Uncertainty as an Unavoidable Good

Michael Piotrowski
University of Lausanne, Lausanne, Switzerland
ORCID: http://orcid.org/0000-0003-3307-5386

ABSTRACT

In digital history, uncertainty is generally regarded as an unavoidable evil. One generally aims to reduce—and ideally resolve—uncertainty in data as much as possible. However, information systems are not designed to handle the absence of information; we discuss how both SQL's seemingly simple Null marker and the TEI Guideline's elaborate facilities for recording "certainty" fail to address the challenges posed by uncertainty. Neither is big data and a "digital historical positivism" a satisfactory answer: the causal models that underpin historical narratives do not simply emerge from a collection of facts. Here, it is necessary to distinguish between two types of uncertainty: *historical uncertainty*, which concerns the facts of the past, and *historiographical uncertainty*, which concerns the causal models constructed by historians. The latter results from different interpretations of the causal relations between the facts; given our limited knowledge of the past, it is ultimately irreducible. But it is also this uncertainty that allows us to construct the narratives we need for sense-making. We argue that in this sense uncertainty may be regarded as an unavoidable good and that we should aim to design computational frameworks that treat it as an asset rather than an obstacle.

Keywords: uncertainty, historiography, causal models, epistemology

1 Introduction

As Bertrand Russell famously remarked, "All human knowledge is uncertain, inexact, and partial."[1] For obvious reasons, historical research is particularly confronted with uncertainty: our knowledge of the past is always limited. It is overwhelmingly not based on any kind of first-hand experience but it can only be gained indirectly. The human actions studied by historians are not governed by the laws of nature: while one can certainly identify patterns of human behavior through the ages, human actions are not predictable. In contrast to uncertainty in everyday life, much, if not all, of the original context is missing. Nor can we usually ask historical actors for clarification, and even when we can, their views are highly subjective. What is more, not only is our knowledge of the past always fragmentary, but as Manfred Thaller[2] has pointed out, history is particularly

[1] Bertrand Russell, Human knowledge: Its scope and limits, London 1948, p. 527.

[2] Decoding what the sender did not want to transmit. Information technology and historical data; or something, in: Michael Piotrowski (ed.), Proceedings of the Workshop on Computational Methods in the Humanities (COMHUM 2018), 2018.

concerned with "decoding what the sender did not want to transmit," i.e., the most relevant sources used by historians were not intended as messages to them.

Historians are generally aware of this. When studying sources, they take into account that things may not mean (or have meant) what they seem to mean at first glance, that there are very likely inaccuracies, errors, fabrications, and forgeries. Even when they are convinced they have correctly identified a person, a place, or the time of an event, they know that there is never absolute certainty, and the narratives created as results of historical research tend to reflect this, at least when and where authors consider the uncertainty relevant for the understanding.

2 Uncertainty in Digital History

Uncertainty is usually discussed in negative terms, as something brought about by the *absence* of information, a type of *ignorance*[3] or *imperfection*,[4] which is often discussed together with related concepts like *error*, *ambiguity*, or *vagueness*. This often implies that the missing information exists or has at some point existed somewhere, and that the uncertainty is at least potentially resolvable. In general, however, uncertainty is unavoidable for the same reasons that make Laplace's demon an impossibility; in particular, the storage capacity of both living and technical systems is finite, so no system can ever hold *all* information.[5]

Since information systems—and this applies not only to today's digital information systems, but also to previous technologies such as card files—are designed for storing, retrieving, and processing *information*, absent information is largely implicit as "everything that is *not* in the system." As an illustrative example of some of the problems that arise when trying to represent missing information, take the notion of *Null*, as implemented by relational database management systems based on SQL. Null can be considered information about missing information and may thus a representation or at least an indication of uncertainty. The concept of Null was not part of Codd's original relational model;[6] Codd introduced in 1979, remarking that the "two most important types of null value have the meanings 'value at present unknown' and 'property inapplicable.'"[7]

This already goes to show that the notion of Null is not as straightforward as it may seem: in the prototypical case of a personnel database, actually missing information is an anomaly that must eventually be corrected; inapplicable properties, however, are perfectly fine. Thus, Null is on the one hand very unspecific—there is only a single Null and it is not possible to distinguish between different situations—on the other hand, Null is very specific with regard to the locus of uncertainty: it indicates the absence of a particular piece of information in a particular field of a particular record. The introduction of Null changes the logical basis of the relational model from two-valued to three-valued logic (with the truth values *true*, *false*, and *unknown*); effectively, three-valued logic is used as a mechanism for handling missing information.[8] However, opinions differ whether

[3] Michael Smithson, Ignorance and uncertainty, New York, NY, USA 1989, p. 9.

[4] Philippe Smets, Imperfect information: Imprecision and uncertainty, in: Amihai Motro/Philippe Smets (eds.), Uncertainty management in information systems, Boston, MA, USA 1997, pp. 225–254.

[5] Simon Parsons, Qualitative approaches for reasoning under uncertainty, Cambridge, MA, USA 2001, pp. 7–8.

[6] E. F. Codd, A relational model of data for large shared data banks, in, Communications of the ACM 13 (1970) 6, pp. 377–387.

[7] E. F. Codd, Extending the database relational model to capture more meaning, in, ACM Transactions on Database Systems 4 (1979) 4, p. 403.

[8] David McGoveran, Nothing from nothing: Part I: What's logic got to do with it?, in, Database Programming & Design 6 (1993) 12, p. 33.

this approach actually solves the problem of handling missing information in an information system. Whereas Codd advocated the use of many-valued logic as a basis for dealing with missing information, others criticize this approach as inherently flawed; one of the most prominent critics is probably C. J. Date, who has called it "a mistake" that leads to "wrong answers."[9] An important theoretical observation by Date is that Null violates the (implicit) closed-world assumption underlying relational databases, effectively replacing it by an open-world assumption.

This example shows that the representation of missing information—uncertainty—in an information system is not easy and has potentially far-reaching consequences. Under certain circumstances, the consequences can be said to be limited, notably when no (non-trivial) computation takes place. If, for example, a database of historical persons contains "exact" dates of birth, even though they may be uncertain, this is relatively unproblematic if it serves merely as an *aide-mémoire* to users who are aware of this issue. It does become problematic, though, when this data is to be processed automatically.

There is increasing awareness of this problem in digital humanities. So far, the focus has been primarily on the identification of the referents of linguistic expressions—especially named entities, dates, and locations—or things like manuscript readings. The objective is usually to *reduce* uncertainty as much as possible in order to obtain unambiguous data—such as geographical coordinates—that's amenable to computational processing. In other words, if, for example, coordinates are known to be "good," one can also be sure that identical coordinates do in fact refer to the same place.

As this is often not possible, there has also been work on making uncertainty explicit and machine-readable, so that it can be associated with the data—which requires *modeling* uncertainty—[10]or on making the uncertainty evident in the results of computational processing, e.g., in visualizations.[11] Metrology (the scientific study of measurement) has developed sophisticated approaches for working with measurement uncertainty and to avoid or at least control the propagation of errors, but since history is studying phenomena of a different kind, these methods cannot be transferred, and we lack rigorous methods for dealing with uncertainty in digital history and digital humanities in general.

The TEI Guideline's chapter on "Certainty, Precision, and Responsibility"[12] tries hard to define a formal framework for adding information about "certainty" to annotations; but the notion of "certainty" remains unclear, and it is variously described as "probability" and as "degree of confidence." One of the examples concerns uncertainty about the interpretation of an occurrence of "Essex" in some text as either referring to a place or a person; the Guidelines explain: "We may wish to record the probability, assessed in some subjective way, that 'Essex' really is a place name here. The `@degree` attribute is used to indicate the degree of confidence associated with the certainty element, expressed as a number between 0 and 1." The attribute `degree="0.6"` is then supposed to express

[9] C. J. Date, Why three-valued logic is a mistake, in: C. J. Date (ed.), Relational database writings 1991–1994, Reading, MA 1995 [1992], pp. 22–29; for a discussion between Codd and Date, see E. F. Codd/C. J. Date, Much ado about nothing, in: C. J. Date (ed.), Relational database writings 1991–1994, Reading, MA 1995 [1993], pp. 341–362.

[10] See, e.g., Michael Piotrowski, Accepting and Modeling Uncertainty, in, Zeitschrift für digitale Geisteswissenschaften (2019); Jennifer Edmond, Strategies and recommendations for the management of uncertainty in research tools and environments for digital history, in, Informatics 6 (2019) 3.

[11] See, e.g., Florian Kräutli/Stephen Boyd Davis, Known unknowns: Representing uncertainty in historical time, in, Proceedings of electronic visualisation and the arts 2013, London 2013, pp. 61–68; Roberto Therón/Antonio G. Losada/Alejandro Benito et al., Toward supporting decision-making under uncertainty in digital humanities with progressive visualization, in, Proceedings of the sixth international conference on technological ecosystems for enhancing multiculturality (TEEM'18), New York, NY, USA 2018, pp. 826–832.

[12] https://tei-c.org/release/doc/tei-p5-doc/en/html/CE.html

"the point of view that there is a 60 percent chance of 'Essex' being a place name here, and hence a 40 percent chance of its being a personal name." But what does this actually mean, given that this is not a bet or prediction that could be tested against the actual outcome? Indeed, while the calculus of probabilities is mathematically well understood, "there are sharp disagreements on the meaning of 'probability'."[13] It seems that the authors of the Guidelines recognize, on the one hand, that what we are dealing with here is a subjective degree of belief; on the other hand, many descriptions suggest an objective understanding of probability. Moreover, it remains an open question whether the uncertainty encountered here is in fact usefully described in terms of probability—or whether, say, *plausibility*[14] might be a more appropriate framework.

In practice, these TEI constructs are exceedingly rare. A much more popular approach is what might be called "digital historical positivism," which aims to overcome uncertainty with big data. A prime example is the Venice Time Machine project,[15] but this idea implicitly underlies much of DH research. The tenet of historical positivism formulated by Langlois and Seignobos in 1898 applies to both its 19th and 21st century forms:[16]

> On peut penser qu'un jour viendra où, grâce à l'organisation du travail, tous les documents auront été découverts, purifiés et mis en ordre, et tous les faits dont la trace n'a pas été effacée, établis. — Ce jour-là l'histoire sera constituée [...].[17]

However, this is a serious misunderstanding of what history is: reconstructing the past is not the same as writing history, even if the reconstruction were to be "complete." Wink's analogy of "the historian as detective"[18] is very fitting: historians are confronted with a "crime scene," the current or a past state of the world. Their task is to build a "case," i.e., a narrative that explains how, through which events, this state of the world came to be. It is only in the narrative that selected traces become evidence.

Hull calls historical narratives "descriptions of historical entities as they persist through time";[19] even more generally one could say that a narrative puts entities into relation with each other. Entities can be persons, places, objects, but also ideas or events. While there are many possible types of relations between entities, Hull points out that the "main sort of continuity and unity envisaged so far by philosophers for historical narratives has been causal."[20] In other words, we might say that every historical narrative is underpinned by a *causal model*, which is traditionally not made explicit and exists only in the minds of historians.

---

[13] Paul E. Lehner/Kathryn B. Laskey/Didier Dubois, An introduction to issues in higher order uncertainty, in, IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 26 (1996) 3, p. 289.

[14] Nicholas Rescher, Plausible reasoning: An introduction to the theory and practice of plausibilistic inference, Assen 1976.

[15] Frédéric Kaplan/Isabella di Lenardo, Building a mirror world for Venice, in, The aura in the age of digital materiality: Rethinking preservation in the shadow of an uncertain future, Milan 2020, pp. 197–201.

[16] Author's translation: "It is conceivable that the day will come when, thanks to work organization, all documents will have been discovered, purified, and put in order, and all the facts whose traces have not been erased, established.—On that day, history will be settled [...]."

[17] Charles-Victor Langlois/Charles Seignobos, Introduction aux études historiques, Paris 1898, p. 277.

[18] Robin W. Winks (ed.), The historian as detective: Essays on evidence, New York, NY 1969.

[19] David L. Hull, Central subjects and historical narratives, in, History and Theory 14 (1975) 3, p. 254.

[20] Ibid.

3 Two Types of Uncertainty

At this point we need to make a distinction between digital and computational history. Like Mullen, I consider computational history as "a particular kind of work one can do within the much larger domains of digital history."[21] In line with my definitions of digital and computational humanities,[22] computational history is concerned with the construction of computational models to study historical research questions. Consequently, the goal is neither to digitize sources, nor to digitally reconstruct the past, but rather to formalize historians' causal models that underpin historical narratives, or, as Jean-Claude Gardin[23] put it, *la formalisation du discours savant*, the formalization of the scholarly discourse. It is in this specific context in which we need to consider uncertainty.

Hull further notes that, "[a]ssuming for the moment that history could be analyzed completely into a single set of atomistic elements, there are indefinitely many ways in which these elements can be organized into historical sequences."[24] It would be ludicrous to do so, but could be taken as a rough model of historiography: in crafting the narrative, the historian—like the detective—works with a limited number of elements and determines their relation to each other. Uncertainty can thus concern the elements themselves, the relations between the entities, or both. We therefore need to distinguish two types of uncertainty, which I have come to call *historical uncertainty*, which concerns the facts of past, and *historiographical uncertainty*, which concerns the construction of causal models that link these facts together, and which depends heavily on the interpretation of the facts.

This distinction is not specific to history. For example, in the context of decision theory, these two types of uncertainty are referred to as *parameter uncertainty*, i.e., uncertainty about the parameters of a model, and *model uncertainty*, i.e., uncertainty about the structure of a model, "the situation in which there is no single, agreed upon model for a problem."[25] As in historical research, "it may be necessary to acknowledge that in the presence of irreducible model uncertainty there may be no single 'right answer' and reasonable people may disagree."[26]

Given the unavoidable historical uncertainty and the lack of "laws" of human behavior, more often than not, we are confronted with what Paul Ricœur called the *conflit des interprétations*;[27] historiographical uncertainty is thus often irreducible. Historiographical uncertainty can therefore also be described as "multi-interpretation."[28]

For a long time, digital history has been primarily concerned with the digitization of sources and the derivation of data from sources.[29] For practical reasons, the focus has thus been primarily on historical uncertainty, typically due to missing, inexact, partial,

[21] Lincoln A. Mullen, Computational historical thinking: With applications in R, 2018–, sec. 1.2.

[22] Michael Piotrowski, Digital humanities: An explication, in: Manuel Burghardt/Claudia Müller-Birn (eds.), Proceedings of INF-DH 2018, Berlin 2018; Michael Piotrowski, Ain't no way around it: Why we need to be clear about what we mean by "digital humanities," in: Martin Huber/Sybille Krämer/Claus Pias (eds.), Wozu digitale geisteswissenschaften? Innovationen, revisionen, binnenkonflikte, n.d.

[23] Le calcul et la raison: essais sur la formalisation du discours savant, Paris 1991.

[24] Hull, Central subjects and historical narratives, p. 255.

[25] Kathryn B. Laskey, Model uncertainty: Theory and practical implications, in, IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 26 (1996) 3, p. 340.

[26] Ibid., p. 347.

[27] Paul Ricœur, Le conflit des interprétations: Essais d'herméneutique, Paris 2017.

[28] Jean-Claude Gardin, L'interprétation dans les humanités: réflexions sur la troisième voie, in: Richard Ennals/Jean-Claude Gardin (eds.), Interpretation in the Humanities: Perspectives from Artificial Intelligence, London 1990 (Library and Information Research Report, 71), pp. 22–59.

[29] Jesse W. Torgerson, Historical practice in the era of digital history, in, History and Theory 61 (2022) 4, pp. 37–63.

ambiguous, etc., information. So far, there has been little work on formalizing historiographical uncertainty or multi-interpretation.

4 Uncertainty as an Unavoidable Good

I have previously said that uncertainty must be accepted and modeled.[30] However, in the light of the above, I would now go even further and argue that if we want to construct computational historiographical models, the way forward is not to just accept uncertainty in as an unavoidable evil, caused by the unfortunate fact that our knowledge of the past is incomplete. While it does play a role, historiographical uncertainty is *not* merely a side effect of historical uncertainty. Uncertainty resulting from different interpretations of the facts is a defining characteristic of the humanities; designating them as "the uncertain sciences"[31] or "les sciences de l'imprécis"[32] is justified. This is not due to sloppiness, but due to the inevitability of uncertainty, which, when it concerns human actions, cannot be reduced by falling back on universal laws, i.e., ready-made, proven models.

At the same time, historiography responds to the innate human need for causal explanations. In this sense we note that, yes, uncertainty is unavoidable, but at the same time it is this uncertainty that *enables* the writing of history, and we should treat it correspondingly. If we were Laplace's demon and knew everything, we still would not understand. In the same way, digitizing all documents held by, say, the state archives of Venice, the cadaster and all the tax records is certainly useful, but the causal links between the facts do *not* simply emerge from the facts. This was already a fallacy of the 19$^{\text{th}}$-century historical positivists; for example, the French historian Fustel de Coulanges affirmed, "It is not I who speak, but history which speaks through me."[33]

The narrative is *always* constructed, and more sources or a better knowledge do not necessarily imply a better approximation of the "truth." Thus, what Neil Gershenfeld calls the "most common misunderstanding about science"[34] is also the most common misunderstanding about history: neither scientists nor historians seek and find truth—they make and test models.

To repeat, uncertainty is unavoidable. But that is precisely why we should not consider it as "bad." Trying to model uncertainty to make it manageable is ultimately also a dead end. If historiography is only possible under uncertainty, we should perhaps think of uncertainty not as an obstacle but rather as a lever. The challenge is to come up with computational frameworks that allow us to do this in a rigorous fashion.

Acknowledgements

This paper is based upon work supported by the Swiss National Science Foundation (SNSF) under grant nos. 190306 and 204305.

[30] Piotrowski, Accepting and Modeling Uncertainty.

[31] Bruce Mazlish, The uncertain sciences. With a new introduction by the author, London; New York 2017 [1998].

[32] Abraham A. Moles, Les sciences de l'imprécis, Paris 1995 [1990].

[33] "Ne m'applaudissez pas, ce n'est pas moi qui vous parle ; c'est l'histoire qui parle par ma bouche." (as quoted in Gabriel Monod, M. Fustel de Coulanges, in, Revue historique 42 (1889) 2, p. 278)

[34] Neil Gershenfeld, Truth Is a Model, response to the thematic issue "What Scientific Concept Would Improve Everybody's Toolkit?" *Edge* (2011), www.edge.org.

References

Codd, E. F., Extending the database relational model to capture more meaning, in: ACM Transactions on Database Systems 4 (1979) 4, pp. 397–434.

Codd, E. F., A relational model of data for large shared data banks, in: Communications of the ACM 13 (1970) 6, pp. 377–387.

Codd, E. F./Date, C. J., Much ado about nothing, in: C. J. Date (ed.): Relational database writings 1991–1994, Reading, MA 1995 [1993], pp. 341–362.

Date, C. J., Why three-valued logic is a mistake, in: C. J. Date (ed.): Relational database writings 1991–1994, Reading, MA 1995 [1992], pp. 22–29.

Edmond, Jennifer, Strategies and recommendations for the management of uncertainty in research tools and environments for digital history, in: Informatics 6 (2019) 3.

Gardin, Jean-Claude, Le calcul et la raison: essais sur la formalisation du discours savant, Paris 1991.

Gardin, Jean-Claude, L'interprétation dans les humanités: réflexions sur la troisième voie, in: Richard Ennals/Jean-Claude Gardin (eds.): Interpretation in the Humanities: Perspectives from Artificial Intelligence, London 1990 (Library and Information Research Report, 71), pp. 22–59.

Hull, David L., Central subjects and historical narratives, in: History and Theory 14 (1975) 3, pp. 253–274.

Kaplan, Frédéric/Lenardo, Isabella di, Building a mirror world for Venice, in: The aura in the age of digital materiality: Rethinking preservation in the shadow of an uncertain future, Milan 2020, pp. 197–201.

Kräutli, Florian/Boyd Davis, Stephen, Known unknowns: Representing uncertainty in historical time, in: Proceedings of electronic visualisation and the arts 2013, London 2013, pp. 61–68.

Langlois, Charles-Victor/Seignobos, Charles, Introduction aux études historiques, Paris 1898.

Laskey, Kathryn B., Model uncertainty: Theory and practical implications, in: IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 26 (1996) 3, pp. 340–348.

Lehner, Paul E./Laskey, Kathryn B./Dubois, Didier, An introduction to issues in higher order uncertainty, in: IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 26 (1996) 3, pp. 289–293.

Mazlish, Bruce, The uncertain sciences. With a new introduction by the author, London; New York 2017 [1998].

McGoveran, David, Nothing from nothing: Part I: What's logic got to do with it?, in: Database Programming & Design 6 (1993) 12, pp. 33–41.

Moles, Abraham A., Les sciences de l'imprécis, Paris 1995 [1990].

Monod, Gabriel, M. Fustel de Coulanges, in: Revue historique 42 (1889) 2, pp. 277–285.

Mullen, Lincoln A., Computational historical thinking: With applications in R, 2018–.

Parsons, Simon, Qualitative approaches for reasoning under uncertainty, Cambridge, MA, USA 2001.

Piotrowski, Michael, Accepting and Modeling Uncertainty, in: Zeitschrift für digitale Geisteswissenschaften (2019).

Piotrowski, Michael, Digital humanities: An explication, in: Manuel Burghardt/Claudia Müller-Birn (eds.): Proceedings of INF-DH 2018, Berlin 2018.

Piotrowski, Michael, Ain't no way around it: Why we need to be clear about what we mean by "digital humanities", in: Martin Huber/Sybille Krämer/Claus Pias (eds.): Wozu digitale geisteswissenschaften? Innovationen, revisionen, binnenkonflikte, n.d.

Rescher, Nicholas, Plausible reasoning: An introduction to the theory and practice of plausibilistic inference, Assen 1976.

Ricœur, Paul, Le conflit des interprétations: Essais d'herméneutique, Paris 2017.

Russell, Bertrand, Human knowledge: Its scope and limits, London 1948.

Smets, Philippe, Imperfect information: Imprecision and uncertainty, in: Amihai Motro/Philippe Smets (eds.): Uncertainty management in information systems, Boston, MA, USA 1997, pp. 225–254.

Smithson, Michael, Ignorance and uncertainty, New York, NY, USA 1989.

Thaller, Manfred, Decoding what the sender did not want to transmit. Information technology and historical data; or something, in: Michael Piotrowski (ed.): Proceedings of the Workshop on Computational Methods in the Humanities (COMHUM 2018), 2018.

Therón, Roberto/Losada, Antonio G./Benito, Alejandro et al., Toward supporting decision-making under uncertainty in digital humanities with progressive visualization, in: Proceedings of the sixth international conference on technological ecosystems for enhancing multiculturality (TEEM'18), New York, NY, USA 2018, pp. 826–832.

Torgerson, Jesse W., Historical practice in the era of digital history, in: History and Theory 61 (2022) 4, pp. 37–63.

Winks, Robin W. (Ed.), The historian as detective: Essays on evidence, New York, NY 1969.