

Summer 2019

How Evidence of Subsequent Remedial Measures Matters

Bernard Chao

Kylie Santos

Follow this and additional works at: <https://scholarship.law.missouri.edu/mlr>



Part of the [Law Commons](#)

Recommended Citation

Bernard Chao and Kylie Santos, *How Evidence of Subsequent Remedial Measures Matters*, 84 Mo. L. REV. (2019)

Available at: <https://scholarship.law.missouri.edu/mlr/vol84/iss3/5>

This Article is brought to you for free and open access by the Law Journals at University of Missouri School of Law Scholarship Repository. It has been accepted for inclusion in Missouri Law Review by an authorized editor of University of Missouri School of Law Scholarship Repository. For more information, please contact bassettcw@missouri.edu.

MISSOURI LAW REVIEW

VOLUME 84

SUMMER 2019

NUMBER 3

How Evidence of Subsequent Remedial Measures Matters

*Bernard Chao & Kylie Santos**

ABSTRACT

Federal Rule of Evidence 407 prohibits plaintiffs from introducing evidence of subsequent remedial measures to show that the defendant is to blame. Among its purported justifications, the rule prevents hindsight bias from unduly influencing jury decisions. Nonetheless, plaintiffs often take advantage of the rule's numerous exceptions to introduce evidence of remedial measures for other purposes (e.g. to prove feasibility). Fearing that the exceptions could swallow the rule, some courts will even exclude evidence that fits into one of these exceptions because it is ostensibly too prejudicial. Alternatively, other courts instruct juries that they should only use the evidence for the limited permissible purpose, but not for proving blameworthiness.

This complex scheme makes several assumptions about how evidence of remedial measures and the accompanying limiting instructions influence juries. Although many studies have examined hindsight bias in other contexts, and one older study looked at Rule 407 in particular, these studies typically used short, written vignettes with small sample sizes. Moreover, none of these studies examined how subsequent measures impact damages. We sought to test these concepts in robust fashion by conducting two separate experiments using videos that included rich fact patterns including arguments from both parties and jury instructions. In the end, over one thousand seven hundred mock jurors rendered verdicts on liability, contributory negligence, and damages.

As expected, evidence of subsequent remedial measures helped plaintiffs win more often. But surprisingly, our results also suggested that taking remedial measures may lower damages, thereby counteracting the increased liability findings. We also studied the efficacy of two limiting jury instructions. In

* Professor Chao is a Professor at the University of Denver, Director of its Intellectual Property certificate program and co-Director for the Denver Empirical Justice Institute. Kylie Santos J.D. '19. The authors would like to thank Arthur Best, John Campbell, Viva Moffat, Christopher Robertson, and Maggie Wittlin for their helpful comments at different phases of this project. We are also indebted to Catherine Durso, Mayuri Mylarisetti, Rosy Sahu and Pooran Singh Negi for their help with the statistical analysis.

one experiment, a limiting instruction with an explanation reduced but did not eliminate the effects of evidence of subsequent remedial measures. The instruction with explanation was also consistently more effective than the simple limiting instructions, but these results were not statistically significant. We hypothesize about the potential reasons for our various results and discuss what they mean for policymakers, litigants, and future researchers that may wish to explore this subject in more depth.

TABLE OF CONTENTS

ABSTRACT	609
INTRODUCTION	613
I. BACKGROUND	615
A. <i>Federal Rule of Evidence 407</i>	615
B. <i>Prior Studies</i>	621
C. <i>Damages</i>	624
D. <i>Hypothesis</i>	625
II. EXPERIMENT 1: PREMISES LIABILITY	626
A. <i>Scenario 1: The Basic Case</i>	627
B. <i>Scenarios 2 and 3: Not Feasible & Remedial Measures</i>	627
C. <i>Scenarios 4 & 5: Limiting Jury Instructions</i>	628
D. <i>Respondents</i>	629
E. <i>Results</i>	630
1. <i>Liability (Staircase)</i>	630
Table 1	631
Staircase Experiment, Liability	631
2. <i>Damages (Staircase)</i>	632
Table 2.....	633
Staircase Experiment:	633
Damages and Contributory Negligence	633
3. <i>Case Expected Value (Staircase)</i>	634
4. <i>Summary (Staircase)</i>	634
III. EXPERIMENT 2: PRODUCTS LIABILITY	635
A. <i>Scenario 1: The Basic Case</i>	636
B. <i>Scenario 2: The Presentation of Subsequent Remedial Measures</i>	637
C. <i>Scenarios 3 and 4: Limiting Jury Instructions</i>	638
D. <i>Respondents</i>	639
E. <i>Results</i>	639
1. <i>Liability (Snowboard)</i>	639
Table 3.....	640
Snowboard Experiment, Liability.....	640
2. <i>Raw Damages (Snowboard)</i>	641
Table 4.....	642
Snowboard Experiment:	642
Damages and Contributory Negligence (“CN”)	642
3. <i>Contributory Negligence (Snowboard)</i>	643
4. <i>Net Damages (Snowboard)</i>	643
5. <i>Case Expected Value (Snowboard)</i>	644
6. <i>Findings</i>	644
IV. DISCUSSION	645
A. <i>Implications</i>	645
1. <i>Policymakers</i>	645
2. <i>Litigants</i>	646
3. <i>Researchers</i>	647
B. <i>Limitations</i>	648
CONCLUSION	650
APPENDIX A	651

Table A1	651
Table A2	651
APPENDIX B	654
Snowboard Analyses	654
Table B1	654
Table B2	655
Table B3	656
Table B4	657
Table B5	658
APPENDIX C	659

INTRODUCTION

Federal Rule of Evidence 407 (“Rule 407”) prohibits plaintiffs from introducing evidence of subsequent remedial measures to show that the defendant is to blame.¹ Among its purported justifications, the rule prevents hindsight bias from unduly influencing jury decisions.² Nonetheless, plaintiffs often take advantage of the rule’s numerous exceptions to introduce evidence of remedial measures for other purposes (*e.g.* to prove feasibility). Fearing that the exceptions could swallow the rule, some courts will even exclude evidence that fits into one of these exceptions because it is ostensibly too prejudicial.³ Alternatively, other courts instruct juries that they should only use the evidence for the limited permissible purpose, but not for proving blameworthiness.⁴

This complex scheme makes several assumptions about how evidence of remedial measures and the accompanying limiting instructions influence juries. Although many studies have examined hindsight bias in other contexts, and one older study looked at Rule 407 in particular, these studies typically used short, written vignettes with small sample sizes. Moreover, none of these studies examined how subsequent measures impact damages. We sought to test these concepts in robust fashion by conducting two separate experiments using videos that included rich fact patterns with arguments from both parties and jury instructions. In the end, over one thousand seven hundred mock jurors rendered verdicts on liability, contributory negligence, and damages.

As expected, evidence of subsequent remedial measures helped plaintiffs win more often.⁵ But surprisingly, our results also suggested that taking remedial measures may lower damages, thereby counteracting the increased liability findings.⁶ We also studied the efficacy of two limiting jury instructions. In one experiment, a limiting instruction with an explanation reduced but did not eliminate the effects of evidence of subsequent remedial measures. The instruction with explanation was also consistently more effective than the simple limiting instructions, but these results were not statistically significant. We hypothesize about the potential reasons for our various results and discuss what they mean for policymakers, litigants, and future researchers that may wish to explore this subject in more depth.

This Article proceeds in four parts. Part I explains Rule 407 and its underlying justifications. In brief, the rule prohibits the introduction of evidence of subsequent remedial measures to show culpability. The thinking behind the rule is that hindsight bias causes juries to give too much weight to evidence of subsequent remedial measures, a result that would be unfair to defendants.

1. FED. R. EVID. 407.

2. *Id.* at advisory committee notes to 1972 proposed rule.

3. Dan M. Kahan, *The Economics – Conventional, Behavioral, and Political – of “Subsequent Remedial Measures” Evidence*, 110 COLUM. L. REV. 1616, 1628 (2010).

4. WRIGHT ET AL., 23 FED. PRAC. & PROC. EVID. § 5291 (2d ed. 2019).

5. *See infra* Part III.2.E.

6. *Id.*

However, the rule is riddled with exceptions that allow the evidence to be used for other purposes. But even when an exception applies, some courts will not admit the evidence of subsequent remedial measures because it is overly “prejudicial.” Other courts issue limiting instructions to the jury telling them that they can use the evidence for one purpose but not another. This complex scheme makes several assumptions about the effect of evidence of subsequent remedial measures and limiting instructions. Part I then goes on to explain that few studies have examined how Rule 407 affects decision-making. To the extent that studies do exist, they have only considered liability decisions. No studies have looked at how evidence of remedial measures might affect decisions on damages.

Part II then explains our first experiment (the staircase experiment). In this negligence case, the plaintiff hurt herself falling down overly steep stairs. We created five versions of this case, cumulatively layering on a new manipulation for each subsequent version: basic case (version 1); defendant argued that it was not feasible to make the staircase safer (version 2); plaintiff introduced evidence that the defendant later added landing (version 3); judge provided a simple limiting instruction (version 4); and judge explained the reason for the limiting instruction (version 5). For each of these versions, we asked mock jurors to determine liability and if applicable, contributory negligence and damages.

Part II then describes the results of the staircase experiment. As expected, evidence of subsequent remedial measures increased plaintiff’s win rate. The limiting jury instructions with explanation reduced liability findings, while a simple instruction trended in the same direction but was not statistically significant. We did not observe any significant damages effects, but the limiting instruction with explanation appeared to be trending lower. We only mention the null damages findings because of the more significant results found in our second experiment.

Part III then explains our second experiment (the snowboard experiment). In this products liability case, the plaintiff was hurt during a snowboarding accident. He sued the snowboard manufacturer for a defective design. While we manipulated this case in two dimensions, four basic versions of this case are relevant to this article.⁷ Again, we cumulatively layered on a new manipulation for each subsequent version: the defendant argued that it made the design as safe as it could while retaining the ability for customers to choose different bindings (version 1); plaintiff introduced evidence that the defendant later made a safer design (version 2); judge provided a simple limiting instruction (version 3); and judge explained the reason for the limiting instruction (version 4). Again, mock jurors were asked to determine liability and if applicable, contributory negligence and damages.

Part III then describes the results of the snowboard experiment. Like the staircase experiment, the snowboard experiment confirmed that evidence of

7. See *infra* note 122 describing the other “low anchor” manipulation.

subsequent remedial measures increased plaintiffs' win rate. Although the impact was smaller than we saw in experiment 1, it was still sizable and statistically significant. Interestingly, when the mock jurors saw evidence of subsequent remedial measures, the plaintiff's win rate increased, but net damages (raw damages discounted for findings of contributory negligence) decreased. Although the causal mechanism is unclear, we hypothesize that jurors thought the defendant's decision to take remedial measures made the defendant less blameworthy and therefore deserving of smaller damages. Because liability and damages went in opposite directions, we saw no statistically significant differences between any of our scenarios when we examined the case expected value. This effect was only present in the snowboard case. This suggests that, at least in some cases, evidence of subsequent measures may not harm defendants as much as previously thought.

Additionally, while the limiting instruction with explanation appeared to lower liability findings, that result was not statistically significant in the snowboard experiment. Moreover, the simple limiting instruction appeared to have no effect whatsoever. Finally, in Part IV, we describe implications of these different findings and the limitations of our study.

I. BACKGROUND

A. *Federal Rule of Evidence 407*

Rule 407 of the Federal Rules of Evidence prohibits the introduction of subsequent remedial measures to prove blameworthiness in negligence and products liability cases, but the rule allows such evidence to be admitted for other reasons, including to prove the feasibility of such remedial measures.⁸ Most states have identical or substantially similar rules to the federal rule.⁹ Although the rule was originally limited to cases involving claims for negligence, many state courts began applying this rule to products liability cases, which is

8. FED. R. EVID. 407. ("When measures are taken that would have made an earlier injury or harm less likely to occur, evidence of the subsequent measures is not admissible to prove: negligence; culpable conduct; a defect in a product or its design; or a need for a warning or instruction. But the court may admit this evidence for another purpose, such as impeachment or – if disputed – proving ownership, control, or the feasibility of precautionary measures."). This paper uses the term "blameworthiness" to refer to all four issues where evidence of remedial measures is inadmissible.

9. Mark G. Boyko & Ryan G. Vacca, *Who Knew – The Admissibility of Subsequent Remedial Measures When Defendants Are Without Knowledge of the Injuries*, 38 MCGEORGE L. REV. 653, 657–58 (2007); DAVID P. LEONARD, *THE NEW WIGMORE: A TREATISE ON EVIDENCE: SELECTED RULES OF LIMITED ADMISSIBILITY* § 2.3.4 (2018) (discussing the acceptance of the rule throughout the country); *but see* R.I. R. EVID. 407 (specifically allowing evidence of subsequent remedial measures).

a strict liability tort.¹⁰ In 1997, Federal Rule of Evidence 407 followed suit, adding language that also prohibited the use of evidence of subsequent remedial measures to prove “a defect in a product or its design; or a need for a warning or instruction.”¹¹ Many, but not all, states have expanded the rule to products liability.¹²

There are two articulated justifications underlying Rule 407. First, the rule encourages accused defendants to take steps to improve safety without fear that these steps will be used against them in court.¹³ Second, and perhaps more controversially, some commentators have questioned the probative value of evidence of subsequent remedial measures.¹⁴ While there is widespread agreement that the evidence would be admissible to show blameworthiness under the liberal rules of relevance found in contemporary evidence rules, commentators disagree on just how probative the evidence is.¹⁵ Some suggest that such evidence has almost no value.¹⁶ In contrast, others believe that the evidence tends to suggest negligence or a product defect.¹⁷ Regardless of which side

10. *See* *Raymond v. Raymond Corp.*, 938 F.2d 1518, 1522 (1st Cir. 1991); *In re Joint E. Dist. & S. Dist. Asbestos Litig. v. Armstrong World Indus., Inc.*, 995 F.2d 343, 345 (2d Cir. 1993); *Cann v. Ford Motor Co.*, 658 F.2d 54, 60 (2d Cir. 1981); *Kelly v. Crown Equip. Co.*, 970 F.2d 1273, 1275 (3d Cir. 1992); *Werner v. Upjohn, Inc.*, 628 F.2d 848, 857 (4th Cir. 1980); *Grenada Steel Indus., Inc. v. Ala. Oxygen Co.*, 695 F.2d 883, 888 (5th Cir. 1983); *Bauman v. Volkswagenwerk Aktiengesellschaft*, 621 F.2d 230, 232 (6th Cir. 1980); *Flaminio v. Honda Motor Co.*, 733 F.2d 463, 469 (7th Cir. 1984); *Gauthier v. AMF, Inc.*, 788 F.2d 634, 636–37 (9th Cir. 1986).

11. FED. R. EVID. 407 advisory committee’s note to 1997 amendment.

12. *See Raymond*, 938 F.2d at 1522; *In re Joint E. Dist. & S. Dist. Asbestos Litig.*, 995 F.2d at 343; *Cann*, 658 F.2d at 60; *Kelly*, 970 F.2d at 1275; *Werner*, 628 F.2d at 857; *Grenada Steel*, 695 F.2d at 888; *Bauman*, 621 F.2d at 232; *Flaminio*, 733 F.2d at 469; *Gauthier*, 788 F.2d at 636–37; *see also* LEONARD, *supra* note 9, at § 2.6.6.

13. FED. R. EVID. 407 advisory committee’s note to 1972 amendment (discussing the “social policy of encouraging people to take, or at least not discouraging them from taking, steps in furtherance of added safety.”).

14. *Id.*; *see also* Kahan, *supra* note 3, at 1628.

15. Under Federal Rule of Evidence 401, evidence is relevant if it has any tendency to make the existence of any fact of consequence to the determination of the issue more or less probable than it would be without the evidence. Even the notes to Rule 407 acknowledge that it is “possible” to infer blameworthiness from remedial measures. FED. R. EVID. 407 advisory committee’s note to 1972 amendment.

16. *See, e.g.*, Jeffrey J. Rachlinski, *A Positive Psychological Theory of Judging in Hindsight*, 65 U. CHI. L. REV. 571, 617 (1998) (“subsequent remedial measures are almost irrelevant to a determination as to whether the defendant was negligent with respect to the initial accident.”); *see also* Leonard, *supra* at 9 at § 2.3.1 (discussing various cases disagreeing on the relevance of subsequent remedial measures).

17. Kahan, *supra* note 3, at 1633 (arguing that “the adoption of a preventive measure after an accident is more consistent with the conclusion that it would have been reasonable to adopt it before the accident than with the conclusion that it would not

they take, everyone worries that jurors will give too much weight to evidence of subsequent remedial measures and thereby ignore contrary exculpatory evidence.¹⁸

This latter concern is rooted in the tendency of individuals to use hindsight bias in their decision-making:

In hindsight, people consistently exaggerate what could have been anticipated in foresight. They not only tend to view what has happened as having been inevitable but also to view it as having appeared ‘relatively inevitable’ *before* it happened. People believe that others should have been able to anticipate events much better than was actually the case.¹⁹

When assessing negligence or product liability claims, the jury must decide if the defendant acted either negligently or created an unreasonable risk of harm that the defendant should have recognized *at the time* of his or her conduct.²⁰ The problem is that a jury cannot truly view the conduct “at the time” it occurred, but rather, must assess the conduct after they have knowledge that the conduct caused an injury. Hindsight bias is an inherent problem with tort litigation because jurors know that there was an unexpected injury and are looking for someone to blame.²¹ The introduction of subsequent remedial measures can exacerbate this bias.²² Jurors might mistakenly construe such evidence as an admission or proof of guilt when, in fact, making something *even* safer does not necessarily mean that the prior precautions were unsafe or unreasonable to begin.

By prohibiting evidence of remedial measures, Rule 407 appears to eliminate the possibility that evidence of those measures will lead to hindsight bias.²³ This outcome is consistent with views of those who believe evidence of subsequent remedial measures have no, or at least a minimal, tendency to

have been reasonable to adopt it at that time”); *see also* Maggie Wittlin, *Hindsight Evidence*, 116 COLUM. L. REV. 1323, 1333–44 (2016) (using both Bayesian reasoning and reasoning by inference to explain generally the relevance of hindsight evidence).

18. Kahan, *supra* note 3, at 1623 (discussing concern that a factfinder will overestimate the likelihood that a subsequent remedial “measure could have been gauged before the accident.”); Rachlinski, *supra* note 16, at 617–18 (approving the rule excluding evidence of subsequent remedial measures).

19. Baruch Fischhoff, *For Those Condemned to Study the Past: Heuristics and Biases in Hindsight*, in JUDGEMENT UNDER UNCERTAINTY 335, 341 (Daniel Kahneman, Paul Slovic, and Amos Tversky eds., 1982) (describing the way in which humans attempt to use the past to make predictions in the future).

20. RESTATEMENT (SECOND) OF TORTS § 282, cmt. g (1965).

21. *See generally* Kim A. Kamin & Jeffrey J. Rachlinski, *Ex Post ≠ Ex Ante: Determining Liability in Hindsight*, 19 LAW HUM. BEHAV. 89 (1995) (finding that when participants knew of an adverse outcome, they were more likely to find that the conduct taking place prior to the injury was unreasonable).

22. *Id.*

23. FED. R. EVID. 407.

show blameworthiness.²⁴ However, if evidence of subsequent remedial measures has some probative value, the concern is that jurors will give the evidence too much weight. Under this view, the rule may be too blunt an instrument.²⁵ As Dan Kahan has argued, excluding the evidence entirely may be an overcorrection.²⁶ In other words, while allowing the evidence may result in jurors overestimating liability because of hindsight bias, excluding that evidence may result in jurors underestimating liability by taking away relevant information from their consideration.

The rule's significant exceptions further complicate this picture.²⁷ Evidence of subsequent remedial measures is admissible for the purpose of impeachment or, if disputed, proving ownership, control, or the feasibility of precautionary measures.²⁸ Trial attorneys can manipulate witnesses to take advantages of these exceptions.²⁹ Two real world examples illustrate this phenomenon. In *Martinez v. WR Grace & Company*,³⁰ the plaintiff was injured when she failed to see a bump in a grocery store parking lot and tripped over it.³¹ She argued that the defendant was to blame because the bump was not painted a different color.³² The defendant's safety manager testified that the bump's natural coloring was sufficiently different from the rest of the paved parking lot to allow the plaintiff to see it.³³ However, the safety manager also testified that if the color had not been different they would have painted it.³⁴ After the accident occurred, the defendant painted the bump yellow.³⁵ As a result, the court found that evidence of subsequently painting the bump was

24. *Id.* at advisory committee's note to 1972 amendment.

25. Kahan, *supra* note 3, at 1634.

26. *Id.* (arguing that evidence of subsequent remedial measures in some cases "justif[ies] revising upward an estimate of substandard conduct on the part of the defendant.").

27. Leonard, *supra* note 9, at § 2.9 (stating that "Counsel bent on exposing the jury to subsequent repair evidence will often be able to fit the evidence into one of the broadly construed alternative purposes.") (citation omitted); Stephen D. Easton, *The Real World Rules of Evidence*, 7 PRAC. LITIG. 49, 56–57 (Jan. 1996) (suggesting that because of the rule's exceptions, "[s]ubsequent remedial measures are almost always admitted").

28. FED. R. EVID. 407.

29. *See, e.g.*, Kahan, *supra* note 3, at 1626 (describing how a plaintiff's attorney can manipulate a witness into either conceding that the defendant's conduct was unreasonable or opening the door to allow introduction of evidence of subsequent remedial measures).

30. 782 P.2d 827 (Colo. App. 1989).

31. *Id.* at 828.

32. *Id.*

33. *Id.* at 829.

34. *Id.*

35. *Id.*

admissible, not to prove negligence, but to impeach the safety manager's testimony that the bump did not need to be painted.³⁶ *Herndon v. Seven Bar Flying Service* also illustrates how easy it is to introduce evidence of subsequent remedial measures by impeachment.³⁷ In *Herndon*, the Tenth Circuit found that a defendant's refusal to opine or testify as to the feasibility of a safety precaution essentially controverted feasibility, and thus, opened the door to the admissibility of subsequent remedial measures.³⁸

Concerned that the exceptions can swallow the rule, courts use two tactics to prevent this from occurring. First, they often narrowly construe Rule 407's exceptions.³⁹ For example, in *Bauman v. Volkswagen Aktiengesellschaft*,⁴⁰ a defective design of a car door was alleged to have caused the plaintiff's injury.⁴¹ The plaintiff sought to introduce evidence that the defendant redesigned the doors after the accident.⁴² The defendant stated that the redesign occurred simply to comply with new government regulations.⁴³ Since Volkswagen never affirmatively controverted the feasibility of an improved design, the U.S. Court of Appeals for the Sixth Circuit found that the admission of subsequent remedial measures was inappropriate.⁴⁴ Under this view, it appears that evidence of subsequent remedial measures is only admissible to impeach claims that an improved design was not feasible. In other words, a defendant must actively controvert feasibility for the exception to apply.⁴⁵ But the text of Rule 407 suggests that "impeachment" and "if disputed" are two separate reasons why evidence of subsequent remedial measures can be introduced.⁴⁶ The *Volkswagen* analysis collapses these two reasons into a single exception.

Second, even when one of Rule 407's exceptions appears to apply, courts have relied on Rule 403 to exclude evidence of remedial measures, finding the evidence more prejudicial than probative.⁴⁷ Federal Rule of Evidence 403 states that the court "may exclude relevant evidence if its probative value is substantially outweighed" by the danger of such things as unfair prejudice or misleading the jury.⁴⁸ The Advisory Committee's notes suggest that Rule 403

36. *Id.* (explaining that evidence of subsequent remedial measures was permissible for impeachment purposes).

37. *Herndon v. Seven Bar Flying Service, Inc.*, 716 F.2d 1322 (10th Cir. 1983).

38. *Id.* at 1329 (explaining subsequent remedial measure implemented by defense to be redesigning a pitch trim switch, which modifies the planes altitude).

39. Kahan, *supra* note 3, at 1627.

40. 621 F.2d 230 (6th Cir. 1980).

41. *Id.* at 232.

42. *Id.* at 233.

43. *Id.*

44. *Id.*

45. *Knight v. Otis Elevator Co.*, 596 F.2d 84, 91 (3d Cir. 1979) ("Rule 407 requires that the feasibility of precautionary measures be controverted.").

46. FED. R. EVID. 407.

47. Kahan, *supra* note 3, at 1628.

48. FED. R. EVID. 403.

was “designed as a guide for the handling of situations for which no specific rules have been formulated.”⁴⁹ The Rules that follow, like Rule 407, “are concrete applications evolved for particular situations.”⁵⁰ These comments plainly suggest that there is no need to consider Rule 403 if a more specific Rule applies. Nevertheless, the U.S. Court of Appeals for the Third Circuit has explicitly said that courts should look at both Rule 403 and 407 together to determine whether evidence of remedial measures should be admitted.⁵¹

Finally, if the court admits evidence of remedial measures, the rules of evidence entitle a defendant to a limiting jury instruction.⁵² This instruction explains how the jury should use the evidence.⁵³ Namely, they can consider the evidence for any of the issues outlined in Rule 407’s exception, but never to show culpability.⁵⁴ The specific instruction given by the judge can vary depending on the state or even the judge. Notably, instructions generally do not explain the purpose of the rule.⁵⁵ Rather they simply tell the jury how to use the evidence. One exemplar states,

Certain evidence has been admitted in this case for only a limited purpose. This [evidence consists of ___ and] may be considered by you only for the purpose of ___. It may not be considered for any other purpose. Any discussion of the evidence during your deliberations must be consistent with this limitation.⁵⁶

In short, Rule 407 and its exceptions are premised on three beliefs: (1) defendants will be more willing to take remedial measures if the evidence cannot be used against them later on;⁵⁷ (2) evidence of subsequent remedial measures will lead to hindsight bias, thereby causing juries to overestimate blameworthiness;⁵⁸ and (3) limiting instructions will effectively counter any

49. *Id.* at advisory committee’s note to 1972 amendment (discussing the need for an exclusion where certain circumstances give rise to “an undue tendency to suggest decision on improper basis.”).

50. *Id.*

51. *Stecyk v. Bell Helicopter Textron, Inc.*, 295 F.3d 408, 415–16 (3d Cir. 2002); *see also Cyr v. J.I. Case Co.*, 652 A.2d 685, 694 (N.H. 1994) (suggesting that even if evidence of subsequent repairs would be admissible under Rule 407, Rule 403 should be applied without regard to the policies that support 407).

52. WRIGHT, *supra* note 4, at § 5291 (stating, “If the court decides that exclusion is not required under Rule 403, the defendant is entitled, upon request, to an instruction limiting the use of evidence of remedial measures to one of the permissible purposes.”).

53. Soo Yeon Ahn, *Friendly Limiting Instructions*, 52 U.S.F. L. REV. 353, 353 (2018).

54. *Id.* at 354.

55. *Id.* at 355.

56. 6 Wash. Prac., Wash. Pattern Jury Instr. Civ. WPI. 1.06 (6th ed.).

57. FED. R. EVID. 407 advisory committee notes for 1972 proposed rules.

58. *Id.*

hindsight bias.⁵⁹ For those who think evidence of subsequent remedial measures is relevant, the flip side of the second premise is that juries may underestimate blameworthiness if not exposed to the evidence of subsequent remedial measures. To the extent these beliefs relate to how juries do and do not make decisions, these beliefs can be tested experimentally. Of course, this experiment will not answer questions about whether Rule 407 encourages remedial measures. Nonetheless, understanding how jurors view evidence of subsequent remedial measures should aid policymakers in assessing the value of Rule 407 and its exceptions.

B. Prior Studies

Although several studies have looked at the issue of hindsight bias in legal decision-making,⁶⁰ we are only aware of one study that specifically examined how the various issues associated with Rule 407 affect jury decision-making.⁶¹ Stephen Landsman and Richard Rakos had both real judges and mock jurors read short, written vignettes that involved exposure to subsequent remedial measures.⁶² While the primary purpose of the study was to determine whether judges were able to resist biasing information better than jurors, the study also provided some insight into the effect that subsequent remedial measures had on jury decision-making. The study found that mock jurors were more likely to find the defendant liable when the jurors were exposed to evidence of subsequent remedial measures.⁶³ Interestingly, the authors grouped both jurors that had been told that the evidence was admissible with those that were told that the evidence was objectionable and thus should be “disregarded.”⁶⁴

59. Ahn, *supra* note 53, at 365.

60. See, e.g., Bernard Chao et al., *Why Courts Fail to Protect Privacy: Race, Age, Bias, and Technology*, 106 CALIF. L. REV. 263 (2018); Chris Guthrie et al., *Inside the Judicial Mind*, 86 CORNELL L. REV. 777, 818 (2001); Reid Hastie et al., *Juror Judgments in Civil Cases: Hindsight Effects on Judgments of Liability for Punitive Damages*, 23 L. & HUM. BEHAV. 597, 609–10 (1999); Kamin & Rachlinski, *supra* note 21, at 98.

61. See MICHAEL J. SAKS ET AL., *THE PSYCHOLOGICAL FOUNDATIONS OF EVIDENCE LAW* 78 (2016) (stating they were unaware of any “empirical research” testing the different theories on Rule 407 decision-making).

62. Stephan Landsman & Richard F. Rakos, *A Preliminary Inquiry into the Effect of Potentially Biasing Information on Judges and Jurors in Civil Litigation*, 12 BEHAV. SCI. & L. 113, 114 (1994).

63. *Id.* at 122, 124 (Table 3 showing that 91% (31/34) of the mock jurors found the defendant liable when not exposed to evidence of subsequent remedial measures compared to 68% (45/66) of mock jurors that were exposed to evidence of subsequent remedial measures ($p < .03$)). We performed our own analysis using confidence intervals and found that the 23% difference (CI: 5.7% to 36%) in verdict rate was indeed significant.

64. *Id.* at 122.

To examine the effectiveness of limiting instructions, we performed another line of analysis with Landsman & Rakos' data. Specifically, we compared the verdicts of the mock jurors who had been exposed to the evidence of subsequent remedial measures and not told to disregard the evidence against those who were told to disregard the evidence.⁶⁵ Although the 9% reduction in verdict rate suggests that the sustained objection could reduce the effect of the evidence of subsequent remedial measures, we calculated a 95% confidence interval showing that Landsman & Rakos' results were consistent with anywhere between a 15% increase in verdict rate to a 32% reduction in verdict rate. Because the confidence interval crossed zero, the results were not statistically significant. Thus, the study did not have enough power (i.e. number of subjects in the experiment) to prove that sustaining the defendant's motion made the small observed difference.

We also sought to determine whether the sustained objection might neutralize the effect of exposure to evidence of subsequent remedial measures. Comparing the jurors that were never exposed to evidence of subsequent remedial measures to those exposed but with a sustained defense objection, we found that they were statistically insignificant (confidence interval: -2% to 38%). Thus, even though there was an 18% difference in verdict rate, we cannot be sure that this difference was not due to sheer randomness. Again, this suggests the study was underpowered. Nonetheless, the range of the confidence intervals, mostly above zero, suggests that granting the defendants' motion fails to fully neutralize the effect that exposure to evidence of subsequent remedial measures had on the mock jurors. In short, Landsman & Rakos results suggest that a sustained objection helps ameliorate, but not eliminate, the effect of exposure to subsequent remedial measures. But, there was insufficient data to be confident in that conclusion.

As is common to many studies of that era, the study had several serious limitations. The stimulus was short: two or three written paragraphs depending on the experimental condition.⁶⁶ This potentially exaggerated the effect of the few arguments that were present. The sample consisted of volunteers from the Cuyahoga County jury pool in Ohio.⁶⁷ Thus, it is unclear whether the findings apply to other populations. Additionally, as discussed above, the sample size

65. 73% (24/33) of mock jurors who were told that the evidence was inadmissible (i.e. ignore the evidence) found the defendant liable while 64% (21/33) of jurors that were told that evidence was admissible found the defendant liable.

66. Landsman & Rakos, *supra* note 62, at 120.

67. *Id.* at 120, 122 (the paper initially states that there are 104 jurors, but it only reports 100 responses).

was small (n=100) making it difficult to draw conclusions confidently.⁶⁸ Finally, the study only looked at liability, not damages.⁶⁹ While logically, evidence of subsequent remedial measures should only affect liability, other studies have repeatedly found that the two issues have unexpected interactions.⁷⁰

Both judges and commentators have been extremely skeptical of the effectiveness of limiting jury instructions.⁷¹ But David Sklansky has challenged this view and argued that such instructions work, although imperfectly.⁷² With this belief in mind, he advocates for engaging in the “messy but important task of assessing when evidentiary instructions are most likely to fail, how they can be made more effective”⁷³ Although not the main focus, this Article takes up Sklansky’s call and seeks to experimentally evaluate the efficacy of different jury instructions associated with Rule 407.

Others have already begun this work. In one study, Shari Diamond and Jonathan Casper found that jury instructions were more effective when accompanied by an explanation of their purpose.⁷⁴ The Casper & Diamond study involved a simulated antitrust price-fixing case.⁷⁵ In antitrust cases, the jury awards compensatory damages and the judge trebles (3x) those damages to arrive at the final award.⁷⁶ If jurors know that trebling will occur, they may be tempted to reduce their damages award.⁷⁷ The experiment was designed to

68. *Id.*

69. *Id.*

70. *See infra* Part II and Part III.

71. *See* *Krulwich v. United States*, 336 U.S. 440, 453 (1949) (Jackson, J., concurring) (“The naive assumption that prejudicial effects can be overcome, by instructions to the jury, all practicing lawyers know to be unmitigated fiction.” (citation omitted)); SAKS, *supra* note 61, at 102 (“The overwhelming conclusion from empirical research on the ability (or willingness) of jurors to follow instruction to use evidence for a limited purpose is like that for disregarding evidence – instructions fail to accomplish their purpose and can even backfire, making the forbidden use more influential than had no instruction been given.”); *see also* Roselle L. Wissler et al., *On the Inefficacy of Limiting Instructions: When Jurors Use Prior Conviction Evidence to Decide on Guilt*, 9 L. & HUM. BEHAV. 37, 38 (1985); Joel D. Lieberman et al., *Understanding the Limits of Limiting Instructions: Social Psychological Explanations for the Failures of Instructions to Disregard Pretrial Publicity and Other Inadmissible Evidence*, 6 PSYCHOL. PUB. POL’Y & L. 677, 677–78 (2000).

72. David Alan Sklansky, *Evidentiary Instructions and the Jury as Other*, 65 STAN. L. REV. 407, 408–09 (2013) (arguing that evidentiary instructions work imperfectly).

73. *Id.* at 409.

74. Shari Seidman Diamond & Jonathan D. Casper, *Blindfolding the Jury to Verdict Consequences: Damages, Experts, and the Civil Jury*, 26 L. & SOC’Y REV. 513, 515 (1992) (noting most courts, and by virtue those court’s rules, regard jurors as passive participants to the trials they decide).

75. *Id.* at 517.

76. *Id.*

77. *Id.*

assess, among other items, whether different instructions to ignore trebling were effective.⁷⁸ In one version, jurors were admonished to disregard the trebling provision, but they were not given any explanation for the instruction.⁷⁹ In another version, an explanation accompanied the jury instruction.⁸⁰ The latter scenario resulted in significantly higher damages.⁸¹ This suggests that explanations can make limiting instructions more effective. We attempted to replicate Diamond and Casper's findings in a different context – the limiting instructions typically associated with Rule 407.

C. Damages

In addition to examining liability decisions, our two experiments sought to assess whether evidence of subsequent remedial measures affected the amount jurors awarded. Landsman and Rakos only studied liability effects, and at first blush, this choice seems logical.⁸² After all, evidence of what the defendant did after an accident should not affect how juries view the gravity of the plaintiff's injury or what damages should be awarded to compensate for that injury.

But several studies have shown that liability evidence affects damages decisions, and conversely, damages evidence often affects liability decisions. Roselle Wissler *et al.* referred to this as “fusion.”⁸³ Initially, we had no hypothesis for how evidence of subsequent remedial measures might affect damages. We only asked mock jurors to assess damages because one of the authors previously observed unexpected fusion effects in prior works.⁸⁴ But as we analyzed the results and observed that damages appeared to decrease when evidence of subsequent remedial measures were introduced, we looked for an explanation and developed the following hypothesis.

Jurors may award lower damages when the defendant takes subsequent remedial measure to address the cause of plaintiff's injury because the defendant is less morally blameworthy. Several studies have shown that higher damages are awarded when the evidence of defendant's liability actions are

78. *Id.* at 521.

79. *Id.* at 523.

80. *Id.*

81. *Id.* at 534 (detailing $t=-2.63$, $p<.01$).

82. Landsman & Rakos, *supra* note 62, at 120.

83. Roselle L. Wissler *et al.*, *The Impact of Jury Instructions on the Fusion of Liability and Compensatory Damages*, 25 LAW & HUM. BEHAV. 125, 125–39 (2001).

84. See John Campbell *et al.*, *Time is Money: An Empirical Assessment of Non-Economic Damages Arguments*, 95 WASH U. L. REV. 1, 26 (2017) (finding that characterizing plaintiff's injury in granular time units increased plaintiff's likelihood of prevailing on liability); John Campbell *et al.*, *Countering the Plaintiff's Anchor: Jury Simulations to Evaluate Damages Arguments*, 101 IOWA L. REV. 543, 562 (2016) (observing that plaintiff's prevailed on liability slightly less often when they asked for an outrageously high damages award).

stronger.⁸⁵ Mock juries appeared to punish blameworthy defendants with higher damages awards.⁸⁶ The Hans-Reyna gist based model of decision-making explains a potential mechanism for this phenomenon.⁸⁷ The model suggests that once jurors determine damages are warranted, they will make ordinal gist judgment about the amount of damages appropriate (e.g. low or high).⁸⁸ As part of this process, jurors will consider the defendant's blameworthiness (e.g. degree of negligence).⁸⁹

Evidence of remedial measures might work this way too, but in reverse. In some sense, the defendant is "doing the right thing" by taking steps to prevent others from being injured. Juries might find the defendant to be less blameworthy and reduce damages accordingly. From a legal perspective, this result would be illogical. The gravity of the plaintiff's injury is unaffected by any subsequent remedial measures. What's more, these steps are taken after the conduct that led to the original injury. But from a psychological perspective, it is natural for jurors to use this information to assess the blameworthiness of the defendant and award damages based on that determination.

D. Hypothesis

We seek to test several assumptions associated with Rule 407. Our four primary hypotheses are:

(1) *Underlying Justification for Rule 407*: The win rate for the plaintiff will increase when evidence of subsequent remedial measures is introduced.

(2) *Simple Limiting Jury Instruction*: A simple limiting jury instruction will not affect the jury's decision on liability.

85. See, e.g., John M. Darley & Charles W. Huff, *Heightened Damage Assessment as a Result of the Intentionality of the Damaging-Causing Act*, BRITISH J. OF PSYCH. 29, 181–88 (1990) (showing higher damages when defendant's actions were more intentional); Gretchen B. Chapman & Brian H. Bornstein, *The More You Ask For, the More You Get: Anchoring in Personal Injury Verdicts*, APP. COGNITIVE PSYCHOL. 519, 526 (1996) (finding clearer evidence of the danger of birth control pills caused jurors to give higher damages); Dale W. Broeder, *The University of Chicago Jury Project*, 38 NEB. L. REV. 744, 760 (1959) (finding clearer evidence of liability led to higher compensatory damages).

86. See Darley & Huff, *supra* note 85, at 181–88; Chapman & Bornstein, *supra* note 85, at 526; Broeder, *supra* note 85, at 760.

87. Valerie P. Hans & Valerie F. Reyna, *To Dollars from Sense: Qualitative to Quantitative Translation in Jury Damage Awards*, 8 J. EMPIRICAL LEGAL STUD. 120, 126 (2011).

88. *Id.* at 129–30.

89. Valerie F. Reyna et al., *The Gist of Juries: Testing a Model of Damage Award Decision Making*, 21 PSYCHOL. PUB. POL'Y & L. 280, 291 (2015) (describing an experiment where jury perceptions of negligence predicted damages awards).

(3) *Limiting Jury Instruction with Explanation*: Limiting jury instructions that explain the basis for the instruction will reduce the effect of introducing evidence of subsequent remedial measures more than simple limiting jury instructions with no explanation.

(4) *Fusion of Liability with Damages*: When evidence of subsequent remedial measures is introduced, jurors will award lower damages.

Others have suggested that the best practice for researchers is to use multiple scenarios to test the same phenomena.⁹⁰ We follow that practice by testing our hypotheses using two experiments, a premises liability case where the theory of liability is based on a claim for negligence (“the staircase lawsuit”),⁹¹ and a products liability case where the plaintiff relies on strict liability theory (“the snowboard lawsuit”).⁹²

II. EXPERIMENT 1: PREMISES LIABILITY

In our first experiment, we performed a 5x1 between-subjects experiment⁹³ using Amazon Mechanical Turk (“MTurk”), an online crowd-sourcing marketplace.⁹⁴ We loosely modeled our first experiment after the events in *Duncan v. Mill Management Company*, a premises liability case.⁹⁵ In our experiment the plaintiff, Mackenzie Dunn, was a resident of Skyline Vista Apartments. Ms. Dunn was injured falling down a set of steep stairs leading to the building’s basement. Arguing that the stairs were too steep, she sued the building’s manager, Mesa Management, for negligence.

In order to test the different assumptions associated with Rule 407, we created five variations of this basic case. Each variation consisted of three PowerPoint presentations, one each for the presiding judge (with both an opening explanation of the dispute and jury instructions that followed the attorneys’ arguments), the plaintiff’s attorney, and the defendant’s attorney. Using three different voices, we then recorded arguments for each slide. The result was

90. See Krin Irvine, David A. Hoffman, & Tess Wilkinson-Ryan, *Law and Psychology Grows Up, Goes Online, and Replicates*, 15 J. EMPIRICAL LEGAL STUD. 320, 346 (2018) (suggesting that “researchers ought to show that the same results appear in response to different stimuli.”).

91. See *infra* Part II.

92. See *infra* Part III.

93. A between-subjects experiment means that each participant is only assigned to one condition.

94. MTurk has become a large and robust platform for social science research, with proven reliability through the replication of many known results. See Irvine et al., *supra* note 90, at 322, 344 (concluding that Amazon MTurk provides a reasonable subject pool for experiments in law and psychology); Adam J. Berinsky et al., *Evaluating Online Labor Markets for Experimental Research: Amazon.com’s Mechanical Turk*, 20 POL. ANALYSIS 351, 362–63 (2012) (stating the authors successfully replicated three experiments using MTurk).

95. 60 A.3d 222, 230 (Conn. 2013).

five videos lasting approximately fifteen to seventeen minutes. Like real jurors, the mock jurors were asked to render verdicts on liability, and, if applicable, contributory negligence and damages.⁹⁶ Subsequently, the mock jurors were asked several debriefing questions.⁹⁷

A. Scenario 1: The Basic Case

In this case, the plaintiff, MacKenzie Dunn, fell down a flight of stairs and brought a lawsuit for negligence against her apartment building, Mesa Management. Ms. Dunn's attorney pointed out that the stairs were so steep that they violated some of the surrounding cities' building codes. In addition, a previous tenant suffered a similar accident on the same set of stairs. Yet, the defendant did nothing. The plaintiff's injuries included a fractured left ankle that required surgery. She also complained about pain, suffering, and loss of enjoyment of her life. Altogether, the plaintiff requested damages in the amount of \$280,000: \$80,000 for her medical expenses and \$200,000 for her pain and suffering.

In response, the defendant, Mesa Management, explained that the stairs complied with the building code that actually governed the location of the building. Mesa Management's expert also pointed out that the stairs would have complied with the other jurisdictions' building codes the plaintiff referenced because those codes did not apply to older stairs like those at issue. The defendant also argued that Ms. Dunn's own negligence was to blame for her fall. With respect to damages, the defendant did not challenge the \$80,000 in medical expenses, but it did argue that \$200,000 for pain and suffering was far too much given that the plaintiff had healed quickly and could do almost all of the activities she had done before.

B. Scenarios 2 and 3: Not Feasible & Remedial Measures

Scenarios 2 and 3 were designed to test a basic premise of Rule 407: how does evidence of remedial measure affect jurors' decisions of liability? All of the evidence and arguments found in scenario 1 are also found in scenarios 2 and 3. However, in scenarios 2 and 3, Mesa Management relied on one more argument. Specifically, the defendant's attorney argued:

[b]ecause of the limited space surrounding the stairs in the basement, it would not have made any sense to add a landing in the middle of the

96. Our mock judge instructed the jury "Mesa Management claims that Mackenzie Dunn's own negligence contributed to her injuries. To succeed on this defense, Mesa Management must prove both of the following: (1) that Mackenzie Dunn was negligent; and (2) that Mackenzie Dunn's negligence was a substantial factor in causing her injuries." See Cal. Jury Instruction 405.

97. To prevent these questions from affecting responses on the primary outcomes of interest, participants could not return to the jury verdict after seeing the debriefing questions.

staircase or change the angle. If you look at the picture of stairs just after Ms. Dunn's fall, you can see it just was not feasible to retrofit the stairs as Ms. Dunn's expert would have you believe. There was simply not enough room. The stairs were as safe as they could have possibly been.

This argument opened the door for Rule 407's impeachment exception. We presented this evidence without impeachment in scenario 2. However, in scenario 3, the plaintiff had a rebuttal after the defendant's argument. In the rebuttal, the plaintiff used evidence of subsequent remedial measures to impeach the feasibility argument. Specifically, the plaintiff showed that the defendant later took steps to make the stairs less steep. These arguments and the corresponding diagrams were introduced in scenario 3, but they were also found in scenarios 4 and 5.

In short, scenario 2 allowed us to make a baseline measurement and observe how much weight mock jurors gave to defendant's argument that remedial measures were not feasible. Presumably, that argument should help the defendant persuade jurors that it was not liable. Scenario 3 then determined whether evidence of subsequent remedial measures would counter the argument of feasibility or perhaps overwhelm the feasibility by swinging the case dramatically in the plaintiff's favor.

Importantly, the evidence of subsequent remedial measures in this experiment was introduced in a separate rebuttal that served to disprove the defendant's argument that there was no way to make the stairs safer. Thus, any differences we observed between scenario 2 and 3 may have been due to a combination of the tendency for evidence of subsequent remedial measures to show the defendant should have taken remedial actions earlier and the fact that this evidence makes the defendant appear less credible.

C. Scenarios 4 & 5: Limiting Jury Instructions

Scenarios 4 and 5 tested the efficacy of different limiting instructions. Scenarios 4 and 5 both contained all the arguments and evidence found in Scenario 3. However, they also contained different limiting instructions that a judge may give to a jury. In scenario 4, the judge stated:

The jury needs to understand what this evidence may and may not be used for. The evidence regarding the remedial measures taken by Mesa Management after the earlier accident is not to be used to find negligence. Rather, it is only allowed to challenge defense counsel's argument that a safer design was not feasible.

This instruction was given to jurors directly after the evidence of the subsequent redesign of the stairs was presented. In other words, like a real trial, the judge instructed the jury in the middle of plaintiff's case. A similar instruction was also given at the end of the trial with the final jury instructions. At that time, the judge said:

During the trial, I told you that evidence of remedial measures was being admitted for the limited purpose of challenging the defendant's argument that a safer design was not feasible. You must consider it for that purpose only and may not consider it as evidence of negligence.

These instructions were similar to the instructions given in *Duncan v. Mill* and were very much like the standard limiting jury instruction given in many jurisdictions.⁹⁸

Scenario 5 tested a version of the slightly more elaborate jury instruction recommended by Diamond & Casper. In scenario 5, the standard instructions were given with one addition. Here, the judge gave a brief explanation for Rule 407. He stated:

We have this rule because we want to encourage people to fix problems and the law will not use evidence of such fixes against them later in court.

This explanation came twice, once directly after the presentation of the evidence and then again during the general jury instruction. Notably, the instruction did not mention Rule 407's second justification – namely, the risk that the jury will give too much weight to the evidence of subsequent remedial measures. The second experiment included this rationale. Thus, scenario 4 and 5 allowed us to test the effect of two possible limiting jury instructions.

D. Respondents

We recruited subjects from the population of workers on Amazon Mechanical Turk in September 2017 and established an initial screen for those that were “jury eligible” (i.e. residents of the United States over age 18 who could read, write, and speak English). Subjects were paid \$2.50 to complete the experiment online. All subjects gave consent in accordance with the Institutional Review Board requirements. At the beginning of the survey, subjects filled out a demographic questionnaire.

1,189 participants started the experiment participant's data. After disqualifying participants for failing attention checks or providing inconsistent responses, 981 participants remained.⁹⁹ This sample was nearly a 50/50 split

98. *Duncan*, 60 A.35 at 237.

99. Of the 207 participants who were not included in our analyses, 108 were eliminated due to participants failing to answer the attention check question correctly. Another seventeen participants were discarded for answering that the defendant was not negligent then answering that the defendant was a substantial part in causing the accident. Two more participants were discarded because they answered that the plaintiff was not negligent but that she was a substantial part in causing the accident. These answers indicated to us a confusion about what was needed to prove liability and were discarded to ensure our data was not contaminated.

between males and females with males representing 49% and females representing 51%. The sample was younger, more educated, and more politically liberal than the population at large; gender, race, and median income, however, were more representative of the U.S. Census data.¹⁰⁰

E. Results

The results of each simulation are presented below in Table 1. The descriptive statistics found in Table 1 provide a general sense of our results but do not indicate how reliable or significant our results are. We also calculated 95% confidence intervals and performed basic t-tests to answer those questions and have included that analysis in the text. In addition, we performed a series of regression analyses and included some of the most important ones in the appendix.

Table 1 describes the win rate for plaintiffs as well as the mean, median, and standard deviation for damages when the jurors awarded the plaintiff damages and the mean, median, and standard deviation for the entire experiment, including when jurors decided in favor of the defendant.

1. Liability (Staircase)

Looking first at liability, we found that evidence of subsequent remedial measures increased liability findings. Limiting jury instructions appeared to reduce this effect, but they did not eliminate it. The descriptive statistics are found in Table 1 below. In scenario 1, the control scenario, mock jurors awarded the plaintiff verdicts in 47.3% of the cases. In scenario 2, the defendant argued that it was not feasible to redesign the stairs. The win rate for the plaintiff increased slightly to 48.7% over scenario 1, but this was not statistically significant. Surprisingly, these results suggest that the defendant's argument that redesigning the staircase was not feasible did not help the defendant.

Evidence of remedial measures was introduced in scenario 3. The win rate increased by 20.1% over scenario 2 to 68.8%. In scenarios 4 and 5, different limiting jury instructions were introduced. In scenario 4, the judge used the basic limiting instruction and the plaintiff's win rate dropped to 60.3%. In scenario 5, the judge used the longer limiting instruction that provided an explanation for Rule 407. The results were very similar to scenario 4, but the effects were slightly larger. The plaintiff's win rate dropped to 57.6%.

100. Specifically, the sample demographics are as follows: 51% female, mean and median age 38 and 36, 78% white, 10% African American, 8% Asian, 2% American Indian or Alaskan Native, and the remainder classified as other; the median income was between \$30,000 to \$49,999; and medical education was a Bachelor's Degree, and 53% leaned toward, or strongly preferred Democrats while 38% leaned toward or strongly preferred Republicans.

Table 1
Staircase Experiment, Liability

Scenario	No. of Mock Jurors	Plaintiff Wins	Win Rate
1. Control	201	95	47.3%
2. Not Feasible	197	96	48.7%
3. Evidence of SRM	186	128	68.8%
4. Simple J-Instr.	204	123	60.3%
5. J-Instr. w/ explanation	191	110	57.6%

Although the descriptive statistics provide an intuitive sense of our liability findings, we also performed rigorous logistic regression analyses controlling for relevant demographics.¹⁰¹ The size of the effects changed modestly. But more importantly, we were able to determine which effects were statistically significant. We used scenario 3 as the baseline case because it allowed us to examine the effect of both subsequent remedial measure (2 vs. 3) and different jury instructions (3 vs. 4 and 3 vs. 5). In our model, the plaintiff's win rate for that was 68.7% ($p=0.000019$).¹⁰² The difference between scenario 2 and 3 was significant. The odds of plaintiff winning decreased by .417 when evidence of subsequent remedial measures was not present ($p=.000048$). When it was present, the introduction of a simple limiting instruction in scenario 4 appeared to reduce the plaintiff's win rate (odds ratio .700 to 1), but these results were not statistically significant ($p=.098$). When we introduced a limiting instruction that included an explanation of Rule 407, the size of the effect was larger than that for the simple jury instruction and statistically significant (odds

101. See App. A. We added demographics to our model whenever they had a statistically significant effect in either the staircase or snowboard experiments with respect to any of the dependent variables we report on. This allowed us to contrast demographic findings. The only demographic that affected liability in the staircase experiment was income. For every increase in our income scale, we found that the odds of plaintiff winning on liability decreased (odds ratio .812 to 1, $p=.002$).

102. See App. A (showing an odds ratio of 2.1987 translates to 68.7 % chance of the event occurring).

ratio .61 to 1, $p=0.0022$).¹⁰³ However, we should note that the differences between the two types of limiting instructions (scenarios 4 and 5) were not statistically significant. Finally, we found statistically significant income effects. As mock jurors' incomes increased on a six-point scale, the plaintiff won less often (odds ratio .81 to 1, $p=002$).¹⁰⁴

2. Damages (Staircase)

Because the ultimate damages a plaintiff actually receives can be reduced if the plaintiff is found to be contributorily negligent, our experiment also measured the various factors that make up the damages a plaintiff actually recovers. Those factors are the raw damages award, the rate at which the plaintiff was found contributory negligent, and the percentage of fault that was allocated to the plaintiff in those cases.

We first looked at raw damage awards without any discount for contributory negligence in Table 2 below. To do so, we excluded all cases in which the defendant won. In scenario 1, the average damages award was \$177,947. As manipulations were layered on in scenarios 2-4 (defendant arguing that remedial measures was not feasible, adding evidence of subsequent remedial measures, adding a short limiting jury instruction), the average damages remained essentially unchanged. However, when the jury was given a more detailed limiting jury instruction explaining the basis for Rule 407, the average damages award dropped to \$164,045.

103. See App. A.

104. Respondents were asked whether their household income was: 1) less than \$10,000, 2) \$10,000 to \$29,999, 3) \$30,000 to \$49,999, 4) \$50,000 to \$99,999, 5) \$100,000 to \$199,999 or 6) \$200,000 or more.

Table 2

Staircase Experiment:

Damages and Contributory Negligence¹⁰⁵

Scenario	Plain-tiff Wins	Contrib. Neg.	Raw Damages		Case Expected Value
			Mean	SD	
1. Control	95	8	\$177,947	\$76,820	\$86,363
2. Not Feasible	96	7	\$178,385	\$73,196	\$84,886
3. Evidence of SRM	128	9	\$177,848	\$87,603	\$120,108
4. J-Instr (Short)	123	6	\$181,094	\$76,202	\$108,102
5. J-Instr w/explain	110	8	\$164,045	\$77,816	\$92,762

When we performed additional regression analysis our model suggested a similar decrease in raw damages, but the results did not reach statistical significance ($p=.189$). Likewise, when we calculated 95% bootstrap confidence intervals, the results also suggested lower damages but were not statistically significant.¹⁰⁶

Interestingly, our model did show statistically significant age effects. For every year older the mock juror was, our regression analysis predicted that raw damages would increase by \$1,258.¹⁰⁷ We did not see a similar effect in experiment two involving a snowboard accident. Apparently, older individuals

105. All figures are in U.S. dollars. Damages exclude zeros (i.e. defense verdicts), while case expected values include zeros.

106. For non-normal distributions, a bootstrapping algorithm is an appropriate method for calculating confidence intervals because it uses the observed distribution rather than assuming normality. *See generally* BRADLEY EFRON & ROBERT TIBSHIRANI, AN INTRODUCTION TO THE BOOTSTRAP 45, 113, 183 (1993). In this case, the 95% confidence interval predicted that damages in scenario 5 would be between \$34,036 *less than* scenario 2 (no remedial measures) and \$6,240 *more than* scenario 2. Because the range crosses zero, these results confirm that our findings are not statistically significant.

107. The 95% bootstrap confidence interval confirmed this effect predicting that damages would increase from \$572 to \$1,882.

determined that the injury caused by falling down stairs deserved more money than younger people.

We do not report on any effects of contributory negligence in our experiment 1. Findings of contributory negligence ranged between a low of six in scenario 4 to a high of nine in scenario 3. These small numbers made any calculations unreliable.

3. Case Expected Value (Staircase)

Finally, we report on case expected value. Case expected value takes into account the damages for both the defense verdicts and the verdicts for the plaintiff. Defendant's wins are treated as zero dollars. For plaintiff wins, we took the raw damages award and reduced the damages for contributory negligence if applicable. As one might expect, the trends appear to track the liability effects. As illustrated in Table 2, the basic case (scenario 2) has a case expected value at \$84,886. When evidence of subsequent remedial measures was introduced (Scenario 3) the value increased to \$120,108. The case expected value with the simple limiting instruction then went back down to \$108,102. That decrease was more substantial when the limiting instruction with explanation was used, decreasing to \$92,762.

We did further analysis to determine if these trends were statistically significant. Our regression analysis predicted that the introduction of subsequent remedial measures increased case expected value by \$36,288. The results were significant ($p=.00066$).¹⁰⁸ While the simple limiting instruction reduced this effect, it did not totally eliminate the effect of the subsequent remedial measures. The case expected value was still \$25,624 higher than scenario 2 (no subsequent remedial measures) ($p=.014$).¹⁰⁹ When the limiting instruction with explanation was used (scenario 5), the case expected value almost returned to its initial value. The case expected value of scenario 5 was only \$7,958 more than Scenario 2, and this difference was not statistically meaningful ($p=.45$).¹¹⁰

4. Summary (Staircase)

In short, our first experiment confirmed intuitions about how evidence of subsequent remedial measures affects liability while also suggesting that limiting instructions may ameliorate those effects. First, as expected, evidence of

108. The 95% bootstrap confidence interval also showed that the case expected value was between \$14,948 and \$57,628 larger when subsequent remedial measures were introduced (Scenario 2 vs. 3).

109. The 95% bootstrap confidence interval shows a similar increase in value of from \$5,087 to \$45,598.

110. The 95% bootstrap confidence interval shows that the case expected value for Scenario 5 is likely to be between \$11,213 less valuable and \$28,401 more valuable than Scenario 2.

subsequent remedial measures increased plaintiff's win rate.¹¹¹ Keep in mind that in this experiment evidence of subsequent remedial measures was introduced in a manner that directly challenged the defendant's credibility. So, there may have been two mechanisms at work: the tendency for the evidence to show that the defendant's behavior was blameworthy, and damage to the defendant's credibility. Second, in this experiment, a limiting jury instruction with explanation reduced liability findings, while a simple jury instruction trended in the same direction but was not statistically significant. Third, we did not observe any significant damages effects, but the limiting jury instructions with explanation appeared to be trending lower. We only mention this observation because of the more significant results found in our second experiment. Because damages did not appear to be affected by our manipulations, our findings on case expected value tracked our findings on liability. Finally, we observed some demographic effects. Older people were more likely to award higher damages, and those with higher income were less likely to find the defendant liable. We suspect the former is likely unique to the facts of experiment one. Older people are likely more concerned about falling down stairs than younger people.

III. EXPERIMENT 2: PRODUCTS LIABILITY

The results of any single experiment may be an artifact of the particular facts presented to the participants. Consequently, we sought to replicate our initial results by testing Rule 407 in a second context with an entirely different fact pattern.

We modeled our second experiment after a products liability case out of the state of Washington, *Hyjek v. Anthony*, which involved a snowboard accident.¹¹² In *Hyjek*, the defendant, Anthony Industries, manufactured a snowboard called the Dan Donnelly XTC, which had no pre-attached bindings and no pre-drilled holes put into the snowboard's fiberglass.¹¹³ Without any pre-attached bindings or pre-drilled holes, Anthony Industries hoped to give more freedom to purchasers in selecting a binding for their snowboard.¹¹⁴ The plaintiff purchased this snowboard and used the threaded screws to affix the bindings of his choice to the Dan Donnelly XTC snowboard.¹¹⁵ While riding the snowboard down the mountain, the threaded screws allowed the binding to come loose and caused the plaintiff to lose control and crash.¹¹⁶ This crash

111. See *supra* Section II.E. (explaining why the defendant's credibility could be playing a part in the increase in liability findings after evidence of subsequent remedial measures is admitted to prove previously controverted feasibility).

112. *Hyjek v. Anthony Indust.*, 944 P.2d 1036, 1036–37 (Wash. 1997).

113. *Id.*

114. *Id.*

115. *Id.*

116. *Id.*

resulted in an injury to the plaintiff.¹¹⁷ After this accident, Anthony Industries redesigned the Dan Donnelly XTC snowboard to have pre-drilled core inserts to affix bindings to the board.¹¹⁸

During the trial, the plaintiff attempted to present evidence of this subsequent redesign to the jury.¹¹⁹ Unlike in our experiment, the plaintiff did *not* argue that an exception to Rule 407 applied, rather the plaintiff argued that Rule 407 did not apply to strict products liability cases.¹²⁰ The Supreme Court of Washington rejected the plaintiff's argument and held that Rule 407 in fact does apply to strict products liability cases.¹²¹

Our experiment used many of the facts mentioned above, but once again, we modified the details and created a fact pattern where subsequent remedial measures would be admitted into evidence. In order to test the underlying premise of Rule 407 and the effect of limiting jury instructions in products liability cases, we created four variations of the same basic case. Besides changing the basic facts, the experiment design also differed from our previous experiment. Our basic scenario already contained an argument where the defendant "opened" the door for the plaintiff to argue that it was entitled to introduce evidence of subsequent remedial measures. Specifically, Scenario 1 included defendant's argument that "the Carve 3000 was the safest board that X5 was able to design that would allow for customers to select their own binding." In Scenario 2, plaintiff introduced evidence of remedial measures in his case in chief, not as a rebuttal as was done in the staircase experiment. Thus, scenarios 1-4 of the snowboard experiment corresponded to scenarios 2-5 of the staircase experiment. We also tested a different exception to Rule 407. Whereas Experiment 1 contained defendant's claims of infeasibility and later impeaching that testimony, Experiment 2 simply allowed an argument regarding safety to be introduced in the plaintiff's case in chief.¹²²

A. Scenario 1: The Basic Case

In our basic case, plaintiff, Connor McNeil, purchased a snowboard, the Carve 3000, from the defendant, X5 Company's ("X5").¹²³ The Carve 3000

117. *Id.*

118. *Id.*

119. *Id.*

120. *Id.*

121. *Id.* at 1042; *see also* WASH. EVID. R. 407.

122. We also tested a second unrelated issue in Experiment 2 which we call low anchoring. Thus, our experiment had eight cells, 4 (Rule 407) x 2 (low anchor). In all the scenarios, the plaintiff asked for approximately \$500,000 in non-economic damages, but in half the scenarios, the plaintiff also demanded \$10,000 economic damages. There is reason to believe that the \$10,000 demand may reduce the ability to obtain \$500,000. For this paper, we collapse scenarios with and without the low anchor together and report on our findings with respect to Rule 407.

123. All names used in our experiment are fictitious.

was designed for snowboarding experts that wanted the flexibility of selecting different bindings. Consequently, the board did not have any predrilled holes. Unlike the real case, the liability issue did not focus on the proper type of screws, but instead involved the size of the hole that needed to be drilled into the snowboard. The failure to drill the correct size made the snowboard unstable. Further, in order to provide an opening for the plaintiff to introduce evidence of subsequent remedial measures, the defendant argued that its product was the “safest” possible that would allow the rider the freedom to pick his or her own bindings.¹²⁴

The plaintiff injured himself during his first run with the new snowboard. The parties agreed that the holes he drilled were too big, thereby making the board difficult to handle. There was also evidence that X5 knew that this could be a problem. A test rider had previously made the same mistake with less serious consequences.

Mr. McNeil alleged that the Carve 3000 was defective for two reasons: (1) it was foreseeable that purchasers of the Carve 3000 would make mistakes drilling holes and injure themselves, and (2) X5 did not warn its customers of what would happen if the holes were not drilled properly. Consequently, McNeil asked for \$500,000 to compensate for his injuries. Those injuries included a concussion and tears of two ligaments in his knee. While most of the damage was surgically repaired, Mr. McNeil was told that he would be unable to snowboard in the future.

X5 denied that the failure to predrill holes made its product dangerous or required a warning and argued that the design with no pre-drilled holes was the safest and only design possible to allow the purchaser to choose any bindings they wanted. Of course, this statement opened the door for the plaintiff to introduce evidence of subsequent remedial measures to impeach this argument.

Notably, unlike the staircase experiment where the theory of liability was based on negligence, Experiment 2 was a products liability case that relied on strict liability. Thus, the jury was charged with slightly different decision – deciding whether the product was unreasonably dangerous.

B. Scenario 2: The Presentation of Subsequent Remedial Measures

In Scenario 2, the plaintiff introduced evidence of subsequent remedial measures. More concretely, the plaintiff explained that the defendant began using “different sized core-inserts” after the accident. The different inserts still allowed the customer the freedom to choose different bindings, but customers no longer had to drill holes. As a result, the mistake that caused the plaintiff’s injury was no longer possible. The evidence was admissible because it contradicted the defendant’s argument that the original Carve 3000 snowboard used the safest possible design.

124. As discussed earlier, attorneys can often manipulate a witness into making this argument. *See supra* note 29.

Unlike the staircase experiment, the evidence of remedial measures was introduced in the plaintiff's case in chief.¹²⁵ Thus, it did not serve to contradict the defendant's argument in the same way as in the staircase experiment. Accordingly, any differences between scenarios 1 and 2 are probably attributable to the tendency for the evidence to show that the defendant's behavior was blameworthy and not to any damage done to the defendant's credibility.

C. Scenarios 3 and 4: Limiting Jury Instructions

Scenarios 3 and 4 added different limiting jury instructions. However, unlike the first experiment, we did not interrupt the plaintiff's presentation with the limiting instruction. Instead, immediately after the plaintiff explained the subsequent remedial measures to the jury, the attorney stated, "I am only using this evidence to show that the defendant's original design was not as safe as they could make it. That is an argument that the defendant has raised. But it is just not true. They clearly made a safer design later. I am *not* using this evidence to show that the original product was defective."¹²⁶

Later, the judge gave a similar instruction to the jury at the end of the trial when he gave all of the jury instructions. The visual in Figure 1 accompanied the instructions. More specifically, the judge stated, "During his argument, Plaintiff's counsel offered evidence that X5 redesigned its product to include holes and use core inserts. This evidence was only admitted for the limited purpose of challenging the defendant's argument that this defendant's original design was the safest one possible. You must consider it for that purpose only and may not consider it as evidence that the product was defective." These instructions were almost exactly the same as the instructions given in the premises liability case, which were modeled after the instructions given in *Duncan v. Mill*.¹²⁷

In Scenario 4, the judge then added his explanation for the limiting jury instruction by stating, "We have this rule for two reasons. First, we have this rule because we want to encourage people to fix problems and the law will not use evidence of such fixes against them later in court. Second, we are concerned that the jury may overweigh evidence of remedial measures. Although it may tend to show that the original version was defective, it is possible that companies can improve products that were not defective." This instruction was slightly different than the corresponding instruction given in Experiment 1. In Experiment 1, the judge only used the first rationale to justify the instruction. The judge did not mention the possibility of hindsight bias.

125. See *supra* Section II.B.; see also Hyjek, 944 P.2d at 1037.

126. See App. C for the slide associated with the limiting instruction.

127. *Duncan v. Mill Mgmt. Co. of Greenwich*, 60 A.3d 222, 230 (Conn. 2013).

D. Respondents

We recruited subjects from the population of workers on Amazon Mechanical Turk in the spring of 2018 and established an initial screen for those that were “jury eligible” (i.e. residents of the United States over age 18 who could read, write, and speak English). Subjects were paid \$3.00 to complete the experiment online. All subjects gave consent in accordance with the Institutional Review Board requirements. At the beginning of the survey, subjects filled out a demographic questionnaire.

864 participants started the experiment. After disqualifying participants for failing to complete the experiment, failing attention checks, or providing inconsistent responses, 729 participants remained. This sample was nearly a 50/50 split between males and females with males representing 51.4%. The sample was younger, more educated, and more politically liberal than the population at large; gender, race, and median income, however, were more representative of the U.S. Census data.¹²⁸

E. Results

In Experiment 2, the defendant argued that it made its snowboard as safe as it could in every scenario. Thus, unlike Experiment 1, there was no basic scenario where the defendant did not open the door to allow the plaintiff to introduce evidence of subsequent remedial measure. Consequently, Scenario 1 in Experiment 2 corresponded to Scenario 2 in Experiment 1. Scenario 2 corresponded to Scenario 3 of Experiment 1 and so on. With that in mind, we report on our results.

1. Liability (Snowboard)

As seen in Table 3, the plaintiff’s win rate was 33% in scenario 1. Scenario 2 introduced the evidence of subsequent remedial measures. As it did in Experiment 1, plaintiff’s win rate increased. Here, the increase was 11.7% resulting in a 44.7% win rate for the plaintiff. Scenarios 3 and 4 introduced the simple and more detailed limiting instructions, respectively. In Scenario 3, the simple limiting instruction appeared to do nothing to the win rate which remained essentially unchanged (44.2%). However, the plaintiff’s win rate dropped (5.8%) to 38.9% when the limiting instruction with explanation was introduced in scenario 4.

128. Specifically, the sample demographics are as follows: 49% female, 76% white, 12% African American, 9% Asian, 1% American Indian or Native American, and the remainder classified as other; mean and median age 37 and 34 respectively; 54% lean toward or strongly prefer democrats with only 25% leaning or strongly preferring republicans; the median education level for our sample was a bachelor’s degree; and the sample’s median income fell between \$50,000–\$99,999.

Table 3

Snowboard Experiment, Liability

Scenario	No. of Mock Jurors	Plaintiff Wins	Win Rate
1. Basic	176	58	33.0%
2. Evidence of SRM	188	84	44.7%
3. Simple J-Instr.	190	84	44.2%
4. J-Instr. w/explanation	175	68	38.9%

Again, we performed a logistic regression controlling for relevant demographics to determine if the effects on plaintiff's win rate were significant.¹²⁹ Introducing subsequent remedial measures in Scenario 2 increased plaintiff's win rate (odds ratio 1.874 to 1). This difference was significant ($p=.0053$). A comparison of Scenarios 2 and 3 found that the simple limiting jury instruction did not have any effect on liability (odds ratio .936 to 1, $p=.76$). When we compared Scenarios 2 and 4, the effect size was larger, but still not statistically significant (odds ratio .787, $p=.28$).¹³⁰

We also combined our data from the two experiments to see if the added power would reveal an effect. We used the scenarios where evidence of subsequent remedial measures had been introduced without a jury instruction as the baseline case. When a simple limiting instruction was added, the trend pointed in the hypothesized direction (a lower plaintiff win rate), but the results were not significant (odds ratio .824 to 1, $p=.20$). However, when limiting instructions with explanation were added, there was a statistically significant decrease in plaintiff's win rate (odds ratio .687 to 1, $p=.014$).

Specific demographics also affected liability in the snowboard experiment. Men were less likely to find in favor of the plaintiff than women (odds ratio .487 to 1, $p < .001$).¹³¹ That may be because men are less willing to place blame on others when the injury is caused by a high risk activity like expert snowboarding. But our experiment was not designed to delve into the reasons behind this unexpected effect, and we cannot be sure about the mechanism. Additionally, the analyses showed that two ethnic groups were more likely to

129. *See infra* app. B.

130. *Id.*

131. *See infra* app. B.

find liability than white mock jurors – African American mock jurors (odds ratio 2.07, $p=.0029$) and Asian mock jurors (odds ratio 1.798, $p=.037$) – but we do not have any hypotheses to explain these results.¹³²

2. Raw Damages (Snowboard)

Table 4 examines the different components of damages starting with raw damages. Raw damages exclude defendant wins and do not discount for findings of contributory negligence. A comparison of scenarios 1 and 2 suggests that evidence of subsequent remedial measures *might* decrease raw damage awards. In scenario 1, the average plaintiff's damage award was \$310,465. When evidence of subsequent remedial measures was introduced in scenario 2, the average award dropped to \$259,446. Perhaps, mock jurors viewed the defendant more favorably (*i.e.*, as a better corporate citizen) when it took corrective action. That view could manifest in reduced damage awards.

132. *Id.*

Table 4

Snowboard Experiment:

Damages and Contributory Negligence (“CN”)¹³³

Scenario	Pl Wins	#CN	% CN	Raw Damages		Case Expected Value
				Mean	SD	
1. Basic	58	18	31.0%	\$310,465	\$192,556	\$91,851
2. Evid. of SRM	84	39	46.4%	\$259,446	\$182,554	\$98,567
3. Simple J-Instr	84	40	47.6%	\$258,762	\$180,295	\$94,718
4. J-Instr w/explain	68	20	29.4%	\$312,934	\$170,942	\$108,744

Our regression analysis suggested that raw damages decreased by \$55,612 when subsequent remedial measures were introduced in scenario 2, but these results were just short of statistical significance ($p=.076$). Because the damage data was not normally disturbed, we also calculated bootstrap confidence intervals. That analysis gave us similar results, predicting that there was a 95% chance that raw damages in scenario 2 would be between \$116,266 *lower* to \$10,782 *higher* than scenario 1. Because this range crossed zero by a small amount, these results were also just short of statistical significance. The simple jury instruction in scenario 3 did not change these results in any meaningful way. The model showed that raw damages in scenario 3 were \$54,055 less than scenario 1, but these results were short of statistical significance again ($p=.085$).¹³⁴ Somewhat unexpectedly, the more complex limiting instruction in scenario 4 appeared to return raw damages back to the value they had in scenario 1. The model showed that raw damages would be \$3,548 less in scenario 4 than in scenario 1. This number was statistically meaningless ($p=.91$).¹³⁵

133. All figures are in U.S. dollars. Damages awards over \$500,000 are transformed to \$500,000. Damages exclude zeros (i.e. defense verdicts), while case expected values include zeros.

134. The 95% confidence interval predicted that raw damages in scenario 3 could be \$116,311 lower to \$10,025 higher than in scenario 1.

135. The 95% confidence interval predicted that raw damages in scenario 4 could be \$67,589 lower to \$60,477 higher than in scenario 1.

3. Contributory Negligence (Snowboard)

When plaintiffs are found to be contributorily negligent, their damages awards are reduced by the percentage of fault. A comparison of Scenarios 1 and 2 suggests that evidence of subsequent remedial measures *might* be increasing findings of contributory negligence. Finding that the plaintiff also bears some responsibility for an injury might be another way for juries to express approval for a defendant that takes corrective actions.

Table 3 shows that as evidence of subsequent remedial measures was introduced in Scenario 2, the rate of contributory negligence findings increased from 31.0% to 46.4% (39/84). That rate remained substantially unchanged at 47.6% when the simple limiting jury instruction was introduced in Scenario 3. But the rate of contributory negligence dropped to 29.4% (20/68) when the more limiting instruction with explanation was introduced in Scenario 4. Our regression analysis showed that these differences were not quite statistically significant.¹³⁶

There were, however, significant demographic effects based on income and race. For every step increase on our income scale, respondents were more likely to find that the plaintiff was contributorily negligent (odds ratio to 1, $p=.032$). At the same time, African American respondents were less likely than white respondents to find that the plaintiff was contributorily negligent (odds ratio to 1, $p=.022$).

4. Net Damages (Snowboard)

Both raw damages awards and findings of contributory negligence trended in directions helpful to the defendant when evidence of remedial effects were introduced. But neither result reached statistical significance.¹³⁷ This caused us to examine the combined effects of these two outcomes. We call this net damages. Net damages exclude defendant wins and reflect the fact that awards plaintiffs eventually receive are reduced when there are findings of contributory negligence.

The regression analysis for net damages did show statistically significant differences. Damages in Scenario 2 decreased by \$62,690 ($p=.043$) when evidence of subsequent remedial measures was introduced.¹³⁸ The simple limiting

136. See *infra* app. B at Table B3. Comparing scenario 1 vs 2, the odds ratio was 1.822 to 1 ($p=.105$).

137. We also examined whether the percent fault attributed to plaintiff varied when evidence of subsequent remedial measures was introduced. We saw no evidence of this effect and therefore do not report the numerical results here. Because this analysis only looks at plaintiff wins with findings of contributory negligence, our study was inadequately powered to study this question well.

138. See *infra* app. B, Table B4. Our bootstrap confidence interval showed similar results. There was 95% likelihood that net damages in Scenario 2 was between \$124,753 and \$653 lower than scenario 1.

instruction introduced in Scenario 3 did not change this effect. Net damages were still \$66,780 less than Scenario 1 ($p=.039$).¹³⁹ Interestingly, when we introduced the jury instruction with explanation in Scenario 4, the effect appeared to substantially disappear. In our model, net damages in Scenario 4 were only \$4,393 less than Scenario 1.¹⁴⁰ Thus, it appears that defendants were able to reduce net damages by taking subsequent remedial measures. Oddly, a limiting jury instruction with an explanation of the basis for Rule 407 returned net damages to baseline.

5. Case Expected Value (Snowboard)

Table 4 above reports on the case expected value for each different scenario. Interestingly, the case expected value did not vary much between scenarios. In Scenario 1, the case expected value was \$91,851. As evidence of subsequent remedial measures was introduced in Scenario 2, the case expected value increased to \$98,567. The value dropped a small amount to \$94,718 as the simple limiting instruction was added in Scenario 3. The highest case expected value, \$108,744, was found when the limiting instruction with explanation was used in Scenario 4. However, our regression analysis suggested that none of these differences were statistically significant.

The only significant effects were related to demographics. The case expected value decreased \$47,388 when males were mock jurors as opposed to females ($p=.00013$) and increased \$54,286 when African American respondents decided the case as opposed to white respondents.

6. Findings

We draw several conclusions from Experiment 2. First, Experiment 2 confirmed that evidence of subsequent remedial measures increased plaintiff's win rate. Although the impact was smaller than we saw in Experiment 1, it was still sizable and significant. Interestingly, when the mock jurors saw evidence of subsequent remedial measures, the plaintiff's win rate increased but net damages decreased. Although we certainly cannot identify the causal mechanism, it may well be that the defendant's decision to take remedial measures made them less blameworthy and therefore deserving of smaller damages. Because liability and damages went in opposite directions, we saw no statistically significant differences between any of our scenarios when we examined case expected value.

139. *See supra* Table 4. The 95% confidence interval looked substantially like it did for Scenario 2. Net damages for Scenario 2 was between \$128,041 and \$3,419 lower than scenario 1.

140. *Id.* The 95% confidence interval suggested that net damages in Scenario 4 could be anywhere between \$68,477 lower and \$59,308 higher than Scenario 1.

IV. DISCUSSION

In this section, we discuss what our results mean for policymakers, litigators, and future researchers. As part of this discussion, we describe the limitations of our work.

A. Implications

1. Policymakers

Our results confirm that evidence of subsequent remedial measures substantially increases the likelihood that a defendant will be found liable. The size of the effect varied in our two experiments. Important differences in the two experiments likely caused this variance. First, in the staircase experiment, two cognitive mechanisms were likely at work, but one of these mechanisms was not present in the snowboard experiment. In both experiments, evidence of subsequent remedial measures likely suggested that the defendant should have taken the action earlier. This is the kind of hindsight bias that often justifies Rule 407. But in the staircase experiment, the evidence of subsequent remedial measures (a new staircase with a landing) was introduced by the plaintiff in a separate rebuttal that focused on impeaching the defendant's argument of lack of feasibility (space was too limited for such a staircase). By contrast in our snowboard experiment, the evidence of subsequent remedial measures was introduced in the plaintiff's case in chief to suggest that there was a safer snowboard design (pre-drilled holes). This evidence did not place defendant's credibility at issue in the same way as the staircase experiment. This might explain why the effect of evidence of subsequent remedial measures was larger in the staircase experiment. Alternatively, the difference in effect size might be attributable to the particular subsequent remedial measure found in each case or even case specific facts.

Our results also suggest that limiting jury instructions can reduce, but not eliminate, the effect of subsequent remedial measures on liability findings. Moreover, consistent with Diamond and Casper,¹⁴¹ the instruction that explained the basis for Rule 407 appeared to be more effective than the simple instruction. But our conclusions about limiting instructions are tentative at best. While the results trended in directions consistent with our hypotheses, the only difference that reached statistical significance was found in the staircase experiment. The difference in liability rates of the jury instruction with explanation and the scenario where subsequent remedial measures were introduced but no limiting instruction was given was significant.¹⁴² Moreover, the difference between the simple limiting jury instruction and the instruction with

141. See *supra* note 74 and accompanying text.

142. See *supra* Part II.E.1 for the precise statistical analysis and *infra* Table 1 for an intuitive understanding of this effect.

explanation was only suggestive and did not reach statistical significance in either of our two experiments.

Unfortunately, these results do not provide clear answers for judges. The problem is that no one knows precisely how much weight jurors should give to evidence of subsequent remedial measures when deciding liability. If a judge believes that jurors should give some limited weight to evidence of subsequent remedial measures, then it may make sense to admit the evidence and provide a limiting instruction with an explanation of the basis of the rule. If the judge believes that the evidence deserves more weight, the judge can issue a simple instruction instead. On the other hand, if the judge believes that the evidence has no value, our results suggest that limiting instructions will be unable to counteract the entire effect that evidence of subsequent remedial measures has on liability decisions. But the problem with all these recommendations is that our results merely point in directions and do not give precise estimates. Moreover, even if these estimates were more precise, different facts undoubtedly would change these estimates.

Damages effects complicate this picture even further. In the snowboard experiment, damages decreased and essentially counteracted the increased findings of liability.¹⁴³ That result did not happen in the staircase experiment. It is unclear how, or even if, courts should take into account how evidence of subsequent remedial measures affects damages. Perhaps, admitting evidence of subsequent remedial measures might actually incentivize corrective action because it will lower damage awards.

2. Litigants

Our results confirm the widely held notion that evidence of subsequent remedial measures leads jurors to find defendants liable at a higher rate.¹⁴⁴ That might suggest that defendants should be wary about taking remedial measures. But there are two reasons why this concern might be overstated. First, jury instructions appear to reduce these liability effects.¹⁴⁵ Moreover, an instruction that explains the basis for Rule 407 appears to be slightly more effective than a simple jury instruction.¹⁴⁶ We provide this recommendation with an important caveat. For the most part, our results on jury instructions did not reach statistical significance, but only suggested possible trends. Researchers certainly will not (and should not) accept these findings as conclusive. But at this point, there is no other data for litigants to consider. Second, and perhaps more surprising, evidence of remedial measures might actually reduce a defendant's damages.¹⁴⁷ This effect was present (and statistically significant) in

143. See *supra* Part III, Table 3, Table 4, and accompanying text.

144. See *supra* Table 1, Table 3, and accompanying text.

145. *Id.*

146. *Id.*

147. See *supra* Table 2, Table 4, and accompanying text.

the snowboard experiment, but not the staircase experiment. So how is a litigant to use this information? We suspect that damages will only decrease if the specific remedial measures at issue make the defendant look like a good citizen. At this point, we cannot prove that this is the mechanism for lower damages awards, however, this hypothesis is consistent with other studies that found that juries award higher damages when the conduct was more blameworthy.¹⁴⁸

For plaintiffs, evidence of subsequent remedial measures appears helpful on liability. Thus, if liability is hotly contested, a plaintiff's attorney should attempt to find an exception to introduce evidence of subsequent remedial measures. On the other hand, if liability is straightforward and the real dispute lies in the proper amount of damages, a plaintiff's attorney may want to steer clear of presenting such evidence as it may lead to a lower damages award. Obviously, defense attorneys should consider the opposite approach. In addition, if evidence of subsequent remedial measures is admitted, they should ask for a limiting instruction that explains the basis underlying Rule 407.

3. Researchers

With one notable exception, our study did not provide entirely consistent results. That is probably because cases in our legal system are complex and decision-making often depends on the particular facts being tested. The staircase and snowboard experiments provide a good example of how the same kinds of decisions can take place in very different contexts. The two experiments had very different accused tortious acts (steep stairs vs. defective binding). The plaintiffs were also engaged in very different activities (walking down stairs vs. extreme snowboarding). The defendants took different subsequent remedial measures (rebuilding staircase vs. redesigning binds). Procedurally, the evidence of subsequent remedial measures was introduced at different stages in the case. Thus, it should come as no surprise that the results in these experiments differed. Unfortunately, that means the researcher's role is not easy. To the extent any study suggests particular insights, the results needs to be tested in different contexts to determine how well the insight holds up when combined with different factors. It's quite likely that there are factors that may interact in ways that enhance, reduce, or even eliminate many such findings.

While most studies have suggested that mock jurors are unable to disregard information on command, particularly when the instruction suggests that they only ignore evidence for one purpose but not another, our results were not consistent with this view. Even though we only had a single statistically significant finding, three of our four results suggested that limiting instructions

148. *See supra* note 85 and accompanying text.

had the desired effect, albeit modestly.¹⁴⁹ While our results were not conclusive, they seemed to suggest that a limiting instruction with an explanation of the instruction's purpose can have some of the desired effect. Thus, our result appears to support Casper and Diamond's findings.¹⁵⁰

Our experiment was not designed to identify the precise mechanism behind this effect. Surely individuals cannot delete evidence from their thoughts.¹⁵¹ However, if jurors believe that the limiting instruction makes good sense, they may try to make up for the inability to forget the evidence by simply putting a thumb on the scale in the other direction. If we consider this possibility in terms of the Hans-Reyna gist based model of decision-making, a limiting instruction could cause a juror to simply lower the degree of defendant's fault on the ordinal scale. Future researchers might want to explore this possibility.

B. Limitations

Our study had several limitations. First, two experiments exposed mock jurors to one of several different scenarios. The effects we tested may be moderated or enhanced by different arguments and presentations. We did not test how different arguments or presentation graphics might change our results. For example, our first experiment involved a moderate injury involving stairs, but an example with more serious injuries involving a more complex mechanism for injury, such as a car, may have different effects on juries and how they perceive subsequent remedial measures. Further, our findings from version one and two (which indicated that the win rate went up slightly for the plaintiff when the defendant controverted the feasibility of making the stairs less steep) may have been affected by the types of arguments used and may render different results with a different style or argument. Indeed, just within our study, we found several effects that were only found in one experiment. For example, we found that older people tended to award more damages in the staircase experiment, but not the snowboard experiment.¹⁵² Similarly, African Americans viewed liability differently in snowboard experiment than whites, but race effects were not present in the staircase experiment.¹⁵³ This suggests that other

149. In the staircase experiment, the limiting instruction with explanation yielded a statistically significant decrease in plaintiff's win rate ($p=.0022$) and the simple limiting instruction was trending in that direction ($p=.098$). In the snowboard experiment, the limiting instruction with explanation also trended in the same direction ($p=.28$). Admittedly, a fourth result, the simple limiting instruction in the snowboard experiment, did not see almost any change in plaintiff's win rate ($p=.76$).

150. See Diamond & Casper, *supra* notes 74 and accompanying text.

151. See SAKS & SPELLMAN, *supra* note 64, at 89 (suggesting that one reason jurors do not disregard information is because they cannot wipe it from their memory).

152. See *supra* Part II.E.2.

153. See *supra* Part III.E.3.

unknown factors may be interacting with the variables we were testing. Put more simply, the hypotheses we tested only tell part of a more complex story.

Second, our experiments were limited to an abridged civil trial that lasted from approximately fifteen to seventeen minutes where the jurors could only hear the voices of the lawyers, experts, and the judge. Although the respondents viewed PowerPoint presentations, they were not able to see the attorneys, experts, or judge. Thus, the respondents could not witness the different presenters' demeanor or consider other visual cues. This might affect verdicts, but this design also had advantages. Because our experiment consisted of summaries of both parties' arguments, rather than actually having the jurors listen to a full trial in the normal format, we were able to utilize a randomized controlled trial experimental design, which is the gold-standard for scientific research.¹⁵⁴

However, there are still reasonable concerns about external validity. One might expect that the plaintiff's win rates may be increased or decreased if the respondents were to hear more detailed trial evidence. However, we structured the abridged case to give both sides a comparably equal amount of persuasive evidence through two different expert witnesses. Thus, a longer trial with more evidence may not have affected the results in any significant way. Thus, for mock jury research, a 15-minute audio stimulus with a PowerPoint presentation, complete with jury instructions, opening statements, expert witnesses, and arguments, is at the high end of the range of external validity, compared to other studies which might use a 5-minute paper-and-pencil task.

Third, we did not study real jurors. Prior research has shown that "the population of Mechanical Turk is at least as representative of the U.S. population as traditional subject pools."¹⁵⁵ Known experimental results have been replicated using the MTurk population.¹⁵⁶ Nonetheless, MTurkers may be more easily distracted from the trial compared to real jurors and may even provide junk responses (e.g., those who failed to watch the entire video without hearing all the arguments and rendered a verdict). It may be that real jurors are more earnest in their efforts to providing meaningful responses or that real jurors determine liability differently knowing that the outcomes will impact real individuals and companies.

Lastly, our study involved single jurors and the outcomes they arrived at on their own. Our study did not allow jurors to deliberate with other jurors in the manner that would occur if a real trial were conducted. We are unaware of how any such difference would interact with the hypotheses here tested. More-

154. Leon Gordis, *Introduction to Scientific Methodology*, 9 KAN. J. L. & PUB. POL'Y, 34, 43-44 (1999) (noting that the randomized trial is the "gold standard" of study design and explaining how randomization shields the study from the investigator's preconceptions and biases).

155. Gabriele Paolacci et al., *Running Experiments on Amazon Mechanical Turk*, 5 JUDGMENT & DECISION MAKING 411, 411 (2010).

156. See Irvine et al., *supra* note 90 at 344; Berinsky, *supra* note 94, at 361-65.

over, “Research has consistently shown a strong and robust relationship between the verdict preferred by the majority of jurors at the start of deliberation and the jury’s ultimate verdict.”¹⁵⁷ The issue of how *juror* decisions on damages predict *jury* decisions on damages is less developed. The work that has been done suggests that individual juror decisions modestly underestimate what juries will decide after deliberation.¹⁵⁸ But the point of our experiment is to test the difference in damages between experimental conditions, not to estimate absolute values.

CONCLUSION

We conducted two large scale experiments to test the assumptions underlying Federal Rule of Evidence 407 and how it is applied by the courts. Unlike prior studies, we used longer video vignettes that included arguments from the plaintiff and the defendant as well as jury instructions. Like prior studies, mock jurors determined whether the defendant was liable, but unlike any other studies to date, we also asked mock jurors to decide issues of contributory negligence and damages. As expected, evidence of subsequent remedial measures helped plaintiffs win more often. But surprisingly, our results also suggested that taking remedial measures may lower net damages (damages after considering contributory negligence) under certain conditions, thereby counteracting the increased liability findings. We also studied the efficacy of two limiting jury instructions. In one experiment, a limiting instruction with an explanation reduced but did not eliminate the effects of evidence of subsequent remedial measures. The instructions with explanation were also consistently more effective than the simple limiting instructions, but these results were not statistically significant.

157. DENNIS J. DEVINE, *JURY DECISION MAKING: THE STATE OF THE SCIENCE*, 158 (2018); see also Maggie Wittlin, *The Results of Deliberation*, 15 U.N.H. L. Rev. 161, 185 (2017) (summarizing studies that show that individual juror decisions are predictive of jury decisions as a whole).

158. *Id.*; see also, Shari Seidman Diamond et al., *Juror Judgments About Liability and Damages: Sources of Variability and Ways to Increase Consistency*, 48 DEPAUL L. REV. 301, 316–17 (1998) (“jury awards in this case were higher than the average mean and median juror awards, a pattern found in several other studies of damage awards.”); Diamond & Casper, *supra* note 74, at 553 (“A clear inflation of damage awards occurred between the individual and the group level. On average the juries produced awards about \$56,000 (or 26%) higher than the average of their members prior to deliberation.”).

APPENDIX A

Staircase Analyses

Table A1 shows the results of a regression analysis to find the odds ratio of a plaintiff winning on liability. The (Intercept) value corresponds to the odds of finding of liability in the reference case where there was evidence of remedial measures but no limiting jury instructions with a white female whose age and income are the mean of the sample. Coefficients for categorical variables show the odds ratio is situation/demographics are changed from the reference case. The ethnicity odd ratios are given relative to the category “White.” The gender odds ratio is given relative to the reference category of female.

Table A1

Staircase, Liability (Plaintiff Prevailing)

Variable	Odds Ratio	p value
Intercept	2.20	0.0000184***
No Subsequent Remedial Measures	0.417	0.0000482***
Simple Limiting Jury Instruction	0.700	0.0979 ‘.’
Limiting Jury Instruction with Explanation	0.611	0.0240 *
Income	0.812	0.00215 **
Age	0.993	0.288
Gender (Male)	0.912	0.543
African American	1.369	0.221
American Indian, Alaska, Native Hawaiian, Pacific Islander, and Other	1.101	0.796
Asian	1.259	0.412

Significance codes for all tables: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘.’ 1

Table A2
Staircase, Raw Damages

Variable	Damages in Dollars	p value
Intercept	182,882	< 2e-16 ***
No Subsequent Remedial Measures	655.1	0.950
Simple Limiting Jury Instruction	3631.7	0.733
Limiting Jury Instruction with Explanation	-14,339	0.189
Income	-3,013	0.370
Age	1,258	0.000101 ***
Gender (Male)	-2,875	0.700
African American	-10,370	0.373
American Indian, Alaska, Native Hawaiian, Pacific Islander, and Other	-23,750	0.178
Asian	-21,280	0.119

Table A3

Staircase, Case Expected Value

Variable	Case Expected Value (Dollars)	p value
Intercept	86,758	< 2e-16 ***
No Subsequent Remedial Measures	36,288	0.000639 ***
Simple Limiting Jury Instruction	25,623	0.0138 *
Limiting Jury Instruction with Explanation	7,958	0.451
Income	-10,990	0.000955 ***
Age	440	0.186
Gender (Male)	-7,062	0.349
African American	5,315	0.670
American Indian, Alaska, Native Hawaiian, Pacific Islander, and Other	-10,689	0.562
Asian	-170	0.990

Multiple R-squared: 0.03326, Adjusted R-squared: 0.02321

APPENDIX B

Snowboard Analyses

Table B1

Snowboard, Liability

Variable	Odds Ratio	p value
Intercept	1.134	0.520
No Subsequent Remedial Measures	0.534	0.00531 **
Simple Limiting Jury Instruction	0.936	0.758
Limiting Jury Instruction with Explanation	0.787	0.275
Income	0.931	0.321
Age	0.988	0.0906 ‘.’
Gender (Male)	0.487	0.00000765 ***
African American	2.0660	0.00286 **
American Indian, Alaska, Native Hawaiian, Pacific Islander, and Other	0.994	0.989
Asian	1.798	0.0372 *
Low Anchor (True)	0.819	0.202

Table B2
Snowboard, Raw Damages

Variable	Damages in Dollars	p value
Intercept	305,421	<2e-16 ***
Subsequent Remedial Measures	-55,612	0.0763 ‘.’
Simple Limiting Jury Instruction	-54,055	0.0847 ‘.’
Limiting Jury Instruction with Explanation	-3,548	0.914
Income	6,759	0.481
Gender (Male)	-19,639	0.336
African American	4,844	0.867
American Indian, Alaska, Native Hawaiian, Pacific Islander, and Other	-15,110	0.799
Asian	-48,569	0.184
Low Anchor (True)	44,475	0.0375 *

Table B3
Snowboard, Contributory Negligence

Variable	Odds Ratio	p value
Intercept	0.513	0.0473 *
Subsequent Remedial Measures	1.822	0.105
Simple Limiting Jury Instruction	1.863	0.0922
Limiting Jury Instruction with Explanation	0.842	.669
Income	1.322	0.0150 *
African American	0.419	0.0167 *
American Indian, Alaska, Native Hawaiian, Pacific Islander, and Other	0.324	0.172
Asian	0.795	0.586
Male	1.267	0.360
Low Anchor	1.0360	0.887

Table B4

Snowboard, Net Damages

Variable	Net Damages in Dollars	p value
Intercept	279,495	<2e-16 ***
Subsequent Remedial Measures	-62,690	0.0427 *
Simple Limiting Jury Instruction	-66,780	0.0309 *
Limiting Jury Instruction with Explanation	-4,393	0.892
Income	-514	0.957
African American	20,734	0.467
American Indian, Alaska, Native Hawaiian, Pacific Islander, and Other	-23,187	0.692
Asian	-52,384	.146
Male	-18,526	0.394
Low Anchor	25,852	0.219

Table B5
Snowboard, Case Expected Value

Variable	Case Ex- pected Value in Dollars	p value
Intercept	104,747	1.79e-11 ***
Subsequent Remedial Measures	12,635	0.464
Simple Limiting Jury Instruction	4,829	0.778
Limiting Jury Instruction with Explanation	22,811	0.194
Income	-5,054	0.366
African American	54,286	0.00465 **
American Indian, Alaska, Native Hawaiian, Pacific Islander, and Other	-7,072	0.824
Asian	12,250	0.580
Male	-47,389	0.000126 ***
Low Anchor	-432	0.972

APPENDIX C



