Revision 2: Responses to Reviewers Round 2

We would again like to thank the reviewers for their in-depth reviews and the suggestions regarding the manuscript. Please find our detailed response addressing each of the points below.

## Editorial changes

1. Headers are included in the abstract and they do not need to be. *Done*
2. Several references include the DOI and year, volume page citations. Currently we only need the latter if that is available. Please correct. *Done*
3. Reference 4 is a URL and when checked it indicated that it was no linger updated. It also did not take an interested reader to the PCST report. Please use a different reference. *Done*
4. Reference 9 is another URL and my check indicated that the page could not be found. Please correct. *Done*
5. Please follow the formats included in the Information for Authors and abbreviate the journal titles according to Index Medicus *Done*
6. Reference 8 was incomplete. Please complete per the Information for Authors. *Done*

## Reviewer #1

Comments to the Author

This revision is a major improvement from the initial submission, congratulations.

1. Pg.4, ln 14: Change “impurities” to “surface imperfections”

*Thank you, we have incorporated this change*

1. Pg. 4, ln. 45; change to “(because the heel typically has the most contact with the rifling in the barrel)”

*Thank you, we have incorporated this change*

1. Pg.4, last sentence. Please check for grammar or incomplete information. Also capitalize “Lowess”. (?) Perhaps: “From these profiles, bullet signatures (16; 15) are extracted as residuals …”

*Thank you, we have changed the grammar and incorporated this change*

1. Pg.5, ln. 36: Change to “because of the resulting loss in discriminating power.”

*Thank you, we have incorporated this change*

1. Pg.7 and on; Please add some text to clarify your definition of “sub-class”. You are designating this term as a general form or structure that has to be removed to get at the signature. However, the term is used quite differently by firearm examiners to describe features that may reproduce in a number of tool surfaces, and its influence if detected must be considered prior to a comparison conclusion.

\*Thank you for this comment. You are right, the term sub-class characteristics is used by FTEs to describe markings that are shared across multiple tools. Some of these sub-class characteristics will be expressed in markings that can be removed by using the normalization methods suggested. However, normalization can del with more general structure, but at the same time will not address all sub-class structures. We have tried to clarify this point by introducing the following paragraph on page 6:

The implementation of the Chumbley score in Hadler and Morris (1) uses a normalization step before going into the optimization and validation step described below. Normalization is done by using a s LOWESS smooth to reduce extraneous structure in the markings, such as a drift, or spatial trends introduced during the barreling. To a degree, this normalization can also be used to address problems stemming from sub-class characteristics, i.e. markings in a pattern that are not unique to a single barrel but shared across a group of barrels introduced by specifics in the manufacturing process. \*

1. Pg. 8; Thanks for pointing out what “failed test” is, and what it is not.

*We are happy that we could clarify this*

1. Pg. 10, ln. 32: “tank rash” is a is a colloquialism that should be in quote marks since it would be unknown to many scientists outside of the forensic discipline. Or longer: “random tool marks on the fired bullet surface caused by the impact with the interior surfaces of the bullet capture tank.”

*Thanks a lot for pointing this out, we have made the changes in the paper reflecting this*

1. Pg. 10, ln. 7: fix error “toolmaRkpack-age” ln. 48 space is needed between “profiles of0.854”

*Thank you, we have incorporated this change*

## Reviewer #2

Comments to the Author Thank you for making the requested revisions. I appreciate that this takes time. A few issues remain but I believe we are getting close.

1. p3 “These digitized markings allow the use of statistical methods to quantify the scientific mechanism of comparing markings and serve as basis for an error rate calculation.” Still a little awkward.

*Thank you for the comment. We have changed the sentence to: This data forms the basis of a statistical analysis that allow us to quantify similarity of markings and serves as a basis for error rate calculations*

1. The authors claim that they changed all numerical results to percentages but there are still a number of inconsistencies. The authors should use consistent terminology. Sometimes they say FPR, sometimes %FP, sometimes Type I, sometimes Type 1 error rate. It is ok to list all these as being the same term in the definition, but then they should pick one. For example, in the conclusions CS1 is reported with a type 2 error rate of 0.272 which I assume is 27.2%. In the Results section the expected failures are listed as 0.854 and 0.120, should these be 85.5% and 12.0%? Figure 14 y-axes are not percentages and must be updated. Figure 8 both axes need to be updated. Figure 9 (left, EER) also likely needs to be updated.

*Thank you for pointing this out. We now have changed all reported numbers regarding errors to percentages, including all figures.*

1. Original comment: p8 “This severe limitation in the amount of available data poses the main challenge in”. It’s unclear what ‘this’ refers to. Do the authors mean the number of bullets and LEAs in the dataset or do they mean the number of sample points per striation profile (due to the fact that LEAs are shorter than screwdriver profiles)? Both would seem to be limitations but just because the profile is shorter doesn’t inherently mean that there are fewer or less informative striaes. You could have a shorter profile with more informative features. The authors replied: “We were indeed referring to the shorter length of the marks, which given the same amount of information decreases the power of the test, a purely technical consequence. Obviously you are right that this power calculation changes with a change in the amount of information.

No change was apparently made to the manuscript. I feel an edit is required. The text is still ambiguous and the reader will not have the benefit of the author’s response letter. In addition I believe the response is incorrect. We do not have enough information to make this statement. As I mentioned, it’s possible to have a very informative short profile. Perhaps bullet lands are inherently more structured than longer screwdriver marks. For example, a small but focused photo a person is more informative than a larger blurry image.

*Thank you for the comment. We have softened the sentence at the end of the paragraph. We also did not mean to make this sentence an absolute statement but were rather speculating ourselves. We have made the change to the manuscript to reflect that better.*

1. I originally wrote: The authors state “this kind of assessment is only feasible in the setting of a large study, such as the one we presented”. I strongly disagree with the use of the word “large”. The presented work is not a large study. Ten firearms were used in each of two studies (it’s not clear if both datasets used the same 10 firearms). A large study would include at least 100 different firearms and even 100 is not likely to provide statistical confidence. This is a very important point. If the paper is allowed to state that the described work is a large study it will set a precedent that will be detrimental to the field.” The authors answered: We agree the data available to us is by no means large. Our comment was referring to the availability of 35 bullets rater than just 2 0r 3 as might be seen in actual case work. We have changed the sentence accordingly.

The updated text does not appear to be changed here. The text still states: “Unfortunately, this kind of assessment is only feasible in the setting of a large study, such as the one we presented.” This needs to be changed prior to acceptance. This is not a large study. Please see my original comment.

*We are very sorry - we are completely in agreement. The original sentence was left in because of an unfortunate technical hiccup on our side. It is removed now.*