A Computable Representation of the Babylonian Astronomical Diaries

Christopher Wolfram

1 Introduction

The term "astronomical diaries" is a modern one, referring to a genre of Babylonian tablets which record astronomical observations, although they also contain other information.[1] The earliest known diary-like tablet was produced in the 7th century BC, with the diary tradition likely continuing until the first century AD, about 800 years later, making the astronomical diaries the longest running data collection project in history.[2] Over that period, the diaries record information about weather, economics, major events, astronomical phenomena, and more, making them a rich source of information in a variety of areas.

The astronomical diaries were first comprehensively published from 1988 to 1996 by the combined work of Abraham Sachs and Hermann Hunger.[3] They produced transliterations of the Akkadian and translations into English. They also aligned the diaries to the Julian calendar by using existing chronologies and by aligning astronomical observations with modern calculations. All together, they produced the standard editions of the diaries and made them accessible to many kinds of research. Since then, their work has been digitized and is now accessible through Oracc (http://oracc.museum.upenn.edu/adsd/).

For many purposes, this textual representation of the diaries is sufficient. For example, in studying the historical notes or the phrasing the astronomical observations, a researcher might only need access to the raw text of the diaries. This kind of textual analysis is useful, but it is largely restricted to studying the framing of the diaries while avoiding their content. Text is good at representing unstructured data, but the contents of the diaries are filled with structured information. Therefore, in order to effectively study the contents of the diaries, we need a systematic representation that reflects the structure of the diaries themselves. With such a systematic representation, it would be much easier to study the content of diaries as data, as opposed to studying the framing of the data.

The goal of this project is to develop a computational representation of the systematic information content of the astronomical diaries, with the purpose of bridging the gap between ancient and modern representations of data. The first step in this process is to design a format for representing the systematic information content of the diaries. This format would impose structure on the diaries that would make it possible to study them more systematically. The next step is to encode the diaries in this format. We will investigate text analysis, existing structured data collected from the diaries, and human curation. Third, once we have a computational representation of the diaries, we need tools for interpreting that data and making it accessible to modern analysis. This includes converting between Babylonian and modern units, dealing with the Babylonian calendar, as well as computing the real configurations of astronomical phenomena so they can be compared with

---

[1] Abraham J. Sachs and Hermann Hunger. *Astronomical Diaries and Related Texts from Babylonia.* Vol. 1. Verlag der Österreichischen Akademie der Wissenschaften, 1988, p. 11–12.

[2] John M. Steele. *A Brief Introduction to Astronomy in the Middle East.* Saqi, 2008, p. 41.

[3] Sachs and Hunger, *Astronomical Diaries and Related Texts from Babylonia.*

ancient observations. Finally, we will perform some simple computer analysis of the diaries as a demonstration of how a computational representation is useful. We will start by discussing how to systematize the astronomical diaries.

2 Systematic Representation of Observations

2.1 Scope and Systematic Observations

Before we talk about how we should systematize the data in the diaries, we should talk about what data we want to systematize. For the purposes of this project, we will restrict ourselves to *systematic observations*. This will include observations of the relative positions of celestial bodies, the prices of commodities in the market, historical notes, and so on, but it will not include the phrasing of those observations or choices of terminology, unless that phrasing or terminology carries systematic meaning. In the relatively rare cases where it is unclear whether some phrase carries systematic meaning, we will err on the side of caution and encode information about it.

Why should we restrict ourselves in this way? Why not encode information about phrasing and terminology? There are both practical and theoretical reasons for this decision. Practically, it requires more time and resources to encode the huge amount of phrasing information. Any observation might be phrased in dozens of ways, and encoding the details of exactly how it was phrased can drastically increase the amount of information being encoded. Even if the time and resources were available, it is unclear whether such a representation would be desirable. There is a tension between imposing useful structure and capturing details. The stricter and more rigid our representation is, the fewer details we can encode. For example, if we restricted ourselves to complete observations (that is, observations that have been preserved in their entirety), we would be unable to represent a large fraction of the useful data in the diaries. On the other hand, such a restriction would make the data easier to work with. Similarly, if we tried to encode every detail of every observation, we would need a very flexible representation that does not impose much structure. Our representation would essentially encode the same information as the raw text of the diaries, and it would not be much more useful for systematic analysis. We have to find a middle ground where we impose as much structure as possible while also preserving as many details as possible, by only encoding information that is naturally structured in the original text, that is, the systematic observations.

We will include provenance information that will make it possible to connect our systematic representation of observations with line numbers in the original text. That way, systematic and computational studies of phrasing and terminology will still be possible by accessing the raw text through Oracc.

2.2 Quantities and Metrology

There are many parts to an observation in the astronomical diaries. Each observation may include dates, quantities, distances, prices, references to celestial objects, and so on. Before we can represent an entire observation, we need to have a framework for representing the constituent parts. Consider this very typical observation from a tablet covering the year 345 BC.[4]

$$\underbrace{\text{Night of the 9th, [... first part of the night,]}}_{\text{date and time}} \underbrace{\text{Mars}}_{\text{object}} \text{ was } \underbrace{\text{1 cubit 8 fingers}}_{\text{distance}} \text{ below } \underbrace{\eta \text{ Tauri}}_{\text{object}}$$

---

[4]Sachs and Hunger, *Astronomical Diaries and Related Texts from Babylonia*, vol. 1, p. 145, Obv. 18–19. Also available on Oracc at `http://oracc.iaas.upenn.edu/adsd/adart1/X103460`.

Here we have dates, celestial objects, and quantities with units (in this case, a distance on the sky). We need to have a consistent and systematic way of representing each of these quantities before we can represent the whole observation.

As a general principle, we will try to represent quantities in a way similar to their Babylonian representations. That means we will use Babylonian units, dates, and so on. There will be code for converting between ancient and modern representations, so this allows us to keep detail without making the data much harder to work with.

In addition, many numbers appearing in the original text are written in base 60 or include fractions. Representing either of these with binary floating point numbers would often be an approximation. Thus, we will represent all numbers with fractions, so there will be no need to approximate. This also means that there will be no rounding error in the computations that we might perform on this data.

2.2.1 Celestial Objects

All celestial objects (the Moon, planets, stars, etc.) and signs of the zodiac will be represented using the entity system in the Wolfram Language (also known as Mathematica).[5] In the case of the example above, Mars would be represented `Entity["Planet","Mars"]` and η Tauri by `Entity["Star","Alcyone"]` ("Alcyone" is the canonical name used for η Tauri). This allows for simple interaction with outside programs as well as access to curated data about these objects.

2.2.2 Dates and Times

Dates will be represented simultaneously in the Babylonian and Julian calendars. Each date will contain the Babylonian regnal year, the Babylonian month (including whether it was intercalary), the Babylonian day, as well as the corresponding Julian year, month, and day. However, Julian and Babylonian days do not perfectly align. The Julian day starts at midnight, but the Babylonian day starts at sunset.[6] For our format, we will align days so that their daylight hours agree. This has a few advantages. First, it means that we maximize the hours of agreement between the calendars because there will only be a few hours (between sunset and midnight) when they do not agree. Second, it means that daylight observations (like the weather) will be recorded as occurring on the correct Julian day. Finally, this alignment makes it easier to use other chronologies, because most chronologies give the correspondence between Babylonian and Julian days based on their daylight hours.

It might seem like including the Julian date goes against our principle of always using the Babylonian representation. However, it is necessary because of the difficulty of converting between the two calendars. While some chronologies from this period have been digitized,[7] they do not perfectly align the the chronology of Sachs and Hunger, which was developed to align with the diaries.[8] Eventually, it will be possible to integrate with a unified, digital chronology like Shanati[9], but at the moment no such chronology exists.

Often, the time of an observation is given in addition to the date. The day and night are each divided into four parts. At night, there is the beginning part, the first part, the middle part, and

---

[5] https://www.wolfram.com/knowledgebase/

[6] Sachs and Hunger, *Astronomical Diaries and Related Texts from Babylonia*, p. 15.

[7] Richard A. Parker and Waldo H. Dubberstein. *Babylonian Chronology, 626 B.C. – A.D. 45.* University of Chicago Press, 1942. Available in digital form at https://www.staff.science.uu.nl/~gent0113/babylon/babycal.htm as well as http://baptiste.meles.free.fr/site/mesocalc.html#dates

[8] Sachs and Hunger, *Astronomical Diaries and Related Texts from Babylonia*, p. 15.

[9] http://shanati.org

the last part. During the following day, there is the morning, noon, the afternoon, and sunset. When representing a date and time, we will encode the relevant division of the day or night.

2.2.3 Celestial Distances

Many observations, like the example above, specify a distance on the sky. Celestial distance in the diaries are measured in cubits (KÙŠ) and fingers (SI or U).[10] The relationship between cubits and fingers has been subject to some debate. Neugebauer believed that, by the time of the diaries, both the relationships

$$1 \text{ cubit} = 24 \text{ fingers and } 1 \text{ cubit} = 30 \text{ fingers}$$

were in use.[11] However, Sachs and Hunger only give the relationship of 24 fingers per cubit.[12] Now with the work of Jones[13] and Steele,[14] it is mostly agreed that there are 24 fingers per cubit in the diaries.

There has also been debate on the relation between cubits and degrees. In mathematical texts, the value 1 cubit $= 2°$ is used.[15] However, statistical analysis of the diaries shows that in practice 1 cubit $\approx 2.27°$.[16]

With this in mind, we will represent distance on the sky with a pair of numbers, with the first representing the reported cubits and the second the reported fingers. For convenience, there will be code for converting this distance into a number of cubits using 1 cubit = 24 fingers. There will also be two functions for converting into degrees: one that uses the ideal 1 cubit $= 2°$ and another that uses the more realistic 1 cubit $= 2.27°$.

We will also support named distances like "a little", which do not come with an explicit numerical value.

2.2.4 Durations

Most durations in the diaries are recorded in degrees (UŠ) and NINDA, where[17]

$$1° \text{ (UŠ)} = 60 \text{ NINDA}$$

Because of this relationship, durations are often written as a base-60 number without reference to units. 1° (UŠ) corresponds to 4 minutes.

A few durations, however, also give the number of bēr, where 1 bēr $= 30°$ (making 1 bēr equivalent to 2 hours).[18] This notation is essentially only used in eclipse observations.

We will represent durations as a triplet of numbers, representing the bēr component, the degree (UŠ) component and the NINDA component. There will be code for combining all components to get the total degrees (UŠ), as well as code for converting the whole duration into minutes.

---

[10]Sachs and Hunger, *Astronomical Diaries and Related Texts from Babylonia*, p. 22.

[11]Otto Neugebauer. *Astronomical Cuneiform Texts*. Lund Humphries and Springer Science, 1955, p. 39.

[12]Sachs and Hunger, *Astronomical Diaries and Related Texts from Babylonia*, p. 22.

[13]Alexander Jones. "A Study of Babylonian Observations of Planets Near Normal Stars". In: *Archive for History of Exact Sciences* 58.6 (2004), pp. 475–536, p. 515.

[14]John M. Steele. "Planetary Latitudes in Babylonian Mathematical Astronomy". In: *Journal for the History of Astronomy* 34.3 (2003), pp. 269–289, p. 283–286.

[15]Ibid., p. 286.

[16]Jones, "A Study of Babylonian Observations of Planets Near Normal Stars", p. 520.

[17]Sachs and Hunger, *Astronomical Diaries and Related Texts from Babylonia*, p. 16.

[18]Otto Neugebauer and Abraham J. Sachs. "Some Atypical Astronomical Cuneiform Texts. I". in: *Journal of Cuneiform Studies* 21 (1967), pp. 183–218. URL: http://www.jstor.org/stable/1359371, p. 212. An example can be seen at line (r 3) of http://oracc.museum.upenn.edu/adsd/adart3/X301620

2.2.5 Culmination Times

In some observations (particularly of eclipses), the time of night is given relative to the culmination of a (ziqpu) star. Here is an example of such an observation.[19]

When (the point) $\underbrace{5°}_{\text{duration}}$ $\underbrace{\text{behind}}_{\text{direction}}$ $\underbrace{\mu \text{ Herculis}}_{\text{ziqpu star}}$ culminated, lunar eclipse

This phrase refers to the moment 5° (time degrees) after the culmination of the star μ Herculis, which is referred to as a ziqpu star. If it had said "in front of" instead of "behind", it would have have referred to 5° before the culmination of the ziqpu star.

We will represent times of this form with a duration (from the previous section), a direction "in front of" or "behind", and a culminating (ziqpu) star.

2.2.6 Capacities and Commodities

At the end of each month, the diaries record the amount of barley, dates, mustard, cress, sesame, and wool that can be purchased for 1 shekel of silver. Five out of six of these commodities have their capacities measured by volume. In particular, they are measured in kur, pān, sūt, and qa, where

$$1 \text{ kur} = 6 \text{ pān} = 36 \text{ sūt} = 180 \text{ qa}$$

and where 1 qa roughly corresponds with 1 liter.[20]

We will represent these capacities as four numbers corresponding to kur, pān, sūt, and qa.

2.3 Observations and Observation Types

The core of our format will be a list of observations. Each observation record will carry general metadata as well as a payload specific to its observation type. The general metadata will consist of the date of the observation, whether that observation was partially inferred (that is, whether it relies on information in square brackets that was imputed by Sachs and Hunger), a UUID (universally unique identifier, used to refer to observations), and provenance information. The provenance information will contain the creator of the record, the ID of the text in Oracc (for example, X103460), the line number corresponding to the start of the observation, whether this observation has been reviewed and checked, and any notes. Line numbers are represented by ascending integers counting lines in the Oracc editions of the diaries. A reference copy of the translations in Oracc have been included in order to protect against changes to the Oracc editions, as well as to annotate the lines with ascending integers. Finally, each observation will contain an observation type, and a type-specific payload. Let us now look at the payloads associated with each type.

2.3.1 Relative Positions

One of the most common types of observations in the diaries records the relative position of two celestial bodies, most often the position of the Moon or planets relative to a normal star. Here is an example of a typical such observation.[21]

---

[19]Line (o 8') of http://oracc.museum.upenn.edu/adsd/adart3/X301224
[20]Sachs and Hunger, *Astronomical Diaries and Related Texts from Babylonia*, p. 34.
[21]Line (o 11) of http://oracc.iaas.upenn.edu/adsd/adart1/X103460

the moon was 1 1/2 cubits above β Scorpii, the m[oon] being 8 fingers
back to the west

(object) (vertical distance) (vertical relation) (reference) (horizontal distance) (horizontal relation)

Essentially, such observations contain two distances corresponding to a vertical and horizontal displacement. It should be noted that the coordinate system of these observations (and even whether there was a consistent coordinate system), has been the subject of much debate.[22] However, it is clear that one of the parameters roughly corresponds with a vertical (perhaps latitudinal) displacement, while the other roughly corresponds with a horizontal (perhaps longitudinal) displacement. It should be noted that often only one coordinate is given, in which case it has been proposed that the other parameter was observed as being zero.

We will represent observations of relative positions by a pair of distances corresponding to the horizontal and vertical displacement, as well as the terms used for the relation (in front of, behind, east, and west for horizontal displacement, and above, below, north, and south for vertical displacement). When distances are not mentioned, they will be marked as such (see the section on missing data below).

2.3.2 Zodiacal Positions

While some positions are given relative to other celestial objects, others are given relative to signs of the zodiac. For example, it is often reported when a planet enters a sign of the zodiac, or which sign it was in when it first became visible above the horizon, etc. Sometimes it is also reported which part of the sign the object is in (usually the "beginning" or the "end").

Zodiacal positions will be represented with the object being observed, the sign of the zodiac containing that object, whether this is the first day when the object enters that sign of the zodiac, and the position of the object within the sign (as either "beginning" or "end").

2.3.3 Named Positions

Though much less common than observations of relative or zodiacal positions, the diaries contain some observations of objects standing in named regions of the sky. Here is an example of such an observation from the tablet for 650 BC.[23]

Mars became stationary in the area of the Lip of the Scorpion

(object) (named region)

Even though it is not always known where each named region is in the sky, we can still systematize these observations. If nothing else, systematizing these observations will allow us to easily check where these named regions are.

We represent each observation of this type with an object and the named region which contains the object.

2.3.4 Not Visible

At the end of each month, the diaries contain a summary of the movements of the planets. In particular, they record in which signs of the zodiac the planets stand at the end of the month, and in later tablets, they also record the days on which planets moved from one sign to another.

---

[22]Jones, "A Study of Babylonian Observations of Planets Near Normal Stars".

[23]Line (o i 10) of http://oracc.iaas.upenn.edu/adsd/adart1/X106510

However, when a planet is not visible at the end of the month, it is impossible to say which sign it is in. In these cases, the diaries note explicitly that a planet is "not visible".

This observation type contains an object, and represents an observation that that object was not visible on the day of the observation.

2.3.5 Solstices and Equinoxes

The diaries also contain observations of solstices and equinoxes. Generally, only the day of the event is recorded. Sometimes, however, it is also recorded that "I did not watch". This pattern appears in a number of observation types, particularly when it was possible to predict or interpolate when an event would happen. It is believed that almost all of the solstices and equinoxes were computed instead of observed, so almost every observation of this type is marked with "I did not watch".[24]

2.3.6 Greek-Letter Phenomena

The diaries contain observations of certain key points in the movement of the planets (and sometimes Sirius). In particular, first appearances, last appearances, stationary points, and acronychal risings. In the case of inner planets, they separately observed first and last appearances in the east and west. These are often referred to as the Greek-letter phenomena because Neugebauer assigned each a greek letter in order to avoid "clumsy terminology".[25]

Let us look at an example. Here is an observation of the first appearance of Jupiter.[26]

$$\underbrace{\text{The 12th}}_{\text{date}}, \underbrace{\text{Jupiter's}}_{\text{object}} \overbrace{\text{first appearance in Pisces}}^{\text{position (sub-observation)}}; \text{it was bright (and) high, rising of Jupiter}$$

$$\text{to sunrise:} \quad \underbrace{10° \; 30'}_{\text{visibility duration}} \quad ; \text{(ideal) first appearance on the} \quad \underbrace{\text{11th}}_{\text{ideal date}}.$$

Here we see a few new structures. First, we have mention of an ideal state separate from the observed one. This indicates that Jupiter was first seen on the 12th, but its visibility duration implied that it was already visible on the 11th, although it was not directly observed.

We also have a sub-observation, where one observation references another. That is, within this observation of the first appearance, we also have a reference to an observation of the zodiacal position. We could model this by adding a field to first appearance observations which records the sign of the zodiac. However, this has several disadvantages. First, the position is not always given with the sign of the zodiac. With stationary points, for example, the position is often given relative to a normal star.[27] By allowing for sub-observations, we are able to use different position specifications while still being easy to work with. Second, it avoids redundant observations. For example, if we wanted to study all observations of zodiacal positions, but many references to zodiacal positions were inside of first appearance observations, we would have to extract data from both observation types. By having a single zodiacal position observation and allowing the first appearance observation to reference it, we avoid this problem.

Thus, we will represent first and last appearances with the celestial object, the duration of visibility, the ideal date, a sub-observation of the position (represented by its UUID), and a flag for "I did not watch". Stationary points are the same as first and last appearances but without

---

[24] Sachs and Hunger, *Astronomical Diaries and Related Texts from Babylonia*, p. 26.

[25] Neugebauer, *Astronomical Cuneiform Texts*, p. 280–281.

[26] Line (r 9') of http://oracc.iaas.upenn.edu/adsd/adart2/X201731

[27] Example on line (r 12') of http://oracc.iaas.upenn.edu/adsd/adart2/X201783

the duration of visibility, and acronychal risings are the same as stationary points but without the sub-observation of position.

2.3.7 Eclipses

Eclipses are the most complicated observations regularly recorded in the diaries with over a dozen parameters that can be given for each observation. We will use a list of 17 parameters that are mostly derived from the list compiled by Sachs and Hunger.[28] This includes the eclipse type (lunar or solar), whether it was observed or predicted, sub-observation of the duration from moonrise to sunset, sub-observation of the duration from sunrise to moonset, the start time relative to a culminating star, the entrance direction, the duration between the start of the eclipse and its maximal phase, the magnitude of the maximal phase, the duration of the maximal phase, the duration between the end of the maximal phase and the end of the eclipse, the direction in which the shadow moved, the total duration of the eclipse, a list of weather-related sub-observations, a list of object visibility sub-observations, a sub-observation of the lunar position, the duration from sunset to the start of the eclipse, and finally the duration from the start of the eclipse to sunrise.

2.3.8 Lunar Six Phenomena

The Lunar Six are a group of six observation types measuring the duration between the appearance and disappearance of the sun in relation to the appearance and disappearance of the Moon. In particular, they are the durations from sunset to moonset, from moonset to sunrise, from sunrise to moonset, from moonrise to sunset, from sunset to moonrise, and from moonrise to sunrise.

We will represent each of these with a duration (in time degrees), and a few flags for "I did not watch" and "measured".

2.3.9 Wind Directions

One of the most common observation types recorded in the diaries is of the wind direction. These follow a very simple format, just giving the direction of the wind as north, south, east, or west.

We will represent observations of wind direction with a single parameters indicating the direction of the wind.

2.3.10 Generalized Weather

In addition to the wind direction, the diaries also contain many other observations of weather. For example, some days are observed to be "cloudy", and on others it was "very overcast".

We will represent each such observation with a tag indicating the type of weather. We use a list of tags derived from the list provided by Sachs and Hunger.[29]

2.3.11 Market Rates

As mentioned previously, the market rates for various commodities were given at the end of each month. We represent these observations by the capacities of barley, dates, mustard, cress, and sesame, as well as the weight of wool in minas, that can be purchased for 1 shekel of silver in the market.

---

[28]Sachs and Hunger, *Astronomical Diaries and Related Texts from Babylonia*, p. 23–24.
[29]Ibid., p. 27–34.

2.3.12 River Levels

At the end of each month, the level of the Euphrates river was often recorded.[30] There were three formats for such observations.

First, the earliest tablets record when the river switched from rising, to falling, to standing still. We can see an example of this in the second observation of the earliest preserved tablet.[31]

$$\underbrace{\text{Night of the 1st}}_{\text{date}}, \underbrace{\text{the river level rose}}_{\text{direction}}.$$

We will represent these observations with the direction of the river level change as rising, falling, or standing still.

In many later tablets, this observation type was replaced by one where the numerical change in river level was reported over an explicit interval. Take this observation from 371 BC.[32]

$$\text{From} \underbrace{\text{the 11th}}_{\text{start date}} \text{until} \underbrace{\text{the 22nd}}_{\text{observation date}}, \text{the river level - remainder} \underbrace{\text{3 cubits}}_{\text{remainder}} - \underbrace{\text{rose}}_{\text{direction}} \underbrace{\text{4 2/3 cubits}}_{\text{distance}}.$$

Here we have a start and end date for the river rising, as well as a numerical record of how much it rose. We are also given the "remainder" (bābtu), which seems to give some absolute measure of the river level. However, Sachs and Hunger found no obvious relationship between the remainder and the reported change in river level, making it unclear what it represents.[33] We will represent these observations by their start date, end date, the direction, the distance, and the remainder.

Finally, many later tablets report the absolute river level in terms of the *na* (translated as the "*na* gauge" by Sachs and Hunger). The numbers on the *na* align with the reported changes in river level so that 1 unit on the *na* corresponds with 4 fingers. We will represent these observations simply with the value of the *na* (gauge).

2.3.13 Historical Events

Throughout the diaries, although most often at the end of the month, major historical events are reported. For example, on the diary for 331 BC, we have an account of the battle of Gaugamela.[34][35]

> That month, on the 11th, panic [occurred] in the camp before the king [...]
>
> lay? opposite the king. On the 24th, in the morning, the king of the world [...] the standard? [...]
>
> they fought with each other, and a heavy? defeat of the troops of [...]
>
> the troops of the king deserted him and [went?] to their cities [...]
>
> they fled to the land of the Guti [...]
>
> ...
>
> That month, from the 1st to [...]
>
> came to Babylon saying: "Esangila [...]"

---

[30] Sachs and Hunger, *Astronomical Diaries and Related Texts from Babylonia*, p. 34–36.

[31] Line (o i 2) of http://oracc.iaas.upenn.edu/adsd/adart1/X106510

[32] Line (o i 14) of http://oracc.iaas.upenn.edu/adsd/adart1/X103721

[33] Sachs and Hunger, *Astronomical Diaries and Related Texts from Babylonia*, p. 35.

[34] Lines (o 14') through (r 15') of http://oracc.museum.upenn.edu/adsd/adart1/X103301

[35] Sachs and Hunger, *Astronomical Diaries and Related Texts from Babylonia*, p. 179 notes that Alexander is the subject of this account.

and the Babylonians for the property of Esangila [...]

On the 11th, in Sippar an order of Al[exander ...]

["...] I shall not enter your houses". On the 13th, [...]

[...] to$^?$ the outer gate of Esangila and [...]

[...] On the 14th, these$^?$ Ionians a bull [...]

short, fatty tissue [...]

[...] Alexander, king of the world, [came$^?$ in]to Babylon [...]

[... hor]ses and equipment of [...]

[...] and the Babylonians and the people of [...]

[...] a message to [...]

The unstructured nature of these historical notes means that there is not much that can be represented systematically. Therefore, the only parameter specific to this observation type will be a list of general tags like "War" or "ReligiousCeremony". Otherwise, text analysis of historical notes will be possible by using the provenance information to get the raw text from Oracc.

2.3.14 Errors

The diaries occasionally contain errors which make observations impossible to represent systematically. For example, consider this observation of the position of the Moon relative to γ Capricorni.[36]

the moon was $\underbrace{\text{1 cubit above}}_{\text{vertical distance}}$ γ Capricorni, the moon being $\underbrace{\text{1/2 cubit [low] to the south}}_{\text{vertical distance again}}$

Here, the vertical distance between the Moon and γ Capricorni is specified in two conflicting ways. The writer of this tablet probably miswrote a vertical relation when they intended a horizontal one. However, our representation of relative positions can only contain at most one vertical distance, so we cannot systematically represent this observation with our format.

We represent such errors with the type of observation that was incorrectly recorded (if it can be inferred), and any notes about the error.

2.3.15 Other

It is inevitable that there will be some systematic information which will not representable in our format (at least initially). We want to be able to mark such observations so that if the format is updated, we can easily find observations that were previously skipped and which can now be represented.

With this in mind, we mark such observations with a list of tags and any notes.

2.4 Missing data

Given what is preserved of the astronomical diaries, we need methods for representing missing data. There are a few ways in which data can be missing, which we will encode as "destroyed" and "not mentioned". "Destroyed" means that the section of the tablet that might contain this data has been lost, and we cannot know whether it was mentioned or not. For example, if we saw

---

[36]Line number (o B23') of http://oracc.iaas.upenn.edu/adsd/adart3/X301051

an observation that "the moon was 1 cubit [...]", we would know that there was an observation of the relative position of the moon, but we can say nothing about what the reference object was, or the orientation of the relation between the two. "Not mentioned" means that the observation is preserved, but the data in question is not mentioned. For example, consider this observation of the relative position of the Mars and α Leonis.[37]

$$\underbrace{\text{Mars}}_{\text{object}} \text{ was } \underbrace{\text{8 fingers above}}_{\text{vertical relation}} \underbrace{\alpha \text{ Leonis}}_{\text{reference}}. \text{ Night of the 18th ...}$$

The entire observation is preserved, but there is no mention of the horizontal relationship between Mars and α Leonis. This example demonstrates the importance of the distinction between these two kinds of missing information. The explicit lack of a horizontal displacement might be interpreted to mean that Mars was directly above α Leonis, meaning that the observed horizontal displacement was zero. On the other hand, if the observation read "Mars was 8 fingers above α Leonis [...]", we could say nothing about the observed horizontal displacement. It may have been reported with some value, or it may have not been mentioned at all. Thus, whenever data is missing, it will be marked as either destroyed or not mentioned.

2.5 Summary

This concludes our discussion of how we will represent the systematic information contained in the diaries. To summarize, we represent the complete dataset as a list of observations. Each observation contains provenance information, as well as a payload that is specific to its observation type. There are 30 supported observation types, which each have their own specialized formats which have been designed to resemble the formats of these observations in the original text.

See appendix 1 for the formal specification of the format.

3 Data Curation

Now that we have a format for systematically representing the diaries, we should discuss the practical problem of getting data into this format. There are three avenues that we will consider: text analysis, repurposing existing projects, and human curation.

3.1 Text Analysis

In many ways, the diaries seem well suited for automated text analysis. Most observations follow a very formulaic structure, and the organization of the diaries is relatively simple. However, there are some difficulties that make large-scale text analysis impractical. First, while some observation types are phrased formulaically (like the wind direction), others can be much more complicated (like eclipses). In addition, even when it is possible to parse an observation, context like the day and the month are needed to put that observation into our format.

Text analysis on the diaries can be useful for quick statistics and for studying phrasing and terminology, but it is not a good way to systematize the full contents of the diaries.

For an example of some simple and practical text analysis, see appendix 2.

---

[37]Line (o 5) of http://oracc.iaas.upenn.edu/adsd/adart1/X102613

3.2 Existing Data

There have been a several projects for which some data from the diaries has already been curated. In particular, Jones,[38] and Graßhoff and Fischer[39] both curated fairly large datasets which they have made publicly available. It is not too difficult to approximately put these datasets into our format, but there is information that our format requires that is often not provided. For example, we record the order of the vertical and horizontal relations when encoding observations of relative celestial positions, but Jones does not. Similarly, we make a distinction between data that is missing because it is unmentioned and data that is missing because it is destroyed, but neither Jones nor Graßhoff encodes this in their dataset.

There are also questions of quality control when using these datasets. In particular, the Graßhoff dataset contains a number of errors, like the miscalculation of the Modified Julian Day, or the recording of many observations in an intercalary month during a year with no intercalation, and so on.

3.3 Human Curation

Ultimately, the only way to reliably convert the raw text of the diaries into this format would be through human curation. This would mean going through every line of every diary and recording any systematic observations. However, we can develop tools that make this process quicker. In particular, we can use a simplified format which has been designed for inputting data, and which can be automatically converted to the full format.

In order to test the input format, I used it to input all of diary -170A.[40] This process took about 1 hour and produced records of 69 observations. Given that there are about 12000 lines in all of the diaries,[41] and there are about 60 lines in -170A, we can naively compute that it would take about 200 hours to input all of the diaries. This number could probably be reduced with an improved input interface, however, the realities of data curation could also cause it to take considerably longer.

4 Conclusion and Future Work

The astronomical diaries are one of the richest bodies of systematic data from antiquity. The goal of this project is to expose that systematic data, and to make it computable. At the moment, the diaries are only available in unstructured form, despite the fact that they have always represented structured data. With the computerization of the diaries, it would become much easier to study them for what they are: systematic data.

The primary avenues for future work lie in optimizing the curation process, and further developing the support programs that make analysis possible. In the case of the curation process, the input format can likely be simplified, reducing the time required to input data. For the support programs, many more could be implemented, including code for computing the ideal times of Greek-letter phenomena, aggregating river level observations, and so on.

All code for this project is available in the public git repository at `https://github.com/chriswolfram/ComputationalDiaries`.

---

[38]Originally available at `http://www.chass.utoronto.ca/~ajones/normal_stars/`, although now only available through the Internet Archive at `https://web.archive.org/web/20070808121203/http://www.chass.utoronto.ca/~ajones/normal_stars/`

[39]Available on Edition Topoi at `http://doi.org/10.17171/1-3-383`

[40]Full text available at `http://oracc.iaas.upenn.edu/adsd/adart2/X201701`

[41]See appendix 2.

References

Jones, Alexander. "A Study of Babylonian Observations of Planets Near Normal Stars". In: *Archive for History of Exact Sciences* 58.6 (2004), pp. 475–536.

Neugebauer, Otto. *Astronomical Cuneiform Texts*. Lund Humphries and Springer Science, 1955.

Neugebauer, Otto and Abraham J. Sachs. "Some Atypical Astronomical Cuneiform Texts. I". In: *Journal of Cuneiform Studies* 21 (1967), pp. 183–218. URL: http://www.jstor.org/stable/1359371.

Parker, Richard A. and Waldo H. Dubberstein. *Babylonian Chronology, 626 B.C. – A.D. 45*. University of Chicago Press, 1942.

Sachs, Abraham J. and Hermann Hunger. *Astronomical Diaries and Related Texts from Babylonia.* Vol. 1. Verlag der Österreichischen Akademie der Wissenschaften, 1988.

Steele, John M. *A Brief Introduction to Astronomy in the Middle East*. Saqi, 2008.

— "Planetary Latitudes in Babylonian Mathematical Astronomy". In: *Journal for the History of Astronomy* 34.3 (2003), pp. 269–289.

Appendix 2: Text Analysis

First, we use the reference copy of Oracc that is included in the ComputationalDiaries package to get the list of all lines that appear in the diaries:

```
fullLines = DeleteCases[Catenate[getTabletLines[]], Delimiter][[All, 3]];
lines = StringDelete[fullLines, "[" | "]" | "?"];
```

Basic Statistics

There are about 12 000 lines in total:

```
Length[lines]
```

12 021

We can look at the distribution of line lengths:

```
Histogram[StringLength[lines], {10}, ··· + ]
```
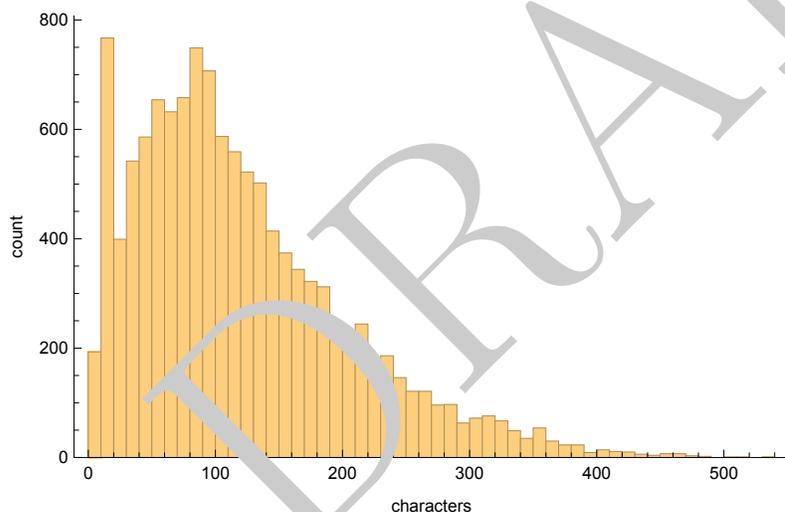

Wind Direction

We can write a simple program to extract mentions of the wind direction, and to count how often the wind blew in different directions:

```
ReverseSort[Counts[Flatten[StringCases[lines,
    "the " ~~ d : Shortest[Except[WhitespaceCharacter] ..] ~~ " wind blew" :> d]]]]
```

⟨| north → 1413, south → 363, west → 58, east → 47, ... → 1 |⟩

We can see that the north wind blew most of the time.

Relative Position

We can easily see how often different relations were applied to different celestial bodies:

```
rels = AssociationMap[
  ReverseSort[Counts[Flatten[StringCases[lines,
      # ~~ " was " ~~ Shortest[___] ~~
        rel : ("above" | "below" | "in front of" | "behind") ⧴ rel]]]] &,
  {"moon", "Mercury", "Venus", "Mars", "Jupiter", "Saturn"}]
```

⟨|moon → ⟨|in front of → 1501, behind → 1305, below → 895, above → 489|⟩,
 Mercury → ⟨|above → 63, below → 44, in front of → 3, behind → 3|⟩,
 Venus → ⟨|below → 227, above → 189, behind → 5, in front of → 4|⟩,
 Mars → ⟨|above → 110, below → 102, behind → 6, in front of → 2|⟩,
 Jupiter → ⟨|above → 37, below → 21, behind → 4, in front of → 2|⟩,
 Saturn → ⟨|above → 13, below → 13, in front of → 3, behind → 1|⟩|⟩

```
BarChart[Lookup[#, {"above", "below", "in front of", "behind"}, 0], ChartLabels →
    (Rotate[#, Pi / 2] & /@ {"above", "below", "in front of", "behind"})] & /@ rels
```
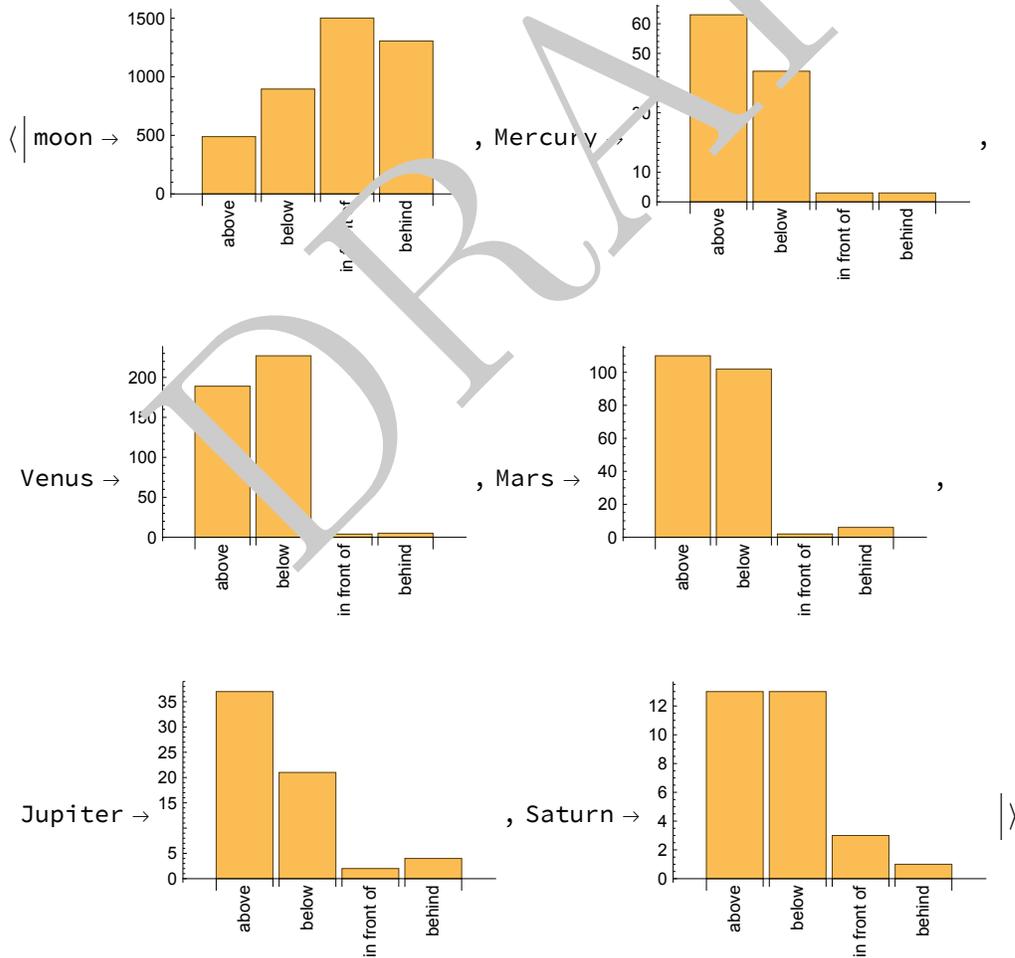
Horizontal relations are observed much more often for the Moon, but the Moon also just has many more observations.

Cubits and Fingers

Find all mentions of *c* cubits and *f* fingers:

```
cubitsFingers = ReverseSort[Counts[DeleteCases[Catenate[StringCases[lines,
      Longest[(c : (DigitCharacter ..) ~~ " cubit " | " cubits ") | ""] ~~
       Longest[(f : (DigitCharacter ..) ~~ " finger" | " fingers") | ""] :>
      ToExpression /@ ({c, f} /. "" → 0)]], {0, 0}]]]
```

⟨| {2, 0} → 3304, {3, 0} → 957, {1, 0} → 917, {4, 0} → 240, {0, 4} → 201, {0, 8} → 199,
 {5, 0} → 136, {6, 0} → 119, {0, 2} → 106, {1, 8} → 101, {0, 20} → 85, {0, 6} → 85,
 {1, 4} → 52, {0, 1} → 40, {2, 8} → 40, {0, 10} → 31, {1, 4} → 29, {   3} → 22,
 {0, 5} → 16, {0, 14} → 15, {1, 20} → 15, {2, 4} → 12,   0} → 10, {3, 8} → 5,
 {2, 6} → 4, {0, 18} → 4, {3, 6} → 3, {2, 20} → 3, {0, 22} → 2, {3, 4} → 1
 {1, 10} → 1, {1, 18} → 1, {1, 5} → 1, {3, 20} →   {1, 2} → 1, {8, 0} → 1 |⟩

Find the total number of cubit and finger measurements:

```
Total[cubitsFingers]
```

6760

Look at the frequencies of different combinations of cubits and fingers:

```
Column[KeyValueMap[StringTemplate[
  "<*#1[[1]]*> cubits and <*#1[[2]]*> fingers: `2` occurrences"], cubitsFingers]]
```

```
2 cubits and 0 fingers: 3304 occurrences
3 cubits and 0 fingers: 957 occurrences
1 cubits and 0 fingers: 917 occurrences
4 cubits and 0 fingers: 240 occurrences
0 cubits and 4 fingers: 201 occurrences
0 cubits and 8 fingers: 199 occurrences
5 cubits and 0 fingers: 136 occurrences
6 cubits and 0 fingers: 119 occurrences
0 cubits and 2 fingers: 106 occurrences
1 cubits and 8 fingers: 101 occurrences
0 cubits and 20 fingers: 85 occurrences
0 cubits and 6 fingers: 85 occurrences
1 cubits and 4 fingers: 52 occurrences
0 cubits and 1 fingers: 40 occurrences
2 cubits and 8 fingers: 40 occurrences
0 cubits and 10 fingers: 31 occurrences
1 cubits and 6 fingers: 29 occurrences
0 cubits and 3 fingers: 22 occurrences
0 cubits and 5 fingers: 16 occurrences
0 cubits and 14 fingers: 15 occurrences
1 cubits and 20 fingers: 15 occurrences
2 cubits and 4 fingers: 12 occurrences
7 cubits and 0 fingers: 10 occurrences
3 cubits and 8 fingers: 5 occurrences
2 cubits and 6 fingers: 4 occurrences
0 cubits and 18 fingers: 4 occurrences
3 cubits and 6 fingers: 3 occurrences
2 cubits and 20 fingers: 3 occurrences
0 cubits and 22 fingers: 2 occurrences
3 cubits and 4 fingers: 1 occurrence
1 cubits and 10 fingers: 1 occurrences
1 cubits and 18 fingers: 1 occurrences
1 cubits and 5 fingers: 1 occurrences
3 cubits and 20 fingers: 1 occurrences
1 cubits and 2 fingers: 1 occurrences
8 cubits and 0 fingers: 1 occurrences
```

Illustrate some simple evidence for there being 24 fingers in a cubit by getting the number of occurrences where the number of fingers divides 24:

```
fingerFrequencies =
  ReverseSort[Merge[MapAt[Last, Normal[cubitsFingers], {All, 1}], Total]];
```

```
Total[KeySelect[fingerFrequencies, # =!= 0 && Divisible[24, #] &]]
```

```
901
```

And compare with the number of occurrences where the number of fingers divides 30:

```
Total[KeySelect[fingerFrequencies, # =!= 0 && Divisible[30, #] &]]
```

```
339
```

Also look at the simpler fractions you get when dividing the number of fingers by 24:

(This is the map of [fingers divided by 24]→[# of occurrences])

```
KeyMap[#/24 &, fingerFrequencies]
```

$$\left\langle \left| 0 \to 5684, \frac{1}{3} \to 345, \frac{1}{6} \to 266, \frac{1}{4} \to 121, \frac{1}{12} \to 107, \frac{5}{6} \to 104, \right.\right.$$

$$\left.\left. \frac{1}{24} \to 40, \frac{5}{12} \to 32, \frac{1}{8} \to 22, \frac{5}{24} \to 17, \frac{7}{12} \to 15, \frac{3}{4} \to 5, \frac{11}{12} \to 2 \right| \right\rangle$$

As opposed to when dividing the number of fingers by 30:

(This is the map of [fingers divided by 30]→[# of occurrences])

```
KeyMap[#/30 &, fingerFrequencies]
```

$$\left\langle \left| 0 \to 5684, \frac{4}{15} \to 345, \frac{2}{15} \to 266, \frac{1}{5} \to 121, \frac{1}{15} \to 107, \right.\right.$$

$$\left.\left. \frac{2}{3} \to 104, \frac{1}{30} \to 40, \frac{1}{3} \to 32, \frac{1}{10} \to 22, \frac{1}{6} \to 17, \frac{7}{15} \to 15, \frac{2}{5} \to 5, \frac{11}{15} \to 2 \right| \right\rangle$$

Appendix 3: Basic Analysis

Here are a few examples of very basic analysis that can be done with computational editions of the diaries. The analysis here is heavily restricted to the nature of the data available, as existing data only covers relative positions of celestial objects. More analysis is certainly possible with this data, but this was just meant as a simple demonstration of the ease of performing analysis.

Load the ComputationalDiaries package

```
PacletDirectoryLoad["/Users/christopher/git/ComputationalDiaries/packages/"];

Get["ComputationalDiaries`"]
```

Load data from Graßhoff and Jones

```
dataDirectory = "/Users/christopher/git/ComputationalDiaries/data";

grasshoffObservations =
   Import[FileNameJoin[{dataDirectory, "MX", "grasshoffObservations.mx"}]];
jonesObservations = Import[FileNameJoin[
      {dataDirectory, "MX", "jonesObservations.mx"}]];
```

Visualize the relative positions of the Moon and planets to normal stars

With Graßhoff data:

```
Graphics[{Opacity[0.3], Line[DeleteMissing[Function[obv, {
         astronomicalPosition[obv["ObjectA"], obv["Date"]],
         astronomicalPosition[obv["ObjectB"], obv["Date"]]
      }] /@ Select[grasshoffObservations, #["Type"] === "RelativePosition" &],
    1, 2]]},
  PlotRange → {{0, 360}, {-15, 15}}, ImageSize → {Automatic, 200}]
```
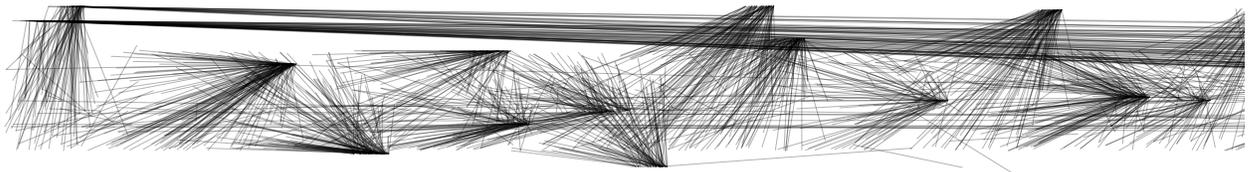

With Jones data:

```
Graphics[{Opacity[0.3], Line[DeleteMissing[Function[obv, {
        astronomicalPosition[obv["ObjectA"], obv["Date"]],
        astronomicalPosition[obv["ObjectB"], obv["Date"]]
      }] /@ Select[jonesObservations, #["Type"] === "RelativePosition" &], 1, 2]]},
 PlotRange → {{0, 360}, {-15, 15}}, ImageSize → {Automatic, 200}]
```
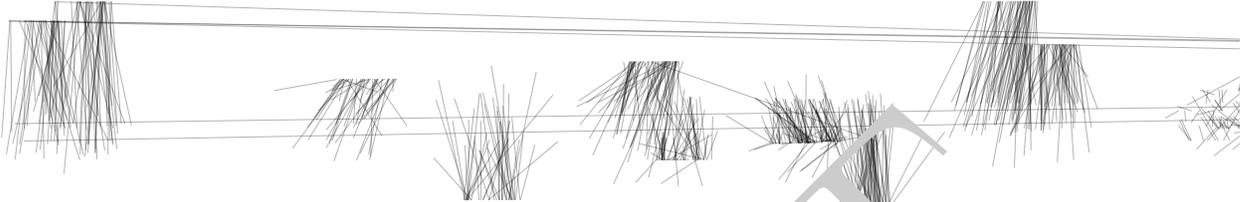

Compute the longitudes for which different normal stars are used

Using Graßhoff data:

```
Histogram[KeyValueMap[Tooltip[#2, #1] &]@Merge[Function[obv,
     obv["ObjectB"] → astronomicalPosition[obv["ObjectA"], obv["Date"]][[1]]
   ] /@ Select[grasshoffObservations, #["Type"] === "RelativePosition" &],
  Identity], {1}, AspectRatio → Automatic]
```


Using Jones data:

```
Histogram[KeyValueMap[Tooltip[#2, #1] &]@Merge[Function[obv,
     obv["ObjectB"] → astronomicalPosition[obv["ObjectA"], obv["Date"]][[1]]
   ] /@ Select[jonesObservations, #["Type"] === "RelativePosition" &], Identity],
 {1}, AspectRatio → Automatic]
```