UvA-DARE (Digital Academic Repository)

Are We Breaking the Social Contract?

Colavizza, G.

DOI
[10.22148/001c.11828](10.22148/001c.11828)

Publication date
2019

Document Version
Final published version

Published in
Journal of cultural analytics

Link to publication

Are We Breaking the Social Contract?

Giovanni Colavizza[a]

[a]*University of Amsterdam, Netherlands*

ABSTRACT

The ambition of scholarship in the humanities is to systematically *understand the human condition* in all its aspects and times. To this end, humanists are more apt to *interpret* specific phenomena than *generalize* to previously unseen observations. When we consider scholarship as a collective effort, this has consequences. I argue that most of *the humanities rely on a distinct social contract*. This contract states that interpretive arguments are expected to be plausible and the grounds on which they are made, verifiable. This is the scholarly purpose (albeit not the rhetorical one) of most of what goes in our footnotes, especially references. Reference verification is mostly a virtual act, i.e., it all too rarely happens in practice, yet it is in principle always possible. Any individual scholar in any domain in the humanities can, by virtue of this contract, verify the evidence supporting any argument in a non-mediated way. This is essential to, at the very least, distinguish between solid and haphazard arguments.

The ambition of scholarship in the humanities is to systematically *understand the human condition* in all its aspects and times.[1] To this end, humanists are more apt to *interpret* specific phenomena than *generalize* to previously unseen observations. When we consider scholarship as a collective effort, this has consequences. I argue that most of *the humanities rely on a distinct social contract*. This contract states that interpretive arguments are expected to be plausible and the grounds on which they are made, verifiable. This is the scholarly purpose (albeit not the rhetorical one) of most of what goes in our footnotes, especially references. Reference verification is mostly a virtual act, i.e., it all too rarely happens in practice, yet it is in principle always possible. Any individual scholar in any domain in the humanities can, by virtue of this contract, verify the evidence supporting any argument in a non-mediated way. This is essential to, at the very least, distinguish between solid and haphazard arguments.

When using computational methods, we run the risk of eliminating this virtual affordance. We effectively tell our peers: in order to verify, and thus assess the grounds on which an argument is made, they need to be able to replicate results and

understand the computational methods they result from. This is still rarely the case in the humanities. I suggest that reactions to computational scholarship in the humanities might be in part due to the breaking of this social contract on verifiability. This is due in part to the challenge of making computational results reproducible, in part to training, and it relates more broadly to a scholarly culture which is not used to division of labor and technical specialization, as much as it practices topical specialization. The consequence is that a scholar without computational training relates with arguments based in part or whole on computational methods as a black-box, on which any scholarly value judgement is not possible. The ensuing perceived inscrutability is not due, as some suggest, to some intrinsic property of these methods, but rests in the eye of the beholder. Reactions to computational scholarship should thus be taken as a symptom of a fundamental tension within our research culture, one that ultimately might lead to intellectual fragmentation instead of mutual enrichment. I conclude by suggesting that a new social contract on verifiability should include: striving for ever higher transparency and reproducibility standards; systematically making computational results falsifiable, by testing their robustness and replicability; explicitly decoupling, for reviewing purposes, the assessment of computational results from those of interpretive work, to allow for broader engagement.

Introduction

The current academic distinction between the sciences and the humanities is but an imperfect instantiation of a long-lasting interest for the general and for the particular, present in both at varying degrees.[2] During the 19th century, Wilhelm Dilthey advanced a distinction between the natural and human sciences grounded in two distinct approaches to knowledge: a focus on explanation and cause-effect relations for the former, on understanding and part-whole relations for the latter. Further to that, Wilhelm Windelband distinguished between nomothetic and idiographic, or a tendency to generalize explanations into laws and a tendency to specify interpretive insights on the particular, respectively.[3] While post-positivist thinking during the 20th century has substantially blurred this clear-cut picture, it is still the case that scholars active in the disciplines traditionally part of the humanities remain mostly focused on interpreting and understanding the specific. Hence, the use of quantitative and computational methods in the humanities remains in the minority,

and is additionally seen with some degree of suspicion, for example when considered at risk of summoning the spirits of naive positivistic determinism or, even more subtly, obliterating the particular in exchange for at times questionable epistemic benefits.[4] The focus on interpreting the particular or on generalizing to the unseen have also led to the development of distinct, mostly unstated traditions on how scholarly communities operate.

I provide here a perspective on how these traditions play a role in motivating, in part, reactions to computational work in the humanities such as those from Nan Da and Stanley Fish.[5] Nan Da attempts a replication study in order to argue against "computational literary studies", yet ultimately both authors take an ideological stance against computational methods as part of humanistic inquiring. They claim computational methods are "atheoretical approaches" that cannot "work magic for you in producing interpretations that are *intentional*". On the one hand, these claims are trivially true: the application of computational methods ought to be informed by a theory and their argumentative use by intention. No scholar using them would argue against this. On the other hand, they are worryingly wrong when they assume there is something intrinsic in those methods forbidding their intentional and theoretically-informed use. As if a photographer, by virtue (or shall we say vice?) of using a machine, could no longer act intentionally and uncover novel aspects of reality.[6] I note that my goal is not to offer a direct reply, which has already been given,[7] and that my background and perspectives come from history rather than literary studies. The reader will, I hope, be indulgent and do the necessary 'porting' of my arguments to other areas of the humanities, as they see fit.

The humanities' social contract on verifiability

Humanities scholars in general do not have inductive prediction tasks to use in order to compare their theories and, over time, gradually improve the state of the art. Nor do they follow the deductive logic common to mathematical sciences. Conversely, scholarship in the humanities comes in the form of arguments mostly stemming from abductive reasoning.[8] An argument, in this setting, is a point of view grounded in evidence, that is to say an inference to the best explanation possible to be made plausibly with respect to the available primary sources and secondary literature. It should be clear, from the text of a contribution, how an explanation 'fits' the

evidence. To be sure, a lot of work goes into interpolating the missing bits and pieces in the evidence, as these indeed leave room for argumentative maneuvering into many open questions of research. Nevertheless, to the extent of its availability, evidence can in principle be verified since the main sources used to support a given argument are referenced and discussed in footnotes:

> *[F]ootnotes [..] are the humanist's rough equivalent of the scientist's report on data: they offer the empirical support for stories told and arguments presented. Without them, [..] theses can be admired or resented, but they cannot be verified or disproved.*[9]

The verifiability of evidence is a crucial part of the social contract specific to the humanities; customarily, verifiability is *implied*, or we could also say virtual, *immediate* and *unmediated*. Virtual verifiability refers to the fact that verification rarely happens in practice: rarely do fellow scholars have the need or time to check every source in a piece of work. Sometimes, a scholar's reputation vouches for their results, making verification seem all the more unnecessary. Nevertheless, verification is in principle possible, and footnotes provide the necessary affordances.[10] Immediate and unmediated verifiability refers instead to the practice of assuming that verification can in principle be performed directly and autonomously by any scholar: "just read the texts"[11]. This is not to say that division of labor and cumulative contributions cannot take place in the humanities, witness for example philological work resulting into critical editions, which are in turn used by literary scholars. Yet, at any given horizontal level of the scholarly debate, e.g., that of sufficiently knit communities of philologists or of literary scholars, verifiability is assumed to be possible in a non-mediated way essentially by reading through sources in view of the given argument. The extent to which this assumption actually holds in practice would make for an interesting research question.

Enter computational results: the breaking of the contract

'Computational results', for a lack of a better term, bring forth a different contract and do so unilaterally. Let us consider as computational result in the humanities any artefact produced from data at the aid of computational methods and used to support, in full or in part, a certain substantive argument. I therefore exclude from my

discussion important but relatively self-contained examples of computational methods used to produce, enrich or retrieve data, for the simple reason that these usually qualify as tasks which can be evaluated relying on established approaches from computer science. These results, which might go under the name of *humanities computing*,[12] can more easily even if not perfectly fit with known models of division of labor in the humanities, such as transcriptions or data entry, and even the more involved critical editions (see above). Computational results as here defined, instead, do not fit footnote' affordances very well, nor do they abide to the social contract I described above.[13]

On the one hand, computational results *require that they be verified explicitly*. Clarity in terms of the correct choice and application of computational methods is necessary during review; the consequent need to make them explicit, alongside the modelling choices they imply or otherwise necessitate, is per se a significant step forward in terms of transparency.[14] Explicit verifiability often entails reproducibility of results, which is only possible when data and code are made available.[15] The verification of computational results also requires skills that are often beyond the traditional training of scholars in the humanities, therefore necessitating *mediation*.[16] Furthermore, there also is another kind of mediation, stemming from the interest that some scholars using computational methods have, to detect and understand patterns or make predictions that *generalize*. Generalization implies robust and replicable results, that is to say comparable outcomes when performing the same analysis on different datasets (robustness) and when performing conceptually related, but methodologically different analyses on the same dataset (replicability), under the assumptions that both analyses and datasets are designed for the same phenomenon being considered.[17] It is common for a community to refine over some time if and how a result generalizes, as a collective effort, instead of attempting to work everything out in one go, as it is common with traditional interpretive work.

Maintaining the same social contract on verifiability we discussed before would thus require a high standard in terms of data sharing, reproducibility of results and understanding of computational methods within the community. This is taking place already.[18] It would also require, to maintain immediate and unmediated verifiability, all scholars part of the conversation to have the necessary skills to

understand and, potentially, reproduce computational results. Lastly, it might require computational results to either renounce their claims to generality or assess them extensively not over time as a community effort, but before publication. This seems unlikely to happen.

We are thus left with a conundrum: on the one hand, the old social contract no longer applies, while a new one requires an adaptation which is taking time and generating frictions, in part because we need to experiment and tinker with it. Most crucially, a majority of scholars in the humanities are technically unable to engage with computational results in an immediate and unmediated way. At the same time, the high standards of transparency and reproducibility of results, application of methods and social division of labor which are required from computational results in the humanities, are still being improved.[19] This historical and cultural conjuncture is, I suggest, creating tensions and counter-reactions. How, if at all, might scholars in the humanities engage with computational results when they cannot immediately verify results, no matter the reason? They can either accept them blindly or refuse them in much the same way. Neither is helpful.

Ways forward

As we work out a new social contract on verifiability which accommodates the use of computational results in humanistic scholarship, we need to provide ways in to all concerned. I propose the following three points for consideration: a) keep working towards ever higher transparency and reproducibility standards; b) confronting more systematically the need to make computational results falsifiable, which translates into testing their robustness and replicability, as previously defined; c) explicitly decoupling the review of computational results from that of interpretive work.

The first point is, perhaps, the least contentious and the most developed. It seems justified to suggest that computational results and resulting scholarship call for the highest transparency and openness standards possible, in order to allow for their reproducibility and to involve in the process as much of the community as possible. If anything, Nan Da's piece is witness to a laudable desire to fully engage with this.

Making results reproducible is also necessary to build confidence for mediated verifiability within a community.

Secondly, *we should embrace falsifiability* via robustness and replication studies. If a result (or an argument, for that matter) claims generality, then it needs to allow itself to be put to test using different yet applicable methods and datasets. This process would first of all build confidence in the result itself: being exposed to falsifiability and, eventually, surviving it is a crucial aspect of verifiability (perhaps *the* crucial one).[20] Furthermore, assessing the robustness and replicability of results as a community allows for cumulative efforts. It is the responsibility of scholars using computational methods to provide for the necessary 'falsifiability affordances', by making the theoretical assumptions, methodological choices and claimed generality of their results explicit.[21] A narrative explaining the argument, and explicit references to the relevant supporting evidence, serve a similar purpose in traditional scholarship.

Yet even if we were to assume complete availability of code and data and full falsifiability for every computational result, we would still need to find in the community a sufficient number of peers with the necessary skills to engage with them. This is a problem of training and, even more so, of choice in what to train for. Nevertheless, there are other options which could be considered in order to guarantee for (somewhat mediated) verifiability. On the one hand, striving for substantive collaborations with the computational sciences, nurturing brokering experiences which can accelerate the development of a new social contract. On the other hand, decoupling the review (and verification) of computational methods used to generate results explicitly and transparently: in essence, establishing a reviewing process by which the methodological aspects of computational results are assessed independently from their argumentative use, and both are considered by respective experts.[22]

I propose that computational methods should be reviewed according to the following criteria: 1) reproducibility, 2) soundness and parsimoniousness (is the method fit-for-purpose and applied according to state-of-the-art practice?), 3) robustness and replicability. I further suggest that the review of *the use* of computational results in humanities scholarship should be valued according to the following criteria: 1)

necessity (is the computational result a necessary part of the argument?), 2) value (is the computational result an integral part of the argument?), 3) well-situatedness (is the computational result organically integrated with all the other non-computational results which contribute to the argument?). These two reviews can, in principle, be performed independently from each other, ideally as part of a transparent process.[23] Such a process would raise the likelihood that: a) the computational result and its argumentative use are considered to be acceptable (or not) separately from each other; b) scholars without computational skills can engage in the reviewing process without having to worry about computational methods, and vice versa for methods' experts; c) computational results can have a separate life than the one of the specific argument they are tied into at publication time. This said, the reviewer for both methods and arguments can indeed be the same person, provided they have the necessary competencies.

Conclusions

I have suggested that the humanities operate by making plausible and grounded arguments. The social contract in use calls for implied, immediate and unmediated verifiability, that evidence is only rarely directly verified and, when this is done, anyone in the community can do it (usually, by reading). Computational results in the humanities, instead, are bringing forth a different social contract: verifiability is explicit and required during review and, more often than not, mediated by someone else. How then, are scholars in the humanities going to be able to engage with computational results? Longer-term objectives could include better training in computational methods and even higher standards in terms of open, reproducible and falsifiable results. Improvements we could start doing now could include more substantive collaborations with the computational sciences and explicitly decoupling the assessment of computational methods from the argumentative use of computational results in the review process.

Nan Da's article is but a recent example of strong, even vehement reaction against computational results in literary studies and, more broadly, the humanities. This malaise should be taken as a symptom of a more fundamental tension within our research culture, one that ultimately might lead to intellectual fragmentation and split

ways, instead of mutual enrichment. It thus seems all the more crucial to keep a constructive, pragmatic and non-ideological conversation going.

Notes

[1] I would like to thank the Living with Machines project team http://livingwithmachines.ac.uk, for having started a conversation that led me to write this piece. I also benefited from insightful comments by Ruth Ahnert, Irene Ceccherini, Michael Piotrowski and Andrew Piper.

[2] Bod, Rens. 2018. "Has There Ever Been a Divide? A Longue Duré Perspective." *History of Humanities* 3 (1): 15-25.

[3] Dilthey, Wilhelm. 1883. *Einleitung in die Geisteswissenschaften*. Leipzig: Duncker & Humblot and Windelband, Wilhelm. 1904. *Geschichte und Naturwissenschaft*. Strassburg: Heitz.

[4] "[Q]uantitative is not nomothetic history, but they invite each other." Weingart, Scott B. 2016. *Lessons from Digital History's Antecedents*. https://scottbot.net/lessons-from-digital-historys-antecedents.

[5] Da, Nan Z. 2019. "The Computational Case against Computational Literary Studies." *Critical Inquiry* 45 (3): 601-39 and Fish, Stanley. 2019. *Afterword*. https://critinq.wordpress.com/2019/04/03/computational-literary-studies-participant-forum-responses-day-3-5.

[6] Daston, Lorraine, and Peter Galison. 1992. "The Image of Objectivity." *Representations*, 40: 81-128 and Daston, Lorraine, and Peter Galison. 2010. *Objectivity*. New York: Zone Books.

[7] See Piper, Andrew. 2019. "Do we know what we are doing?" *Journal of Cultural Analytics* and the contributions at https://critinq.wordpress.com/2019/03/31/computational-literary-studies-a-critical-inquiry-online-forum.

[8] It could be argued that humanists, or at least historians, follow a mix of abductive and inductive reasoning, termed adduction by Fischer: "a process of adducing answers to specific questions, so that a satisfactory explanatory 'fit' is obtained. [..] History is, in short, a problem-solving discipline." Fischer, David Hackett. 1979. *Historians' Fallacies: Toward a Logic of Historical Thought*. New York: Harper & Row, p. xv.

[9] Grafton, Anthony. 1999. *The Footnote: A Curious History*. Cambridge, Mass.: Harvard University Press, p. vii.

[10] I do not discuss here the other side of the footnote, so to say: their rhetorical, even political, use for persuasion or other needs. This is important, but beyond my point. See Grafton, *The Footnote*, p. 22 and Bensman, Joseph. 1988. "The Aesthetics and Politics of Footnoting." *International Journal of Politics, Culture, and Society* 1 (3): 443-470.

[11] Da, "The Computational Case", p. 639.

[12] McCarty, Willard. 2003. "Humanities Computing." *Encyclopedia of Library and Information Science*, 1224-1236.

[13] Consider Judea Pearl's *Ladder of Causation*. The humanities mostly position at the first level, that of observations and empirical associations, and use computational methods there. In so doing, humanists aim to

develop an understanding of human phenomena at the third, counterfactual level. Humanists usually do not have access to the second level, that of interventions and controlled experiments, hence they set themselves to quite a challenging task. The lack of access to the second level prevents most computational results in the humanities to support anything more than observational claims. Pearl, Judea, and Dana Mackenzie. 2018. *The Book of Why*. London: Allen Lane.

[14] See Piotrowski, Michael. 2018. "Digital Humanities: An Explication." In M. Burghardt & C. Müller-Birn (eds.): *INF-DH 2018 - Workshopband*, 25 Sept. 2018, Berlin; Piotrowski, Michael. 2019. "Accepting and Modeling Uncertainty." In: A. Kuczera, T. Wübbena & T. Kollatz (eds.): *Die Modellierung des Zweifels - Schlüsselideen und konzepte zur graphbasierten Modellierung von Unsicherheiten*, Wolfenbüttel, http://www.zfdg.de/sb004_006; Piper, "Do we know what we are doing?".

[15] This is not news to history and even more so to literary studies. "Some of the new forms of history rest on evidence that footnotes cannot accommodate—like the massive analyses of statistical data undertaken by historical demographers, which can be verified only when they agree to let colleagues use their computer files." Grafton, *The Footnote*, p. 15.

[16] The very article from Da was specifically reviewed by an "an expert in quantitative analysis and mathematical modeling". Da, Nan Z. 2019. *Final Comments*. https://critinq.wordpress.com/2019/04/03/computational-literary-studies-participant-forum-responses-day-3-4. Unfortunately, we are left wondering how computationally savvy and not computationally savvy reviewers divided their work, given that Nan Da's use of computational results is so profoundly embedded into her contribution. The mediating role of the editor must have been crucial.

[17] For definitions see The Turing Way Community. 2019, April 26, commit 797eb8ea86965de63a5be3de5227ff6e53c65f60. *The Turing Way*. *Definitions of reproducibility*: https://the-turing-way.netlify.com/reproducibility/03/definitions.

[18] See, for example, the guidelines of the Journal of Cultural Analytics: http://culturalanalytics.org/about/about-ca.

[19] Curiously enough, division of labor is not very practiced in humanities disciplines most acquainted with the model of the lone scholar. While topical specialization is seen as understandable and even desirable, methodological specialization less so. Even technical specialization, especially so in view of enabling cumulative efforts, is considered as not desirable when, as it often happens, it borders on perceived substantive matters of interpretation.

[20] See also Peels, Rik, and Lex Bouter. 2018. "The Possibility and Desirability of Replication in the Humanities." *Palgrave Communications* 4 (1): 1-4 and Ted Underwood's blogpost: https://tedunderwood.com/tag/reproducible-research.

[21] More principled approaches to falsifiability can also be considered, for example by taking inspiration from the field of machine learning. Therein, it is common practice to leave part of the dataset on a side as an ultimate *test set*, to be used to assess how a certain result or method generalizes to the unseen. See, e.g., Hal Daumé III, *A Course in Machine Learning* http://ciml.info. In some areas of the humanities where computational methods are more common, this practice is profitably being used already.

[22] In this sense, Nan Da's proposed guidelines, while necessitating further refinement and experimentation, are a step in the right direction. See Da, "The Computational Case", Appendix, Section 9: "Suggested Guidelines for Reviewing CLS Manuscripts".

[23] See, e.g., https://openreview.net.