduates are attentive to these issues, and particularly if they are attentive to cherry-picked evidence when forming beliefs on scientific and social scientific issues.

Study 1. Comparison between Implicit and Explicit Signs of Cherry Picking

We conducted the first experiment to examine whether undergraduate students adjust two kinds of epistemic judgments in response to cherry-picked evidence: (1) beliefs about the claim that the cherry-picked evidence supports and (2) beliefs about the trustworthiness of the source that presents cherry-picked evidence. There were two conditions that manipulated how explicit the indicators of cherry-picked evidence were.

We developed two documents that presented competing claims on the relationships of rewards, motivation, and performance in social science. Document 1, said to be written by a business journalist, presented evidence from a study that supported a pro-reward claim that “rewards increase workers’ productivity” (pro-reward position) based on a lab study where 36 groups of business people in two conditions worked on personnel selection tasks with or without monetary rewards. The study supported the use of rewards to increase productivity. Document 2, said to be written by an author of five bestselling books on motivation research, presented an opposing claim that “rewards decrease workers’ productivity” (anti-reward position) based on a field study with 87 participants working on 9 varied tasks and a relatively large reward worth several months’ wages. The study’s conclusion was that the higher monetary rewards led to worse performance in most tasks, except for very simple tasks that did not require much cognitive attention. These documents were based on actual positions taken by researchers and authors on this topic (e.g., Document 2 was based on the work of Pink, 2011).

Method

The design was experimental, with two conditions (**implicit** versus **explicit** expressions of cherry picking) that varied according to the content of the two documents. In the *implicit* cherry-picking condition, Document 1 explained the author’s position, mentioning that “monetary rewards tell people that they are valued, and that their work is valued,” and then introduced the study described above as evidence, after writing, “Here is *a study* that supports this view.” The document then concluded, “It is well known rewards are a main motivator of human behavior.....This is *one study* that shows the efficacy of this approach.” (All italics in excerpts from documents have been added to this report for emphasis to highlight differences.) Document 2 first explained the author’s position, reporting that “*nearly all psychological studies* have shown that monetary rewards and prizes decrease motivation and performance.” Then the document introduced the evidence of the study described above after pointing out a problem in other studies that had been conducted because of the relatively small rewards in those studies. In the *explicit* cherry-picking condition, Document 1 added one statement to the document that asserted that the author had searched through many studies until finding one study that supported their claim, thus straightforwardly acknowledging cherry picking. Likewise, another statement was added to Document 2 in which the author warns that if readers find documents that supports a pro-reward position, those documents ignore the large majority of studies that support an anti-reward position. Thus, two corresponding documents presented the same evidence in both conditions, but only the explicit condition contained explicit descriptions of cherry picking and warnings about cherry picking. In both conditions, Document 1 said nothing about the broader body of evidence, and Document 2 asserted that most studies support the anti-reward position.

Participants were 151 undergraduate students in an introductory education course at a public university in Japan; 41 were excluded due to technical problems with the online system, leaving 110 participants. Participants were randomly assigned to conditions. All participants first answered a pre-survey in which they rated (a) the degree of their agreement with the pro-reward or anti-reward positions on 7-point Likert scales (1 = strongly disagree and 7 = strongly agree) (*pre-position*). Next they read versions of the two documents that corresponded to their assigned condition. After this, on a post survey using the same Likert scales, they rated (a) their degree of agreement with the pro-reward and anti-reward positions (*post-position*), (b) the trustworthiness of each of the two documents (*post-trustworthiness*), and (c) the strength of evidence that the pro-reward and anti-reward claims were based on (*post-strength*). In addition, they were asked to report their judgment of the relative amount of evidence supporting the two positions (ranging from 1, indicating that all the studies support the anti-reward position, to 7, indicating that all the studies support the pro-reward position) (*post-amount*). Each pair of responses (on the pro- and anti-reward positions for pre-and post-position, post-trustworthiness, and post-strength) was merged into one variable by subtracting the anti-reward rating from the pro-reward rating, so that a larger value indicated a preference for the pro-reward position, document, and evidence.

Results

The results are summarized in Table 1. There was no significant difference in pre-position between the conditions ($F(1,108) = 0.19, \eta^2 = 0.02$), with both groups tending to favor the pro-rewards position equivalently. On the post surveys, the means for each variable were significantly different (post-position: $F(1, 108) = 7.30, \eta^2 = 0.06$; post-trustworthiness: $F(1,108) = 6.33, \eta^2 = 0.06$; post-strength: $F(1,108) = 7.54, \eta^2 = 0.07$; and post-amount: $F(1,108) = 5.98, \eta^2 = 0.05$). Participants in the explicit condition provided, on average, significantly lower scores on post-position, post-trustworthiness, post-strength, and post-amount for the pro-reward position. That is, the *explicit* markers indicating that the pro-rewards author was cherry picking evidence caused participants to lower their relative preference for the pro-reward position. Explicit markers of cherry picking to support the pro-reward position also led the participants to lower their relative trust in the pro-reward source’s trustworthiness and to lower relative ratings of strength for the pro-reward source’s evidence.

Table 1
Descriptive and ANOVA Statistics for Study 1

Variables	Implicit (n = 62)		Explicit (n = 48)		F (1, 108)	η^2
	M	SD	M	SD		
Pre-position	3.15	2.22	2.96	2.29	0.19 <i>n.s.</i>	0.02
Post-position	1.42	2.46	0.17	2.35	7.30 **	0.06
Post-trustworthiness	-0.26	2.18	-1.27	1.98	6.33 *	0.06
Post-strength	0.68	2.09	-0.50	2.41	7.54 **	0.07
Post-amount	4.34	0.75	3.94	0.98	5.98 *	0.05

* $p < .05$, ** $p < .01$

Study 2. Comparison among Different Signs of Cherry Picking

In Study 1, undergraduate students adjusted their epistemic judgments appropriately when (a) the evidence was explicitly cherry picked and (b) the cherry picking was called out by an opponent. Given this finding, we conducted the second experiment to examine how students responded to differing degrees to which the cherry picking was explicit or implicit.

Method

Two fictional documents were developed that presented opposing claims on the effectiveness of a nutritional supplement called Serenade. Document 1 was said to be written by a doctor who had reported research on nutrition problems for 22 years. It argued for the positive effects of Serenade on health and presented as evidence a recent study showing that Serenade had a positive effect on health (pro-Serenade position). In contrast, Document 2, written by a different doctor who had researched on health promotion by health foods for 21 years, presented an opposing claim with different evidence from another recent study that supported the conclusion that Serenade has a harmful effect on health (anti-Serenade position).

There were four experimental conditions, determined by variations in the content in these two documents. In the first condition (*none*—*no indication of cherry picking*), the two documents were presented as described above. Specifically, both authors reported “a recent study” as evidence. In the second condition (*hint*), the pro-Serenade Document 1 stated, “After some searching of typical recent research reports, I found a study that I would like to share with you.” This statement provided a hint of cherry picking, whereas Document 2 just mentioned that the study presented as evidence was “one typical recent study.” In the third condition (*explicit*), Document 1 stated, “There have been many typical recent studies done on the health effects of Serenade. I have looked at many of these studies and in my search, I found one that supports its positive effects, which I would like to share with you.” This provided a blatant statement of cherry picking. Document 2 was the same as in the *hint* condition. In the last condition (*super explicit*), Document 1 included the same blatant statements as in the *explicit* condition, but also included as background information emphasizing that the author of Document 1 had searched many recent studies that supported his belief that Serenade has positive health benefits, and finally found one that supports his claim out of many studies that did not but that he did not report. Document 2 was the same as in the *hint* and *explicit* conditions, except that one statement was added as background information that the author of Document 2 had written the article after reviewing a variety of recent studies.

Participants were 63 undergraduate students in an introductory course of media and education at a private university in Japan. Participants were randomly assigned to the four conditions. Participants first answered a pre-reading survey in which they were asked to rate (a) their degree of agreement with the pro-supplement or anti-supplement positions for the effectiveness on health in general on 7-point Likert scale (1 = strongly disagree; 7 = strongly agree) (*pre-position*). Next students read the assigned two documents, which varied by condition as described above. After this, they completed a post-reading survey in which they rated their degree of agreement with the pro-Serenade and anti-Serenade positions (*post-position*) and the trustworthiness of the two sources (*post-trustworthiness*). In addition, they rated the relative amount of evidence for each position (1 indicated that all the studies supported the anti-Serenade position, and 7 indicated that all the studies supported the pro-Serenade position) (*post-amount*). Participants also provided written explanations for their level of agreement for both the pro-Serenade and anti-Serenade positions. In the analysis, each pair of responses on the pro- and anti-Serenade stances (*pre-position*, *post-position*, and *post-trustworthiness*) was merged into one variable by subtracting the anti-Serenade rating from the pro-Serenade rating, so that a large value indicates a relative preference for the pro-Serenade stances (see Table 2). In addition, student explanations were analyzed to determine whether students commented on cherry picking as part of their explanation of their level of agreement with the pro-Serenade anti-Serenade positions.

Results

The summary of results is shown in Table 2. The only significant difference between conditions was on the variable *amount of evidence* ($F(3, 59) = 4.47, \eta^2 = 0.19$), whereas differences on the other variables across condition were not statistically significant (Table 2). Participants' pre-position ratings were statistically equivalent across conditions. Post-reading positions and trustworthiness ratings were also not statistically different across conditions, except that students' judgements of the relative amount of evidence supporting Serenade (versus the amount of evidence opposing Serenade) differed across conditions. Post-hoc multiple comparisons with Bonferroni corrections indicated that students in the super-explicit condition rated the relative amount of evidence in favor of Serenade significantly lower than students in the none and hinted conditions. These indicate that participants in this condition recognized that there was more evidence in favor of the anti-Serenade position in this condition, yet they did not make a corresponding adjustment to their evaluations of the trustworthiness of evidence or of the overall benefits/harms of Serenade. However, the lack of statistically significant differences may result in part from insufficient statistical power.

Table 2
Descriptive and ANOVA Statistics for Study 2

Variable	None	Hinted	Explicit	Super Explicit	$F(3, 59)$	η^2
	($n = 15$)	($n = 15$)	($n = 16$)	($n = 17$)		
	$M(SD)$	$M(SD)$	$M(SD)$	$M(SD)$		
Pre-position	1.93 (2.09)	1.93 (1.53)	2.56 (1.63)	2.41 (1.80)	0.53 n.s.	0.03
Post-position	-0.33 (2.19)	-0.67 (1.84)	-0.69 (2.63)	-1.59 (1.77)	1.04 n.s.	0.05
Post-trustworthiness	-0.47 (1.36)	-0.60 (1.96)	-1.44 (2.00)	-0.18 (2.46)	1.19 n.s.	0.06
Post-amount	4.13 (1.06)	4.20 (0.56)	3.69 (0.87)	3.24 (0.83)	4.47 **	0.19

* < .05, ** < .01

In the analysis of students' written explanations for their ratings, only four students mentioned cherry picking of evidence as a problem—two students each from the explicit and super-explicit conditions; there was no statistical difference across conditions). (A fifth student expressed approval of the cherry-picking tactic!) Student explanations were also coded to identify student mentions of larger bodies of evidence favoring one side or the other. There were no such mentions. Combining these results with the results in Table 2, we conclude that students attended very little to the cherry picking of evidence even in the explicit conditions.

Discussion

In two experiments, we examined whether undergraduate students adjust their epistemic judgments in response to cherry-picked evidence when they read multiple texts with conflicting claims. In Study 1, students adjusted their epistemic judgements when cherry picking was explicitly indicated, whereas in Study 2, students did not adjust their judgments (except for ratings of the amount of evidence) when cherry picking was explicitly or super explicitly indicated. Across the two studies, there was one condition in which one source's cherry picking was explicitly criticized by an opponent (Condition 2 in Study 1); this was the only condition in which students adjusted their epistemic judgments in response to cherry picking. This factor may account for the differences across the two studies. Perhaps students are sensitive to the problem of cherry picking only when an opponent explicitly explains that a source as cherry picked evidence that is inconsistent with a much larger body of evidence. Alternatively, the non-significant results for Study 2 may arise from insufficient statistical power; the sample size for Study 2 was much less than that for Study 1. Further research is needed to investigate instructional methods that can help students become more vigilant in noticing cherry-picked evidence and adjusting beliefs accordingly.

References

- Duncan, R. G., Chinn, C. A., & Barzilai, S. (2018). Grasp of evidence: Problematizing and expanding the next generation science standards' conceptualization of evidence. *Journal of Research in Science Teaching*, 55 (7), 907-937.
- Leviton, D. (2017). *Not a scientist? How Politicians Mistake, Misrepresent, and Utterly Mangle Science*. Norton.
- Pink, D. H. (2011). *Drive: The surprising truth about what motivates us*. Riverhead Books.

Acknowledgments

This work was supported by JSPS Grants-in-Aids (#JP20H01729, #JP20K20829) and Uchida Yoko Co, LTD and by a 2019 Fulbright Award to Clark A. Chinn.