MODELING CONFLICTING, UNRELIABLE, AND VARYING INFORMATION

LARS RÖNNBÄCK

20 JUNE 2018

 Most persistent memories in which bodies of information are stored can only provide a view of that information as it currently is, from a single point of view, and with no respect to its reliability. This is a poor reflection of reality, because information changes over time, may have many and possibly disagreeing origins, and is far from often certain. Hereat, this paper introduces a modeling technique that manages conflicting, unreliable, and varying information. In order to do so, the concept of a "single version of the truth" must be abandoned and replaced by an equivocal theory that respects the genuine nature of information. Through such, information can be seen from different and concurrent perspectives, where each statement has been given a reliability ranging from being certain of its truth to being certain of its opposite, and when that reliability or the information itself varies over time, changes are managed non-destructively, making it possible to retrieve everything as it was at any given point in time. As a result, other techniques are, among them third normal form, anchor modeling, and data vault, contained as special cases of the henceforth entitled *transitional modeling*[1].

KEYWORDS

transitional, modeling, information, temporality, concurrency, reliability, variability

Introduction

Most current information modeling techniques result in lossy implementations, in the sense that they cannot preserve combinations of who said what, when, and how sure they were of what they were saying. Modeling such requirements explicitly using traditional techniques is complex and error-prone, and therefore rarely done in practice, resulting in lost information impossible to recover Golfarelli and Rizzi, "A survey on temporal data warehousing"[2]. In fact, every statement has an origin and is made with some degree of confidence Liu, *Uncertainty Theory: An Introduction to its Axiomatic Foundations. 2004*[3], but in most modeling techniques the simplified assumption is that statements are universal, univocal, and unvarying. Riddance of such limitations should be of interest to, for example, institutions subject to financial regulations that stipulate complete auditability, uncertain and complementary clinical results within the health care domain, and conflicting information that may be strengthened or discarded over time within military, policiary, or judicial applications Fisher and Kingma, "Criticality of data quality as ex-

emplified in two disasters"[4]. Modeled information often ends up in databases, and while database vendors have started to add rudimentary temporal capabilities Kulkarni and Michels, "Temporal features in SQL: 2011"[5], these are still insufficient Johnston, *Bitemporal data: theory and practice*[6]. This paper introduces a generally applicable modeling technique, that is lossless with respect to concurrency, reliability and temporality. It is, for example, able to capture the following illustrative story in a structured way:

> The accused, Archie, was seen fleeing the scene of the crime by two witnesses, Donna and Charlie. Donna is certain that the accused was clean shaved, while Charlie thinks he had a red beard. When the victim, Bella, later regained consciousness, she corroborated Charlie's story, at which time Donna retracted her statement. In fact, Archie had been wearing a fake red beard during the attack, but it fell off when he fled, eventually leading to his conviction through a DNA match.

In the terminology introduced, *concurrency* refers to having multiple, possibly conflicting, views of the same state of affairs, *reliability* refers to being able to express a degree of uncertainty about said state of affairs, and *temporality* refers to when it took place, when it changes, when an opinion is had towards it, and when such opinions were recorded. There are approaches that address these aspects of information in isolation or in combination Benjelloun et al., "Databases with uncertainty and lineage"; Dylla, Miliaraki, and Theobald, "A temporal-probabilistic database model for information extraction"[7], but to the best knowledge of the author no existing approach combines all of them. The paper is structured with Section 2 containing a formalization of a technique that will combine all of them, Section 3 defines a conceptual model of the formalized constructs, followed by Section 4 contrasting the work with related research along with the presentation of some special cases, and Section 5 ending the article with conclusions and further research.

Formalization

This formalization introduces a few simple constructs that enables the capturing of information that evolves over time, may have conflicting sources, and that can be less than certain. At its core, it deals with statements pertaining to the properties or relationships of things, where a *thing* for our purposes simply is that which can have properties and participate in relationships. The exact nature of what it entails to be a thing Cumming and Collier, "Change and identity in complex systems"[8] will be left to the reader to judge, but broadly speaking it is something that needs to be sufficiently different from

[4] Craig W Fisher and Bruce R Kingma. "Criticality of data quality as exemplified in two disasters". In: *Information & Management* 39.2 (2001), pp. 109–116

[5] Krishna Kulkarni and Jan-Eike Michels. "Temporal features in SQL: 2011". In: *ACM Sigmod Record* 41.3 (2012), pp. 34–43

[6] Tom Johnston. *Bitemporal data: theory and practice*. Newnes, 2014

[7] Omar Benjelloun et al. "Databases with uncertainty and lineage". In: *The VLDB Journal-The International Journal on Very Large Data Bases* 17.2 (2008), pp. 243–264; Maximilian Dylla, Iris Miliaraki, and Martin Theobald. "A temporal-probabilistic database model for information extraction". In: *Proceedings of the VLDB Endowment* 6.14 (2013), pp. 1810–1821

[8] Graeme S Cumming and John Collier. "Change and identity in complex systems". In: *Ecology and society* 10.1 (2005)

something else to be told apart. Some examples of things are perpe-trators, beards, places, samples, notes, and investigations, but also that which may not come immediately to mind, like thoughts, events, molecules, transactions, and moments in time.

Posits

First some purely syntactical constructions are needed. These will later, through the final construct defined, be given meaning. Their intention will, however, be exemplified along with the introduction of each construct. The distinction here between syntax and semantics is theoretically important but may be neglected in practice. Using syntax at random to create statements will produce a lot of non-sense when semantically tied to the universe of discourse, or in other words, irrelevant to that which is being modeled.

DEF. OF AN APPEARANCE

An *appearance* is a pair (i, r) where the first element is a unique identifier and the second a string.

The intention of an appearance is to capture the identity of a thing along with a role that thing is playing in a statement. If for the sake of simplicity we assume unique identifiers to be capital letters, some examples of appearances are:

$$(A, has\ gender)$$
$$(A, has\ beard)$$
$$(A, is\ accused)$$
$$(B, is\ victim)$$
$$(C, is\ witness)$$

These appearances convey that A is some thing that may have a specific gender, may have a beard of some kind, and that may have been an accused at some point. Similarly B and C may respectively have been a victim and a witness in some, but not necessarily the same, state of affairs. In order to bind appearances to a particular state of affairs, a dereferencing set is needed:

DEF. OF A DEREFERENCING SET

A *dereferencing set*, $\{(i_1, r_1), \ldots (i_n, r_n)\}$, is a set of appearances where any role r_i may only appear once.

The intention of a dereferencing set is to bind a single appearance or several appearances that have appeared or will appear in a state of

affairs. Given the example above, some thinkable dereferencing sets are:

$$\{(A, has\ gender)\}$$
$$\{(A, has\ beard)\}$$
$$\{(A, is\ accused), (B, is\ victim), (C, is\ witness)\}$$

While the first two trivially binds single appearances to some state of affairs, the third states that A, B, and C respectively appeared as the accused, the victim, and the witness in this particular state of affairs. Singularly bound appearances are more commonly referred to as *properties* and multiply bound appearances as *relationships*. Neither of these have yet taken on a value, which for both properties and relationships are expressed through posits:

DEF. OF A POSIT

A *posit* is a triple, $[D, v, t]$, where the first element is a dereferencing set, the second a data value of a simple or complex type, and the third a time point. The domain $\mathbb{t}$ from which time points are taken is called *appearance time*.

While posits still only are syntactical constructs, they will by later constructions be assigned some truth value. In fact, the word "posit" is suitably defined as "a statement that is made on the assumption that it will prove to be true" according to the Oxford English Dictionary Oxford University Press, *The Oxford English Dictionary*[9]. The intention of a posit is to state which value an appearance has taken since the specified point in time. Continuing the example the following are some possible posits:

[9] Oxford University Press. *The Oxford English Dictionary*. 2017. URL: https://en.oxforddictionaries.com/definition/posit (visited on 07/11/2018)

$$[\{(A, has\ gender)\}, male, 1972]$$
$$[\{(A, has\ beard)\}, fluffy\ red, 10{:}00]$$
$$[\{(A, has\ beard)\}, shaved\ clean, 10{:}02]$$
$$[\{(A, is\ accused), (B, is\ victim), (C, is\ witness)\},$$
$$at\ scene\ of\ crime, 09{:}58]$$

The intention of appearance time is to allow the representation of change, which occurs when two posits share the same dereferencing set, but with different values and time points. In the example posits above, Archie's beard (suspiciously quickly) changes from being fluffy red at 10:00 to shaved clean at 10:02. Archie, Bella, and Charlie are also related to each other and have different roles at the scene of the crime, where they all appeared at 09:58. The first posit is from the public records, showing that Archie is male and has remained so since his birth in 1972.

To further exemplify change, the following posit capture the fact that the group had dispersed shortly after the crime.

$$[\{(A, \textit{is accused}), (B, \textit{is victim}), (C, \textit{is witness})\},$$
$$\textit{dispersed}, 10\text{:}02]$$

Posits capture transitions, in which whatever is dereferenced can be said to enter a different state along with when this happened. Some states are likely to remain forever, such as the birth dates of the parties involved.

$$[\{(A, \textit{has birth date})\}, 1972\text{-}08\text{-}20, 1972\text{-}08\text{-}20]$$
$$[\{(B, \textit{has birth date})\}, 1980\text{-}02\text{-}13, 1980\text{-}02\text{-}13]$$

It is easy to be mislead into thinking that if a date of birth had been incorrectly stated, it could be changed by entering a new value with a later appearance time.

$$[\{(A, \textit{has birth date})\}, 1972\text{-}09\text{-}21, 1972\text{-}09\text{-}21]$$

This, however, would lead to ambiguity, since there is no way to tell which of the two coinciding posits is the correct one. If interpreted as a transition between states, as with Archie losing his beard, that would inconceivably mean that there was a time when Archie actually was born in August to be succeeded by a time when he was actually born in September. Corrections, then, must by necessity be handled differently, why more constructs are needed.

Assertions

Rather than to assume posits as meaningful and truthful, they are viewed as syntactical pieces of information towards which it is possible to hold different opinions, be uncertain about, and revise over time. To serve these purposes and to bring posits meaning, a new construct is needed.

DEF. OF AN ASSERTION

An *assertion* is a predicate, $!_A(P, p, \alpha, T)$, taking four arguments, where the first argument is a unique identifier, the second a posit, the third a real number in the range $[-1, 1]$, and the fourth a time point. The domain $\mathbb{T}$ from which time points are taken is called *assertion time*.

Being a predicate, and assertion evaluates to true or false, depending on its parameters. The unique identifier P represents a *positor*, the one holding an opinion about a posit. The real number α represents the *reliability* with which a positor holds its opinion and the time

point T represents when this opinion was asserted. A different asser-
tion with the same positor and posit represents a change of opinion.
Using assertions, concurrency can be expressed, as in the peculiarities
of Archie's beard:

$$p_1 = [\{(A, \textit{has beard})\}, \quad \text{fluffy red}, \quad 10:00]$$
$$p_2 = [\{(A, \textit{has beard})\}, \text{shaved clean}, \quad 10:02]$$
$$!_A(C, p_1, \quad 0.75, \text{Friday})$$
$$!_A(D, p_2, \quad 1, \text{Friday})$$

The two posits p_1 and p_2 are in conflict, since they are stating two
different values for the same dereferencing set, with two assertions
putting them both in effect after 10:02. In plain text, the assertions
say that "Donna stated on Friday that she is certain that Archie's
beard was shaved clean at and since 10:02" and "Charlie stated on
Friday that it is probable that Archie had a fluffy red beard at and
since 10:00". If both of these represent actual facts in the universe of
discourse, the assertions are true, which is the case in the running
example.

α	Linguistic correspondence
1	It is certain that the value is v
0.75	It is probable that the value is v
0.5	It may be that the value is v
0.25	It is possible that the value is v
0	It could be any value whatsoever
−0.25	It is possible the value is not v
−0.5	It may be that the value is not v
−0.75	It is probable that the value is not v
−1	It is certain that the value is not v

Table 1: Suggestions of some reliabilities and their linguistic correspondences for a value v.

Looking at the reliability, it is natural that $\alpha = 1$ corresponds to
the absolute certainty that Donna is expressing. The interpretation of
$\alpha = 0.75$ is more elusive, however. Where machines running proba-
bilistic models may yield exact reliabilities connected to their results,
humans tend to express themselves in a less exact manner. A possible
translation between natural language terms and their corresponding
reliabilities can be found in Table 1. While the meaning of the values
$\{-1, 0, 1\}$ must remain fixed, the values in between may be assigned
according to the requirements at hand.

Not surprisingly, when asked, Charlie was certain that Archie was
not clean shaved when he saw him at 10:00.

$$\bar{p}_3 = [\{(A, \textit{has beard})\}, \text{not shaved clean}, 10:00]$$
$$!_A(C, \bar{p}_3, 1, \text{Friday})$$

Since complementary values, such as "*not* shaved clean", are harder to store than actual values, the *not* in the posit can be removed in favor of a negative reliability:

$$p_3 = [\{(A, \textit{has beard})\}, \text{shaved clean}, 10:00]$$
$$!_{\mathsf{A}}(C, p_3, -1, \text{Friday})$$

This assertion is essentially making the same statement, where $\alpha = -1$ corresponds to being certain of the opposite of a posit. Given what it known so far, it can be concluded that the two positors, Donna and Charlie, are in strong disagreement. Where many of the existing techniques require such conflicts to be resolved at the time information is recorded, these conflicting statements can now be captured and instead reviewed at the time the information is retrieved. In other words, there are many versions of the truth, better reflecting the genuine nature of information.

Characteristics

When information is captured using assertions there are some characteristics that may be desirable. For example, if Charlie is making assertions about both the posit p_3 and its opposite $\bar{p}_3$ it is not desirable that these contradict each other. First a collection of assertions is needed:

DEF. OF A BODY OF INFORMATION

A *body of infomation* is a set of true assertions.

Given a body of information, a set of true assertions that represent actual facts in some universe of discourse,[10] some characteristics that pertain to their interrelationships can be defined.

DEF. OF SYMMETRIC

A body of information is said to be *symmetric* iff the assertions $!_{\mathsf{A}}(P, p, -\alpha, T)$ and $!_{\mathsf{A}}(P, \bar{p}, \alpha, T)$ are equivalent for a posit p and its opposite $\bar{p}$.

The first characteristic is symmetry. In a symmetrical body of information the assertion made by Charlie that "the reliability of Archie having a fluffy red beard is 0.75" is equivalent to asserting that "the reliability of Archie not having a fluffy red beard is -0.75". This quality makes it possible to remove the negation of a value, such as "*not* fluffy red", in a posit by reversing the sign of the reliability.

DEF. OF CANONICAL

A body of information is said to be *canonical* iff all assertions are made against posits without negated values.

[10] Likewise, a body of desinformation would be a set of false assertions, which is beyond the scope of this paper and perhaps less desirable.

It follows that any symmetrical body of information can be transformed to become canonical and thereby easier to manage from a storage perspective. Moving on, it is reasonable that for any assertion that is not certain, for example that "the reliability of Archie having a fluffy red beard is 0.75" also implies that there is a possibility that his beard is not red and fluffy. Intuitively the complementary assertion must be that "the reliability of Archie not having a fluffy red beard is 0.25", since reliabilities should sum up to 1.

DEF. OF BOUNDED

A body of information is said to be *bounded* iff the reliabilities in the assertions $!_A(P, p, \alpha, T)$ and $!_A(P, \bar{p}, \beta, T)$ satisfy $|\alpha + \beta| = |\alpha + \beta|^2$.

Since negative reliabilities are allowed, the formula becomes a little more complicated than just summing up reliabilities. For example, in a symmetrical and bounded body of information, if $\alpha = -0.75$ for $\bar{p}$ then $\beta = -0.25$ for p using boundary, and using symmetry that is equivalent to 0.25 for $\bar{p}$, which using boundary again yields the reliability of p as 0.75. The symmetry and boundary of values and their negation can be illustrated as a graph, seen in Figure 1.

There is still no characteristic preventing a positor from having a multitude of reliabilities for the same posit at the same time. Such a quality will also ensure that even though the reliabilities -0.25 and 0.75 for a posit produce equivalent assertions, only one should exist in the body of information.

Figure 1: The boundary and symmetry of reliability for a value v and its negation *not v*.

DEF. OF EXCLUSIVE

A body of information is said to be *exclusive* iff no assertions in which only the reliability differs exist.

An exclusive body of information prevents a positor from being, for example, certain of a fact and its opposite at the same time. Different positors may be in disagreement though, provided that it can be ensured that positors are having opinions about the same things. If one positor is saying that "Archie has a fluffy red beard" and another that "Archie was clean shaved", can we be sure they are talking about the same Archie and how do we know "*has beard*" means the same thing for them?

DEF. OF UNIVERSAL

A body of information is said to be *universal* iff positors agree on all appearances.

In a universal body of information positors are not allowed to interpret, for example, the appearance $(A, has\ beard)$ differently. Whatever the components represent will be the same for every positor.
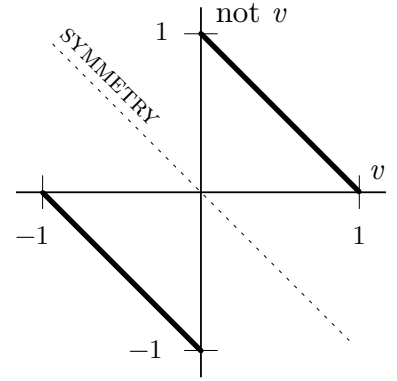
As assertions are predicates it is only through their representation that it can be determined if they are true or false. In order to have "meaningful" assertions, the bodies of information discussed hence-forth will therefore be assumed to have all the previously defined characteristics.

DEF. OF COMPREHENSIVE

A body of information is said to be *comprehensive* iff it is symmetric, canonical, bounded, exclusive and universal.

From here on it will be assumed that all bodies of information discussed are comprehensive.

Corrections

To be able to do corrections a particular, but intuitively understand-able, reliability value is needed.

DEF. OF COMPLETE UNCERTAINTY

An assertion is called *positive* when the reliability is above zero, *negative* when below zero, and *completely uncertain* when zero.

Zero reliability is best understood through an example. Given the following posits:

$$p_2 = [\{(A, has\ beard)\}, \quad \text{shaved clean}, 10\text{:}02]$$
$$\bar{p}_2 = [\{(A, has\ beard)\}, \text{not shaved clean}, 10\text{:}02]$$

For the posit p_2 and its opposite $\bar{p}_2$ the assertions $!_A(P, p_2, 0, T)$ and $!_A(P, \bar{p}_2, 0, T)$ are equivalent given symmetry. Since the reliability for the beard being both shaved clean and *not* shaved clean is the same, the only meaningful interpretation is that zero reliability must be when the positor has no idea what the value is. The positor may have had an idea at some earlier time though.

$$!_A(D, p_2, 1, \text{ Friday})$$
$$!_A(D, p_2, 0, \text{Sunday})$$

In the assertions above Donna is changing her mind from on Fri-day thinking that Archie was shaved clean to on Sunday not having been able to tell whether Archie had a beard or not. The later as-sertion is called a retraction when it changes a positive or negative reliability for an earlier posit to zero.

DEF. OF RETRACTION

Let $\mathbb{A}$ be a body of information. A *retraction* is an assertion $!_A(P, p, 0, T') \in \mathbb{A}$ for which there exists an assertion $!_A(P, p, \alpha, T) \in \mathbb{A}$, with $\alpha \neq 0$ and $T < T'$.

If at the same time another posit is asserted with non-zero reliability, this and the retraction together form a correction.

DEF. OF CORRECTION

Let $\mathbb{A}$ be a body of information. The three assertions $a_1 = !_A(P, p, \alpha, T)$, $a_2 = !_A(P, p, 0, T')$, and $a_3 = !_A(P, p', \alpha', T')$ together form a *correction* when $\{a_1, a_2, a_3\} \subset \mathbb{A}$, $\alpha \neq 0$, $\alpha' \neq 0$, and $T < T'$.

Extending the example with the assertion below, Donna says that it is possible that the perpetrator "Archie" had a fluffy red beard, while at the same time retracting her earlier assertion that his beard was shaved clean.

$$p_4 = [\{(A, \textit{has beard})\}, \text{fluffy red}, \text{10:02}]$$
$$!_A(D, p_4, 0.25, \text{Sunday})$$

Without the retraction Donna would simultaneously be stating that Archie's beard is certainly shaved clean and also that it is possibly fluffy red, so she is contradicting herself. Both assertions will remain in effect without the retraction, but exactly why the retraction will cancel out an earlier assertion has yet to be formalized.

The Information in Effect

With both assertions and posits being bound in time, this warrants the question of what information actually is in effect at any given point in time. The intention is that assertions of posits should capture transitions of states in the domain being modeled, such as people finding themselves at the scene of a crime and later dispersed or Archie removing his fake beard, as well as transitions in the opinions expressed towards such posits. Assertion and appearance time give rise to two axes of time over which it should be possible to "travel", where for all points in the bitemporal plane $\mathbb{T} \times \mathbb{t}$ an unambiguous set of assertions should be in effect.

DEF. OF INFORMATION IN EFFECT

Let $\mathbb{A}$ be a body of information. The *information in effect* is a subset $\mathbb{A}(T_@, t_@) \subset \mathbb{A}$ given a bitemporal point in assertion and appearance time $(T_@, t_@) \in \mathbb{T} \times \mathbb{t}$. Assuming P, α, T, D, v, and t are free variables, all assertions $!_A(P, p, \alpha, T) \in \mathbb{A}(T_@, t_@)$ with $p = [D, v, t]$ are found using the following selection criteria:

1   Let $\mathbb{A}_1 \subset \mathbb{A}$ be the assertions in $\mathbb{A}$ satisfying $T \leq T_@$ and $t \leq t_@$.

2   Let $\mathbb{A}_2 \subset \mathbb{A}_1$ be the assertions in $\mathbb{A}_1$ with $\alpha \neq 0$ and the latest t for each combination of P and D, excluding those for which an assertion $!_A(P, p, 0, T_0)$ exists with $T \leq T_0 \leq T_@$.

3   Let $\mathbb{A}(T_@, t_@) \subset \mathbb{A}_2$ be the assertions from $\mathbb{A}_2$ with the latest T for each combination of P and D.

In less formal terms, the information in effect can be found by first rewinding all assertions to those whose assertion and appearance times are before the specified points in time. Among those, find the ones with the latest appearance times for each positor and dereferencing set, excluding retracted ones. Finally, keep those with the latest assertion times, again for each positor and dereferencing set. This will guarantee that a single assertion for each existing combination of a positor and a posit is in effect at any bitemporal point in time. In a sense, such a "slice" of assertions is static, since all temporal aspects have been removed by the operations described. Had it not been for the remaining reliability it would have been close to what is usually stored in a traditional database. Gathering some assertions from the running example:

$$p_1 = [\{(A, has\ beard)\}, \quad \text{fluffy red}, \quad \text{10:00}]$$
$$p_2 = [\{(A, has\ beard)\}, \text{shaved clean}, \quad \text{10:02}]$$
$$p_3 = [\{(A, has\ beard)\}, \text{shaved clean}, \quad \text{10:00}]$$
$$a_1 = !_A(D, p_2, \qquad\qquad 1, \text{Friday})$$
$$a_2 = !_A(C, p_1, \qquad\qquad 0.75, \text{Friday})$$
$$a_3 = !_A(C, p_3, \qquad\qquad -1, \text{Friday})$$

Given the body of information above, $\{a_1, a_2, a_3\}$, the information in effect on (Saturday, 10:01) is a set with two members $\{a_2, a_3\}$, as they are the only assertions with appearance times on or before 10:01 and assertion times on or before Saturday. The information in effect on (Saturday, 10:03) is the set $\{a_1, a_2, a_3\}$, as all assertions are made before both times, and no retractions made. Although, something happened on Sunday, after Bella woke up:

$$p_4 = [\{(A, has\ beard)\}, \text{fluffy red}, \quad \text{10:02}]$$
$$a_4 = !_A(D, p_2, \qquad\qquad 0, \text{Sunday})$$
$$a_5 = !_A(D, p_4, \qquad 0.25, \text{Sunday})$$

With this added information that extends the body of information, on (Sunday, 10:03) the set $\{a_2, a_3, a_5\}$ is in effect. The assertion a_1 now has a corresponding retraction a_4 and must be excluded, which brings a_5 into effect. The two assertions, a_2 and a_3 made by Charlie are not contradictory, at the same time saying Archie's beard is probably fluffy red and certainly not shaved clean. In fact, "certainly not" may be stated for an arbitrary number of values without a positor contradicting itself. But, what if the following assertion is made:

$$a_6 = !_A(D, p_2, -0.75, \text{Sunday})$$

Suddenly, Donna is stating that Archie's beard is both possibly fluffy red and probably not shaved clean. Could it be the case that

Donna is now contradicting herself? When the same positor is having several assertions in effect with the same dereferencing set, but different values and reliabilities, another characteristic is needed ensuring positors are non-contradictory.

DEF. OF NON-CONTRADICTORY

A body of information is said to be *non-contradictory* iff for every information in effect the the reliabilities in all positive and negative assertions made by the same positor for the same dereferencing set satisfy the inequality:

$$\frac{1}{2} \sum_{i=1}^{n} \left[1 - \frac{\alpha_i}{|\alpha_i|} \right] + \sum_{i=1}^{n} \alpha_i \leq 1$$

Looking at $\{a_5, a_6\}$, these two assertions by Donna concurrently assert different values for the beard of Archie and they are both in effect on (Sunday, 10:03). Calculating the left side of the inequality gives:

$$\frac{1}{2} \left[1 - \frac{0.25}{|0.25|} + 1 - \frac{-0.75}{|-0.75|} \right] + 0.25 - 0.75 = 0.5$$

From this is can be concluded that Donna is far from contradicting herself, since the reliabilities only "sum up" to 0.5. Trivially a positor may make unlimited completely uncertain assertions, $\alpha = 0$, without contradicting itself. As another example, a positor may be 95 percent sure that Archie's beard is not shaved clean, and the same for up to 20 different beard styles and colors, without being contradictory. The 21st beard style Archie has almost certainly not got, however, will bring the left hand side to 1.05. Taking this example all the way, a positor may simultaneously assert with 100 percent certainty that something is *not* the case for an unlimited number of posits.

DEF. OF INDECISIVE AND DECISIVE

When multiple assertions by the same positor are in effect for different posits with the same dereferencing set, the positor is said to be *indecisive* with respect to the set. When a single assertion of a positor is in effect for a dereferencing set, the positor is said to be *decisive* with respect to the set. The information in effect is said to be *indecisive* if there is at least one indecisive positor and *decisive* if not. A body of information is said to be *decisive* if for every information in effect no indecisive positors exist and *indecisive* if there is at least one indecisive positor.

Since contradictions by definition rely on indecisiveness, it follows that a body of information which is decisive is always non-contradictory.[11]

[11] A traditional relational database is decisive as no possible alternative values can be represented, without modeling such behavior explicitly.

Reassertions and Restatements

In a traditional database, updating a value to the same value is an indistinguishable operation. For assertions, it is possible to assert the exact same thing over and over again with increasing assertion times. Such assertions can be chosen to be kept or discarded, depending on the requirements at hand.

DEF. OF REASSERTION AND ASSERTIVE

If two in assertion time successive assertions are otherwise equal, the later is said to be a *reassertion* of the earlier. A body of information is said to be *assertive* if reassertions exist.

It may be noted that in an assertive body of information, the information in effect will contain the latest reassertion, and if desired further searching is necessary in order to find the first assertion in a chain of reassertions. To further exemplify, Charlie will make three additional assertions, $\{a_7, a_8, a_9\}$:

$$p_1 = [\{(A, has\ beard)\}, \text{fluffy red}, \quad \text{10:00}]$$
$$p_4 = [\{(A, has\ beard)\}, \text{fluffy red}, \quad \text{10:02}]$$
$$a_2 = !_\text{A}(C, p_1, \quad 0.75, \text{Friday})$$
$$a_7 = !_\text{A}(C, p_4, \quad 0.75, \text{Friday})$$
$$a_8 = !_\text{A}(C, p_1, \quad 0.75, \text{Sunday})$$
$$a_9 = !_\text{A}(C, p_4, \quad 0.25, \text{Sunday})$$

Here the assertion a_8 is a reassertion of a_2, since only the assertion time differs. The assertions a_7 and a_2 share a different relationship though, where the assertion time is the same, but the appearance time differs while the dereferencing set and value is the same in the posit.

DEF. OF RESTATEMENT AND RESTATABLE

For two assertions, $!_\text{A}(P, p, \alpha, T)$ and $!_\text{A}(P, p', \alpha, T)$, with $p = [D, v, t]$ and $p' = [D, v, t']$, such that $t < t'$, the latter is called a *restatement* of the former. If a body of information contains restatements it is said to be *restatable*.

Given the definition above, the assertion a_7 is a restatement of a_2, and in fact, a_9 is a restatement of a_8. Looking at a_9 and a_7, Charlie is downgrading his belief in the posit, from being probable that the beard was fluffy red to only possible that is was. Perhaps he observed Archie for the full two minutes between 10:00 and 10:02, but could not at first believe his eyes that the beard was gone. With some afterthought, he was less sure on Sunday.

Modeling

While the posits and assertions seen so far have been intuitively understandable, being constructed from a story and simple examples, the exact nature of Archie is still undecided. Is it a perpetrator, an accused, a person, a human being, a fictional character, or something completely different? In order to determine what unique identifiers actually represent, a model is needed.

DEF. OF CLASSIFIER AND CLASS

Let $\mathbb{I}$ be the set of all unique identifiers found in a body of information $\mathbb{A}$. With $i \in \mathbb{I}$, a *classifier* is a predicate $!_C(P, i, C, \alpha, T)$, taking five arguments, where the first and second arguments are unique identifiers, the third a string called *class*, the fourth a real number in the range $[-1, 1]$, and the fifth a time point. The domain $\mathbb{T}^+$ from which time points are taken is called *modeling time*.

The string C in a classifier is intended to be a classification, such that it can be determined to which class any unique identifier belongs. As seen, the structure of a classifier is quite similar to that of an assertion[12]. This is deliberate, making it possible for a positor P to change the model over time, and have different and uncertain opinions about the classification. It also allows for conflicting views of the model, if more than one positor is present.

[12] Since classifiers behave like assertions in almost all respects, further formalization has been left out of the paper.

$$c_1 = !_C(M_1, A, \quad \text{Person}, 1, 2009)$$
$$c_2 = !_C(M_1, B, \quad \text{Person}, 1, 2009)$$
$$c_3 = !_C(M_1, C, \quad \text{Person}, 1, 2009)$$
$$c_4 = !_C(M_2, A, \text{Accused}, 1, 2012)$$
$$c_5 = !_C(M_2, B, \quad \text{Victim}, 1, 2012)$$
$$c_6 = !_C(M_2, C, \text{Witness}, 1, 2009)$$

The two modelers M_1 and M_2 have taken different approaches, with M_1 being more generic and M_2 more specific in their classification.

DEF. OF MODEL

Let $\mathbb{A}$ be a body of information and let $\mathbb{I}$ be the set of all unique identifiers found in $\mathbb{A}$. A *model* of $\mathbb{A}$ is a set of true classifiers with respect to $\mathbb{A}$ and its universe of discourse, that exhaust $\mathbb{I}$.

As seen from the example, there may be many ways to model the same body of information. The question then arises when a model is a "good" model. Unfortunately there is no simple answer. In some cases, few and generic classes are preferable, where in others

many and specific ones are better. To make a personal opinion and philosophical digression, a model is that which creates boundaries between similar and dissimilar things. To model is to define these boundaries by determining when things are similar enough to be of the same class.

DEF. OF POSIT TYPE

A *posit type*, $\tau(p) = [\{(C_1, r_1), \ldots, (C_n, r_n)\}, \tau(v), \tau(t)]$, for a posit $p = [\{(i_1, r_1), \ldots, (i_n, r_n)\}, v, t]$, is constructed by replacing the unique identifiers, i_i with their class, C_i, the value, v, with its data type, $\tau(v)$, and the time point, t, with its data type, $\tau(t)$.

Posit types are descriptive by nature, and will be used to enrich the model through a later construct. First though, given the "Person" class from the classifiers $\{c_1, c_2, c_3\}$, an example posit and its corresponding type are:

$$p_4 = [\{(A, \textit{has beard})\}, \textsf{fluffy red}, \textsf{10:02}]$$
$$\tau(p_4) = [\{(\textsf{Person}, \textit{has beard})\}, \textsf{string}, \textsf{time}]$$

While the posit type may look like a template, constraining which and how posits may be created, this is not their intention. They are only descriptive of what is found in the body of information. It may be the case that the following posits coexist in the same body of information.

$$p_5 = [\{(A, \textit{is accused}),$$
$$(B, \textit{is victim}),$$
$$(C, \textit{is witness})\},$$
$$\textsf{at scene of crime}, \textsf{09:58}]$$
$$p_6 = [\{(A, \textit{is accused}),$$
$$(X, \textit{is victim}),$$
$$(C, \textit{is witness})\},$$
$$\textsf{at scene of crime}, \textsf{09:42}]$$
$$\tau(p_5) = [\{(\textsf{Person}, \textit{is accused}),$$
$$(\textsf{Person}, \textit{is victim}),$$
$$(\textsf{Person}, \textit{is witness})\},$$
$$\textsf{string}, \textsf{time}]$$
$$\tau(p_6) = [\{(\textsf{Person}, \textit{is accused}),$$
$$(\textsf{Animal}, \textit{is victim}),$$
$$(\textsf{Person}, \textit{is witness})\},$$
$$\textsf{string}, \textsf{time}]$$

What seems to have started the kerfuffle between Archie and Bella was in fact that Archie was being mean to her dog, which Charlie witnessed somewhat earlier that day. This results in two simliar, but not equal, posit types $\tau(p_5)$ and $\tau(p_6)$.

DEF. OF ENSEMBLE

Let $\mathbb{A}$ be a body of information and let $\mathbb{P}$ be the set of all posit types constructed from the posits in $\mathbb{A}$. If C is a class from a model of $\mathbb{A}$, the *ensemble* of C is the set of all posit types from $\mathbb{A}$ in which C occurs.

With what we know so far from the running example, the ensemble of the "Person" class would be the following set of posit types:

$$\{[\{(\text{Person}, \textit{has beard})\}, \text{string}, \text{time}],$$
$$[\{(\text{Person}, \textit{has gender})\}, \text{int}, \text{year}],$$
$$[\{(\text{Person}, \textit{has birth date})\}, \text{date}, \text{date}],$$
$$[\{(\text{Person}, \textit{is accused}),$$
$$(\text{Person}, \textit{is victim}),$$
$$(\text{Person}, \textit{is witness})\},$$
$$\text{string}, \text{time}],$$
$$[\{(\text{Person}, \textit{is accused}),$$
$$(\text{Animal}, \textit{is victim}),$$
$$(\text{Person}, \textit{is witness})\},$$
$$\text{string}, \text{time}]\}$$

This ensemble gives a good idea of what it entails to be an instance of the "Person" class. It is important to note that even if one bearded person exists it does not imply that all persons can or must have beards. There will be no posits of type $\tau(p_4)$ for Bella or Donna, ever. If desired, the ensemble can be complemented with counts of how many posits match each type.

While sufficient to provide a model through a set of classifiers, the ensembles enrich the model by providing a more descriptive view of each class. All ensembles can be algorithmically deduced and they will naturally evolve over time, as new information is added. Together they can be likened with a growing schema in a database, but with a complete history of changes, and possibly multiple conflicting models of the same information.

Identification

A body of information is rarely something that comes about in a Big Bang kind of fashion. Rather, they are gradually built, and many may continue to evolve indefinitely. Herein lies a difficulty with the

unique identifiers that have already been assigned within that body. Say, if a new witness steps forward a few weeks later, is the Archie that witness saw the same Archie that already has been assigned the unique identifier *A*?

DEF. OF IDENTIFICATION

Let *C* be a class in a body of information $\mathbb{A}$ and *e* its ensemble. Given some circumstances, where circumstances consist of information not yet or only partly represented in $\mathbb{A}$, an *identification* is a search in *A* in which *e* is gradually populated with values and/or other unique identifiers, until they match existing posits, from some information in effect, such that a single unique identifier of class *C* can be determined or that after exhausting the circumstances no match was found.

In the case of the new witness, the circumstances may have been that that witness also saw a man at the same scene of the crime at the same time as the others and that that man also wore a red beard and fled shortly thereafter. When ensembles themselves have become large, it may be convenient to narrow down searches by providing some subsets of posit types that normally suffice for identification.

DEF. OF IDENTIFIER

Let *e* be the ensemble of a class *C*. An *identifier* of *C* is a subset $e_{id} \subset e$, for which it is deemed sufficient to fill with values and/or other unique identifiers in order to complete an identification.

Identification is, by all means and the definitions above, no exact science. There also may be many ways to reach the same conclusion in an identification and many different identifiers for the same class. A real perpetrator could perhaps be identified through DNA matching, fingerprints, witnesses, motives, means, opportunity, and so on. Since positors are allowed to make retractions, this may make identification even harder.

DEF. OF ANALYTIC AND SYNTHETIC

Let *p* be a posit. If all positors retract *p* and this is debilitating to the identification process, then *p* is said to be *analytic* to *i*, or otherwise *synthetic* to *i*.

Eliot, the investigator, swabbed Archie shortly after capturing him, resulting in his DNA being sequenced.

$$p_7 = [\{(A, has\ DNA)\}, \text{ACGT}\ldots, \text{1972-08-20}]$$

$$a_{10} = !_\mathbb{A}(E, p_7, \qquad 1, \quad \text{Sunday})$$

If Eliot retracts a_{10} and this would seriously debilitate the identification process, that means that Archie's DNA is analytic to Archie.

Analytic posits need to be handled with additional care as the information evolves, since retracting them may have negative side effects[13].

Cardinality

$$p_{1w} = [\{(A, is\ husband), (B, is\ wife), (C, is\ minister), (8, is\ church)\}, marriage, 2004]$$
$$p_{2w} = [\{(A, is\ husband), (B, is\ wife), (C, is\ minister), (8, is\ church)\}, divorce,\ \ \ 2009]$$
$$p_{3w} = [\{(A, is\ husband), (F, is\ wife), (C, is\ minister), (8, is\ church)\}, marriage, 2012]$$
$$p_{4w} = [\{(G, is\ husband), (B, is\ wife), (C, is\ minister), (9, is\ church)\}, marriage, 2012]$$

Figure 2: Some example posits of a quaternary "wedding" relationship.

Eliot's investigation would also show that Archie and Bella knew each other quite well. The posits in Figure 2 describe a series of events, from Bella marrying Archie in 2004, to divorcing him in 2009, and to both remarrying others in 2012. Of these, only the first two constitute a real change, since they have the same dereferencing set. The last two instead constitute different events, as the dereferencing sets involve different unique identifiers.

With combined temporality, concurrency, and reliability, it is of interest to understand constraints under the circumstances where values may change over time, positors may be in conflict, and retractions occur. Describing the specific case of cardinality in transitional modeling, but a common constraint in almost every technique, will help shed light on the general case of how any constraint is handled. Consider enforcing monogamy[14] in the "wedding" relationship, where a married couple may come to divorce, different positors may have conflicting opinions about when it happened, and they may change their minds about whether it happened at all.

DEF. OF LIMITER

A *limiter* is a triple, (R, l, u), where $R = \{r_1, \ldots, r_n\}$ is a set of strings and l and u are non-negative integers, with the special value $*$ also allowed for u. If $u = *$, then l may be any non-negative integer, or otherwise $l \leq u$ must hold.

The intention of a limiter is to express a lower, l, and upper, u, bound for how many times a unique identifier or combination thereof may appear having the roles from R in a relationship. The special value $*$ should be interpreted as an "unlimited" number of times.

DEF. OF CARDINALITY CONSTRAINT

A *cardinality constraint*, $\{L_1, \ldots, L_n\}$ over a posit type, $\tau(p)$, is a set of limiters, $L_i = (R_i, l_i, u_i)$, with $R_i = \{r_{i_1}, \ldots, r_{i_n}\}$ for which every r_{i_j} is a role in the posit type $\tau(p)$.

In order to enforce monogamy, first recognize that in Figure 2 all posits share the same posit type.

$$\tau(p_{1w}) = [\{(\text{Person}, is\ husband),$$
$$(\text{Person}, is\ wife),$$
$$(\text{Person}, is\ minister),$$
$$(\text{Location}, is\ church)\},$$
$$\text{string}, \text{year}]\}$$

Preventing a husband from having several wives and wives from having several husbands needs the following cardinality constraint on $\tau(p_{1w})$.

$$\{(\{is\ husband\}, 1, 1), (\{is\ wife\}, 1, 1)\}$$

In other words, the unique identifier for a husband must appear at least once and at most once at the same time as the unique identifier for a wife also must appear exactly once. This enforces a one-to-one relationship between husbands and wives. However, already the first two posits in Figure 2 seem to break this constraint.

DEF. OF DECISIVE FULFILMENT

A cardinality constraint is said to be *fulfilled* in a decisive body of information $\mathbb{A}$ if for every positor P and by them asserted posits p_k, such that $\tau(p_k) = \tau(p)$, unique identifiers appear at least l times and at most u times in the roles specified in the corresponding limiter, for each positor and every information in effect, $\mathbb{A}(T_@, t_@)$.

This is where it is important to notice that constraint checks are not done against all posits in a body of information, but rather made and maintained for each information in effect in that body. This pattern repeats itself for any type of constraint. Fulfilment, as can be seen, is only defined for decisive bodies of information. Fulfilment in indecisive bodies will not be detailed in this paper, but to give an idea, if Eliot asserts with 0.50 reliability both that Archie is currently married either to Bella or to Fanny, he could be stating that there is a chance Archie is married to both of them or that in reality, one wife is a fact and the other is not, but he is not sure which one it is. Cardinalities in indecisive bodies of information may aid in this, but therefore also need additional treatment, which is left as further

research. Now, let Eliot instead decisevly assert the posits.

$$a_{1w} = !_A(E, p_{1w}, 1, \text{Sunday})$$
$$a_{2w} = !_A(E, p_{2w}, 1, \text{Sunday})$$
$$a_{3w} = !_A(E, p_{3w}, 1, \text{Sunday})$$
$$a_{4w} = !_A(E, p_{4w}, 1, \text{Sunday})$$

The information in effect on $(\text{Sunday}, 2004)$ is $\{a_{1w}\}$, on $(\text{Sunday}, 2009)$ it is $\{a_{2w}\}$, and on $(\text{Sunday}, 2012)$ it is $\{a_{3w}, a_{4w}\}$, exhausting all possible informations in effect. Neither of these break the constraint, so the constraint is therefore fulfilled. Let Both Archie and Bella admit they were married in 2004 through the following two assertions.

$$a_{5w} = !_A(A, p_{1w}, 1, \text{Sunday})$$
$$a_{6w} = !_A(B, p_{1w}, 1, \text{Sunday})$$

Each of the sets with the information in effect above will now also have these two assertions included. This will, however, still not break the constraint, since each assertion that could have done so is made by a different positor. The pattern that constraint checks are made and maintained for each positor also repeats itself for any type of constraint.

Given the quaternary "wedding" relationship, there are a lot more cardinality constraints that can be defined. For the sake of completeness, the following is defined.

DEF. OF COMPLETE CARDINALITY[15]

The *complete cardinality* of a posit type, $\tau(p)$, is a set of all cardinality constraints that can be created using combinations of roles from $\tau(p)$.

[15] The size of the complete set is $n + (3^n - 2^{n+1} + 1)/2$, giving 3, 9, 29, 95, and 307 elements for 2, 3, 4, 5, and 6-ary posits.

Already for a quaternary relationship, the complete cardinality set has 29 members. Three of them are exemplified through the following constraints; two 2-ary and one 4-ary.

$$\{(\{is\ husband\}, 1, *), (\{is\ minister\}, 1, 1)\},$$
$$\{(\{is\ husband\}, 1, *),\quad (\{is\ church\}, 0, 1)\},$$
$$\{(\{is\ husband,\qquad is\ wife\}, 1, *)\ ,$$
$$(\{is\ minister,\quad is\ church\}, 1, 1)\}$$

The first constraint expresses that if a husband is present at a wedding exactly one minister must be as well and that a minister may wed many husbands. The second expresses that if a husband is present at a wedding it may or may not take place in a church and that any church may be the location of many husbands being wed. The third expresses that if a husband and a wife is present at a

wedding in which there is a combination of a minister and a church, there must be only one such combination, and that a combination of a minister and a church may have partaken in the wedding of many husbands and wives.

Intuitively, there should be some relation between the first two constraints and the third. While not immediately obvious, it turns out that the first two, or any pairwise cardinality for that matter, can be used to calculate bounds for those of higher order. Being slightly less formal in how these are calculated, first, define a "cross product" according to the following rules.

$$(l_i, u_i) \times (l_j, u_j) = (max(l_i, l_j), min(u_i, u_j))$$
$$min(u, *) = u$$
$$max(l, *) = l$$
$$min(*, *) = max(*, *) = *$$

To find the bounds for the 4-ary constraint, first calculate from left to right and crosswise roles, (husband side to minister) $\times$ (husband side to church) $\times$ (wife side to minister) $\times$ (wife side to church) = (1, *) $\times$ (1, *) $\times$ (1, *) $\times$ (1, *) = (1, *). The calculate from right to left and crosswise roles, (minister side to husband) $\times$ (minister side to wife) $\times$ (church side to husband) $\times$ (church side to wife) = (1, 1) $\times$ (1, 1) $\times$ (0, 1) $\times$ (0, 1) = (1, 1). Incidentally this gives the cardinality found in the exemplified 4-ary constraint, which cannot pass beyond these bounds without resulting in inconsistent cardinalities. Further restricting the constraint is allowed though, and warrants the need for actually specifying higher order constraints. For example, if a minister is only allowed to wed five couples in any given church the 4-ary constraint can be restricted to:

$$\{(\{is\ husband, \quad is\ wife\}, 1, 5)\ ,$$
$$(\{is\ minister,\ is\ church\}, 1, 1)\}$$

Declaring constraints in transitional modeling turns out to be very similar to traditional techniques, with the complexity stemming from its rich features instead add to how they are enforced.

Related Research

Some 20 years ago Inmon, *Building the data warehouse*[16] defined a Data Warehouse as being "a subject-oriented, integrated, time-variant, nonvolatile collection of data". While companies are belatedly coming to terms with the first two conditions, in our experience, supported by Golfarelli and Rizzi, "A survey on temporal data warehousing"[17], few can say that they fulfil the last two, primarily be-

[16] William H Inmon. *Building the data warehouse*. John wiley & sons, 2005

[17] Matteo Golfarelli and Stefano Rizzi. "A survey on temporal data warehousing". In: *International Journal of Data Warehousing and Mining (IJDWM)* 5.1 (2009), pp. 1–17

cause of lacking methods. If Data Warehouses are our best attempt at integrating bodies of information under evolution, they are currently not well equipped to do so. To address this the presented *transitional modeling* technique, rooted in multiple research areas, makes it easy to communicate, store, manage, and analyze information that is conflicting, uncertain and varying.

Bradley, *Appearance and reality: a metaphysical essay*[18] shares the notion that a thing is defined by its appearance and change is "a bond of identity in differences". Through positors and appearances problems such as "the Ship of Theseus", discussed by Chisholm et al., *Person and object: A metaphysical study*[19], can be largely avoided. Even if every plank in the ship is replaced, the unique identifier once associated with the ship will prevail, and will do so eternally Sider, "Quantifiers and temporal ontology"; Black, "The identity of indiscernibles"[20]. If at some point another identifier is associated with the ship, because a different positor deems it a new ship, that positor is perfectly free to do so. The two ships can then be discussed in parallel, and relations between them defined. To paraphrase the greek philosopher Heraclitus [~500 BC], it is possible for the same man to step into the same river twice, as long as some positor can identify them to still be the same. Being the same man or the same river rarely comes down to following the trail of body cells or water molecules.

Lombard, "Relational change and relational changes"[21] presents the problem of distinguishing relational from non-relational change, which is solved by separating immutable parts, allowing change to happen independently across a model, similar to how our unique identifiers are the only immutables, while everything else may change. Moens and Steedman, "Temporal ontology and temporal reference"[22] define "culmination" as being a point in time "accompanied by a transition to a new state of the world", similar to how our appearance time is used to capture different states. Bradley, *Appearance and reality: a metaphysical essay*[23] also states that shifting to statements and away from reality relaxes the need for absolute truth. Similarly our assertions make it possible to hold different and uncertain opinions about posits. An RDF triple W3C RDF Working Group, *Resource Description Framework*[24] roughly corresponds to our posit, but has no concept beyond it with respect to temporality, reliability, or concurrency. The quality of a posit being "analytic" or "synthetic" is borrowed directly from the philosophical works of Ayer, "Language, Truth, and Logic"; Ramsey, "Truth and Probability. Reprinted in HE Kyburg and HE Smokler"[25].

Johnston, *Bitemporal data: theory and practice*[26] suggests an approach based on speech act theory with actors resembling positors

[18] Francis Herbert Bradley. *Appearance and reality: a metaphysical essay*. Routledge, 2016

[19] Roderick Chisholm et al. *Person and object: A metaphysical study*. Routledge, 2014

[20] Theodore Sider. "Quantifiers and temporal ontology". In: *Mind* 115.457 (2006), pp. 75–97; Max Black. "The identity of indiscernibles". In: *Mind* 61.242 (1952), pp. 153–164

[21] Lawrence Brian Lombard. "Relational change and relational changes". In: *Philosophical Studies* 34.1 (1978), pp. 63–79

[22] Marc Moens and Mark Steedman. "Temporal ontology and temporal reference". In: *Computational linguistics* 14.2 (1988), pp. 15–28

[23] Francis Herbert Bradley. *Appearance and reality: a metaphysical essay*. Routledge, 2016

[24] W3C RDF Working Group. *Resource Description Framework*. 2004. URL: http://www.w3.org/RDF/ (visited on 07/11/2018)

[25] Alfred Ayer. "Language, Truth, and Logic". In: *London: V. Gollancz ltd* (1936); Frank Ramsey. "Truth and Probability. Reprinted in HE Kyburg and HE Smokler". In: *Studies in Subjective Probability* (1926), pp. 25–52

[26] Tom Johnston. *Bitemporal data: theory and practice*. Newnes, 2014

and speech acts bound to two axes of time, but does not take reliability into account. Reichenbach, "The tenses of verbs"[27] was among the first to suggest bitemporality; a difference between the "point of speech" and the "point of event", similar to our assertion time and appearance time. The specifics of bitemporal models in databases was later elaborated upon by Snodgrass and Ahn, "A taxonomy of time databases"[28].

Reliability has likenesses with the Uncertainty Theory of Liu, *Uncertainty Theory: An Introduction to its Axiomatic Foundations. 2004*[29] and its personal belief degree. While Liu states that "when the personal knowledge changes, the belief degree changes too", and that "different people may produce different belief degrees", his theory does not as of yet support these notions. His "normality" and "duality" axioms correspond to our definitions of non-contradiction and boundary.

Massaro and Friedman, "Models of integration given multiple sources of information."; Lin and Mendelzon, "Merging databases under constraints"[30] highlight the difficulty of integrating information from multiple sources, especially when these may be contradictory. The field of information fusion is flourishing, due to requirements to integrate extreme volumes of sensory data, in which current techniques try to integrate in real time Smets and Kennes, "The transferable belief model"; Liu et al., "Sequential adaptive combination of unreliable sources of evidence"[31]. This, naturally, leads to loss of the actual circumstances, and perhaps this field could benefit from storing actual data and delaying integration until decisions need to be made.

Benjelloun et al., "Databases with uncertainty and lineage"[32] introduces databases with uncertainty and lineage, with confidence values similar to reliability and lineage similar to positors, but that lack the concept of temporality. Dylla, Miliaraki, and Theobald, "A temporal-probabilistic database model for information extraction"[33] extends similar databases with "valid time", but does not make the connexion to transitions in opinions, which are tracked over our assertion time. Papafragou, "Epistemic modality and truth conditions"[34] warrants such connexions in her linguistics paper, stating that "subjective epistemic modality is time-dependent".

Thalheim, "The art of conceptual modelling"[35] holds that modeling is an art and that models provide boundaries between similar and dissimilar things. Furthermore, that models need to evolve along with the environment they are meant to capture. Our ability to track the evolution of a model over time is similar to requirements in ontology evolution Noy and Klein, "Ontology evolution: Not the same as schema evolution"[36] and database schema evolution McBrien and

[27] Hans Reichenbach. "The tenses of verbs". In: *Time: From Concept to Narrative Construct: a Reader* (1947)

[28] Richard Snodgrass and Ilsoo Ahn. "A taxonomy of time databases". In: *ACM Sigmod Record* 14.4 (1985), pp. 236–246

[29] B Liu. *Uncertainty Theory: An Introduction to its Axiomatic Foundations. 2004.* Springer-Verlag, Berlin, 2004

[30] Dominic W Massaro and Daniel Friedman. "Models of integration given multiple sources of information." In: *Psychological review* 97.2 (1990), p. 225; Jinxin Lin and Alberto O Mendelzon. "Merging databases under constraints". In: *International Journal of Cooperative Information Systems* 7.01 (1998), pp. 55–76

[31] Philippe Smets and Robert Kennes. "The transferable belief model". In: *Artificial intelligence* 66.2 (1994), pp. 191–234; Zhun-ga Liu et al. "Sequential adaptive combination of unreliable sources of evidence". In: *Advances and Applications of DSmT for Information Fusion* 2 (2015), p. 23

[32] Omar Benjelloun et al. "Databases with uncertainty and lineage". In: *The VLDB Journal-The International Journal on Very Large Data Bases* 17.2 (2008), pp. 243–264

[33] Maximilian Dylla, Iris Miliaraki, and Martin Theobald. "A temporal-probabilistic database model for information extraction". In: *Proceedings of the VLDB Endowment* 6.14 (2013), pp. 1810–1821

[34] Anna Papafragou. "Epistemic modality and truth conditions". In: *Lingua* 116.10 (2006), pp. 1688–1702

[35] Bernhard Thalheim. "The art of conceptual modelling". In: *Information Modelling and Knowledge Bases XXII. Frontiers in Artificial Intelligence and Applications* 237 (2012), pp. 149–168

[36] Natalya F Noy and Michel Klein. "Ontology evolution: Not the same as schema evolution". In: *Knowledge and information systems* 6.4 (2004), pp. 428–440

Poulovassilis, "Schema evolution in heterogeneous database architectures, a schema transformation approach"[37]. Our rich take on classifiers is probably unique, but the concept of an ensemble is directly borrowed from Hultgren, *Modeling the agile data warehouse with data vault*[38].

Concerning constraints, some work on higher-order cardinality in particular has been made by Hartmann, "On the implication problem for cardinality constraints and functional dependencies"[39]. A more generic discussion with respect to cardinalities in conjunction with temporality and the complexity that arises can be found in Allen, "Maintaining knowledge about temporal intervals"[40]. While positors need to be precise with respect to appearance and assertion time, this may be limiting in some cases, and research into imprecise time intervals has been made by Koubarakis, "Representation and querying in temporal databases: the power of temporal constraints"[41]. It is possible that the introduction of complex data types for representing time may satisfy such needs, but that is yet to be researched.

A good overview of research concerning erroneous, imprecise and uncertain information can be found in Smets, "Imperfect information: Imprecision and uncertainty"; Motro, "Imprecision and uncertainty in database systems"[42]. Erroneous data is to some extent handled by our retractions. Imprecise information, such as a value being somewhere within some limits, say for example that Charlie estimated Archie's length to be in the 170–180cm range, can be handled by introducing complex data types for representing values. Uncertainty is handled through our reliability.

Special Cases

The fundamental nature of *transitional modeling* combined with its ability to capture rather generic information results in many other modeling techniques being expressable as special cases of this technique. Three of them are Anchor Modeling, Data Vault, and the third normal form.

Starting with Anchor Modeling **RonnbackEtAl**[43], it is similarly constructed around immutable unique identifiers, called "anchors". Its "attributes" and "ties" correspond to posits with respectively single or multiple appearances in their dereferencing sets. These posits are also limited to primitive data types for values and times, except for "knotted" attributes and ties, for which a complex value type in the form of an enumeration is used. Their "changing time" is over which states change in the domain being modeled, exactly like our appearance time. As it, in the form presented in the referenced paper, has no construct corresponding to assertions, the assumption

[37] Peter McBrien and Alexandra Poulovassilis. "Schema evolution in heterogeneous database architectures, a schema transformation approach". In: *International Conference on Advanced Information Systems Engineering*. Springer. 2002, pp. 484–499

[38] Hans Hultgren. *Modeling the agile data warehouse with data vault*. New Hamilton, 2012

[39] Sven Hartmann. "On the implication problem for cardinality constraints and functional dependencies". In: *Annals of Mathematics and Artificial Intelligence* 33.2-4 (2001), pp. 253–307

[40] James F Allen. "Maintaining knowledge about temporal intervals". In: *Communications of the ACM* 26.11 (1983), pp. 832–843

[41] Manolis Koubarakis. "Representation and querying in temporal databases: the power of temporal constraints". In: *Data Engineering, 1993. Proceedings. Ninth International Conference on*. IEEE. 1993, pp. 327–334

[42] Philippe Smets. "Imperfect information: Imprecision and uncertainty". In: *Uncertainty management in information systems*. Springer, 1997, pp. 225–254; Amihai Motro. "Imprecision and uncertainty in database systems". In: *Fuzziness in Database Management Systems*. Springer, 1995, pp. 3–22

[43] **RonnbackEtAl**

is that there is a single positor asserting every posit with absolute certainty since the dawn of time. Unpublished work is available on their homepage **Anchor**[44], concerning extensions that allow for functionality along the lines of transitional modeling.

[44] **Anchor**

Continuing with Data Vault Linstedt and Graziano, *Super Charge Your Data Warehouse: Invaluable Data Modeling Rules to Implement Your Data Vault*[45], it is similarly constructed around immutable unique identifiers, called "hubs", but where hubs also contain at least one analytic posit with a complex data type for its value, consisting of a set of primitive values. Synthetic posits become either satellites or links, depending on whether they respectively have single or multiple appearances in their dereferencing sets. Both of these also use complex data types for their values, consisting of sets of primitive values. There is no definite rule for when such a set of primitive values may be divided in order to create another satellite or link, but the practice seems to be related to their rate of change Hultgren, *Modeling the agile data warehouse with data vault*[46]. Their "loading time" is over which states change in the database, rather than the domain being modeled, so it is different from our appearance time. There are no constructs corresponding to assertions, but loading time is closer to assertion time, if assumed that the database is the single positor asserting every posit with absolute certainty at the time of loading.

[45] Dan Linstedt and Kent Graziano. *Super Charge Your Data Warehouse: Invaluable Data Modeling Rules to Implement Your Data Vault*. CreateSpace, 2011

[46] Hans Hultgren. *Modeling the agile data warehouse with data vault*. New Hamilton, 2012

Finishing with the third normal form Kent, "A simple guide to five normal forms in relational database theory"; Codd, "A relational model of data for large shared data banks"[47], it is not constructed around immutable unique identifiers, but rather around analytic posits. A "table" in third normal form consists of at least one analytic posit along with optional synthetic posits, both having complex data types for their values, consisting of sets of primitive values. The cardinality of a posit type with multiple roles determines to which tables they belong. For one-to-one and one-to-many the posit is broken apart and put into both tables. For many-to-many or higher order than pairwise cardinalities, a separate table is needed. In third normal form posits with multiple appearances in their dereferencing sets carry no value, but it is likely that if they do, a separate table is always needed. It also lacks constructs to manage state changes, so it assumed that all posits are eternally in the same state, until destructively updated to the next assumed eternal state. There is no construct corresponding to assertions, so the assumption is that there is a single positor asserting every posit with absolute certainty since the dawn of time.

[47] William Kent. "A simple guide to five normal forms in relational database theory". In: *Communications of the ACM* 26.2 (1983), pp. 120–125; Edgar F Codd. "A relational model of data for large shared data banks". In: *Communications of the ACM* 13.6 (1970), pp. 377–387

Conclusions and Further Research

The presented technique, *transitional modeling*, should have a large number of application in a number of different fields of research, as well as in practical applications. It can, for example, be used to bridge the gap between what a Data Warehouse should be and what it currently is in most implementations. It should also be of interest for operational system databases and other types of systems in which it is of importance never to lose information. In research areas where information is influenced by variation, uncertainty, or opinions, this technique may bring a refreshing approach to integration, such that it can be moved from gathering time to the time of the decision making. Classifiers, and their derived ensembles, make it possible to have multiple evolving schemas over the same information, which should prove useful where not only the content, but also the structure, of information changes rapidly. The story in the running example is illustrative of information coming from humans, sensors, predictions, ratings, tracking, and other multifaceted sources. Such can now be fully modeled, stored, communicated, and analyzed, through the introduced concepts of posits and assertions.

More research is needed in order to fully understand constraints in indecisive bodies of information. Another area of interest is which complex data types for values and times could enrich the theory even further. Ranges would be an immediate candidate for dealing with imprecise information, for example. Most of the paper assumes bodies of information to be comprehensive, but there may be many fruitful conclusions that can be drawn from other types of bodies. It is also not unthinkable that false assertions, desinformation, could be interesting to represent, but it is unclear what the repercussions to the theory would be. A new massively parallel processing database engine, built using posits and assertions at its core, is in early stages of development, but is a natural venue for our further research. We believe strongly that it is only when a theory is converted to practice that you see its true benefits and limitations.

References

[1]   James F Allen. "Maintaining knowledge about temporal intervals". In: *Communications of the ACM* 26.11 (1983), pp. 832–843.

[2]   Alfred Ayer. "Language, Truth, and Logic". In: *London: V. Gollancz ltd* (1936).

[3]   Omar Benjelloun et al. "Databases with uncertainty and lineage". In: *The VLDB Journal-The International Journal on Very Large Data Bases* 17.2 (2008), pp. 243–264.

[4]   Max Black. "The identity of indiscernibles". In: *Mind* 61.242 (1952), pp. 153–164.

[5]   Francis Herbert Bradley. *Appearance and reality: a metaphysical essay*. Routledge, 2016.

[6]   Roderick Chisholm et al. *Person and object: A metaphysical study*. Routledge, 2014.

[7]   Edgar F Codd. "A relational model of data for large shared data banks". In: *Communications of the ACM* 13.6 (1970), pp. 377–387.

[8]   Graeme S Cumming and John Collier. "Change and identity in complex systems". In: *Ecology and society* 10.1 (2005).

[9]   Maximilian Dylla, Iris Miliaraki, and Martin Theobald. "A temporal-probabilistic database model for information extraction". In: *Proceedings of the VLDB Endowment* 6.14 (2013), pp. 1810–1821.

[10]   Craig W Fisher and Bruce R Kingma. "Criticality of data quality as exemplified in two disasters". In: *Information & Management* 39.2 (2001), pp. 109–116.

[11]   Matteo Golfarelli and Stefano Rizzi. "A survey on temporal data warehousing". In: *International Journal of Data Warehousing and Mining (IJDWM)* 5.1 (2009), pp. 1–17.

[12]   Sven Hartmann. "On the implication problem for cardinality constraints and functional dependencies". In: *Annals of Mathematics and Artificial Intelligence* 33.2-4 (2001), pp. 253–307.

[13]   Hans Hultgren. *Modeling the agile data warehouse with data vault*. New Hamilton, 2012.

[14]   William H Inmon. *Building the data warehouse*. John wiley & sons, 2005.

[15]   Tom Johnston. *Bitemporal data: theory and practice*. Newnes, 2014.

[16]   William Kent. "A simple guide to five normal forms in relational database theory". In: *Communications of the ACM* 26.2 (1983), pp. 120–125.

[17]   Manolis Koubarakis. "Representation and querying in temporal databases: the power of temporal constraints". In: *Data Engineering, 1993. Proceedings. Ninth International Conference on*. IEEE. 1993, pp. 327–334.

[18]   Krishna Kulkarni and Jan-Eike Michels. "Temporal features in SQL: 2011". In: *ACM Sigmod Record* 41.3 (2012), pp. 34–43.

[19]   Jinxin Lin and Alberto O Mendelzon. "Merging databases under constraints". In: *International Journal of Cooperative Information Systems* 7.01 (1998), pp. 55–76.

[20]   Dan Linstedt and Kent Graziano. *Super Charge Your Data Warehouse: Invaluable Data Modeling Rules to Implement Your Data Vault*. CreateSpace, 2011.

[21]   B Liu. *Uncertainty Theory: An Introduction to its Axiomatic Foundations. 2004*. Springer-Verlag, Berlin, 2004.

[22]   Zhun-ga Liu et al. "Sequential adaptive combination of unreliable sources of evidence". In: *Advances and Applications of DSmT for Information Fusion* 2 (2015), p. 23.

[23]   Lawrence Brian Lombard. "Relational change and relational changes". In: *Philosophical Studies* 34.1 (1978), pp. 63–79.

[24]   Dominic W Massaro and Daniel Friedman. "Models of integration given multiple sources of information." In: *Psychological review* 97.2 (1990), p. 225.

[25]   Peter McBrien and Alexandra Poulovassilis. "Schema evolution in heterogeneous database architectures, a schema transformation approach". In: *International Conference on Advanced Information Systems Engineering*. Springer. 2002, pp. 484–499.

[26]   Marc Moens and Mark Steedman. "Temporal ontology and temporal reference". In: *Computational linguistics* 14.2 (1988), pp. 15–28.

[27]   Amihai Motro. "Imprecision and uncertainty in database systems". In: *Fuzziness in Database Management Systems*. Springer, 1995, pp. 3–22.

[28]   Natalya F Noy and Michel Klein. "Ontology evolution: Not the same as schema evolution". In: *Knowledge and information systems* 6.4 (2004), pp. 428–440.

[29]   Oxford University Press. *The Oxford English Dictionary*. 2017. URL: https://en.oxforddictionaries.com/definition/posit (visited on 07/11/2018).

[30]   Anna Papafragou. "Epistemic modality and truth conditions". In: *Lingua* 116.10 (2006), pp. 1688–1702.

[31]   Frank Ramsey. "Truth and Probability. Reprinted in HE Kyburg and HE Smokler". In: *Studies in Subjective Probability* (1926), pp. 25–52.

[32]   Hans Reichenbach. "The tenses of verbs". In: *Time: From Concept to Narrative Construct: a Reader* (1947).

[33]   Theodore Sider. "Quantifiers and temporal ontology". In: *Mind* 115.457 (2006), pp. 75–97.

[34]   Philippe Smets. "Imperfect information: Imprecision and uncertainty". In: *Uncertainty management in information systems*. Springer, 1997, pp. 225–254.

[35]   Philippe Smets and Robert Kennes. "The transferable belief model". In: *Artificial intelligence* 66.2 (1994), pp. 191–234.

[36]   Richard Snodgrass and Ilsoo Ahn. "A taxonomy of time databases". In: *ACM Sigmod Record* 14.4 (1985), pp. 236–246.

[37]   Bernhard Thalheim. "The art of conceptual modelling". In: *Information Modelling and Knowledge Bases XXII. Frontiers in Artificial Intelligence and Applications* 237 (2012), pp. 149–168.

[38]   W3C RDF Working Group. *Resource Description Framework*. 2004. URL: http://www.w3.org/RDF/ (visited on 07/11/2018).