Tracing the evolution of personality cognition in early human civilisations: A computational analysis of the Gilgamesh epic

Amy He Du[1], Johannes A. Karl[1,2] , Velichko Fetvadjiev[3,4], Markus Luczak-Roesch[5], Reinhard Pirngruber[6] and Ronald Fischer[1,7]

Abstract

Assessing evolution of cognitive structures across historical periods has remained challenging in the absence of direct access to humans from the past. Overcoming some of these challenges, we examined shifts in the implicit cognitive structures in the Epic of Gilgamesh, which is one of the earliest surviving pieces of literature, circulating in various versions over a period of approx. 2000 years in ancient Mesopotamia. Using a canonical English translation, we applied natural language processing (NLP) and human coding to extract low-dimensional representations of the implicit personality structure in three different historical epochs. We found systematic shifts over time with increasing complexity and increasing resemblance of contemporary personality models in later periods. We discuss how lexical analyses of ancient texts using trait co-occurrence analyses can provide novel insights on the evolution of human behaviour of relevance for contemporary social and behavioural science and the study of ancient societies.

Stories open windows into the minds of the distant past, describing salient features of human behaviour that engage the imagination of listeners and revealing what was considered worth passing on. By analysing how key characters in stories are described, we can unpackage what was deemed relevant to convey about a person and in turn, gain rich insights into cognition of past societies. Today, analyses of life stories are the realm of personality science (McAdams et al., 2004), yet personality judgements have been of interest for people throughout history (Diggle & Theophrastus, 2007; Mayer et al., 2011). An important question is whether such person descriptions follow the same logic, that is, whether they show comparable dimensionality across time and place. Although a five-factor model of personality had been assumed to be an evolutionarily stable feature (Buss, 1991, 1997), recent evidence suggests that person descriptions may be shaped by the cultural and ecological conditions (Durkee et al., 2022; Fischer, 2017; Fischer & Karl, 2021). Hence, person descriptions are likely to evolve over historical periods and it may be possible to trace evolutionary patterns across human history.

Nevertheless, attempts to capture change in personality descriptions are complicated because peculiarities of specific stories, texts or formats might obscure potential insights into the structure and changes in underlying cognition. We present an analysis of the Epic of Gilgamesh, which is one of the earliest bodies of literary texts in human history that has survived nearly completely till today. Importantly, for our purposes, recorded versions of the story circulated and evolved over approximately 2000 years in Ancient Mesopotamia with the first documented scripts from around 2200 BC and the last scripts being dated to 100 BC. Having a central story with a common set of characters and key shared elements, it becomes possible to analyse temporal change and to trace developments of cognitive structures through the lens of person descriptions. Using automated text-mining tools and English translations of the key source material by a

[1]School of Psychology, Victoria University of Wellington, Wellington, New Zealand
[2]School of Psychology, Dublin City University, Dublin, Ireland
[3]Department of Social Psychology, University of Amsterdam, Amsterdam, The Netherlands
[4]WorkWell Research Unit, North-West University, Potchefstroom, South Africa
[5]School of Information Management, Wellington School of Business and Government, Victoria University of Wellington, Wellington, New Zealand
[6]Independent Researcher, University Wien, Vienna, Austria
[7]Neuroscience and Neuroinformatics Research Group, Institute D'Or for Research and Education, Rio de Janeiro, Brazil

Corresponding author:
Ronald Fischer, Neuroscience and Neuroinformatics Research Group, Institute D'Or for Research and Teaching, Rua Diniz Cordeiro, 30 - Botafogo, Rio de Janeiro - RJ, 22281-100, Brazil.
Email: ronald.fischer@idor.org

single translator, we start to gain insights into what was deemed worth recording in person descriptions. Our study therefore provides snapshots of the minds of the people at the beginning of the first human civilisations.

Our work builds on the long-standing psycho-lexical hypothesis first sketched by Francis Galton, which proposes that all important personality traits are represented in language as single terms (Goldberg, 1981). This idea has inspired factor-analytical studies of self- and other-ratings of words that resulted in the description of major dimensions of personality (Ashton et al., 2004; Cheung et al., 2008; Goldberg, 1990; Nel et al., 2012). There is general consensus that the Big Five model (including Extraversion, Agreeableness, Conscientiousness, Neuroticism and Openness) captures the major dimensions of personality at least in English and other Germanic languages (Saucier & Goldberg, 2001), although various extensions have been proposed, notably adding the honesty–humility factor in the HEXACO model (Ashton & Lee, 2007, 2020). At the same time, observed personality structure may not be independent of their socio-cultural context. Studies of Chinese (Cheung et al., 2008) and indigenous languages in South Africa (Nel et al., 2012) revealed additional factors that emphasised the social-relational aspects of personality in more group-oriented contexts, but there is also evidence pointing to fewer stable factors (De Raad et al., 2010, 2014; Saucier et al., 2014) as well as substantively different structures in some non-Western languages (Fischer, 2021; Gurven et al., 2013; Singh et al., 2013). Of the contemporary languages that have been studied in psycho-lexical research, Arabic is the closest to Akkadian, the language used for the most complete version of the epic that is available today (whereas Sumerian, the language of the earliest version of the epic, is a language isolate and thus does not have a documented modern cognate). Research on personality concepts in modern Arabic has found a structure similar to the Big Five model but with some additional elements of morality and unconventionality (Zeinoun et al., 2017, 2018).

One potential driver underlying the degree of complexity of person descriptions are socio-economic conditions leading to niche construction effects for personality (Laland et al., 2000; Odling-Smee et al., 2003): in more complex and larger societies, individuals need to differentiate relevant characteristics of a larger number of individuals as they fulfil important roles within a society; this requires finer distinctions aligned with more diverse social roles. These cognitive demands are accompanied by more clearly differentiated behavioural clusters across different niches and together these dynamics lead to more clearly differentiated cognitive representations of individual differences (Fischer, 2017; Lukaszewski et al., 2017; Smaldino et al., 2019).

Niche construction effects in the personality domain raise interesting questions about our evolutionary past. Can we identify salient dimensions in the minds of ancient people? Is it possible to trace cognitive differentiation of implicit personality models that individuals may have held over historical time periods? Descriptions of heroes in epic narratives may offer evidence from the distant past of cognitive changes during early civilisations. Descriptions of characters in cultural stories had to be based on a shared understanding of human behaviour, otherwise, these descriptions would not have been intelligible to their audiences. Based on this assumption, analyses of the occurrences and co-occurrences of descriptors of personality can reveal what the storytellers implicitly believed about which characteristics tend to go together and which do not in order to engage and capture the audience (Rosenberg & Jones, 1972). Since no writing is ever produced in isolation of context, the storytellers' implicit concepts of personality reflect the broader historical and cultural context in which the author was situated (Hamilton, 2003; Porter, 2014).

Previous studies have demonstrated the possibility to extract personality information from ancient texts, using established personality models such as the Big Five (e.g. Passakos & de Raad, 2009). However, studying a text using pre-defined coding categories, behavioural scholars are unable to evaluate the historical developments of personality cognitions. Scholars in historical-philological disciplines examining these ancient texts, on the other hand, aim in the first place at providing accurate translations. They are further interested in the redaction history of single texts (e.g. George, 2003; Sallaberger, 2008; Tigay, 2002), and in the development of literary traditions in a given culture (Beckman et al., 2011; Foster, 2009) as well as the transfer of literary motives across cultures (West, 1999). What little there is of research into the social contexts of literature in Sumerian and Akkadian language has been often confined to rather technical aspects, such as the transmission of literary texts in the context of scribal education during the Old Babylonian period (Tinney, 1999), or to the political configurations to which these texts respond (Delnero, 2016; Konstantopoulos, 2017). However, there has been hardly any research attempting to provide insights into the possible cognitive models of those who produced the texts. To uncover cognitive structure, a bottom-up approach is needed in which co-associations of lexical units are examined and no equivalence of contemporary meaning is imposed, but the patterns can nonetheless be interpreted based on contemporary theories of the human mind and social conditions (Berry, 1989). We suggest that a focus on the character descriptions, especially the co-association of trait-relevant information within narratives, may provide novel insights into possible implicit theories about characters and, ultimately, individuals and social actors (Rosenberg & Jones, 1972). We believe this shift in analysis utilising text-analytical quantitative tools combined with a qualitative, historically informed analysis of the textual remains of Ancient Mesopotamia can offer rich new insights both for scholars of antiquity and for psychologists interested in personality and the evolution of human societies.

The Epic of Gilgamesh provides a good test case because written versions circulated and evolved through a period of more than two millennia in Ancient Mesopotamia and adjacent regions, during a crucial period of the evolution of human civilisation. Currently, scholars distinguish four major stages in the development of the epic (see Table 1). The earliest stratum consists of largely unconnected stories in Sumerian about king Gilgamesh's feats. These poems likely originate at the court of the kings of the Third Dynasty of Ur (ca. 21st century BC; the persona of Gilgamesh is older still, already attested during the Early Dynastic IIIa

Table 1. Summary of the Four Developmental Stages of the Gilgamesh (Approximate Years).

	Sumerian texts	Old Babylonian texts (OB)	Middle Babylonian texts (MB)	Standard Babylonian texts (SB)
Time (BC)	2100–1800	2000–1600	1500–1200	1200–100
Historical context	Akkadian Empire; Third Dynasty of Ur	First Babylonian Dynasty; decline of southern Babylonia	Kassite Dynasty	Kassite Dynasty; Neo-Assyrian Empire; Neo-Babylonian Empire
Language	Sumerian	Akkadian	Akkadian, Hittite and Hurrian	Akkadian
Story structure	Independent poems	Integrated epic	Probably integrated epic, distinct versions	Standardised epic consisting of 12 tablets
Completeness	Relatively well-preserved	Relatively well-preserved	Very fragmented	Best preserved
Geographic origin	Nippur, Ur and Me-Turan	Mainly Mesopotamian cities (e.g. Sippar, Larsa, Nippur, Shaduppum, Diyala and Sulaymaniyah)	Mesopotamia (Ur and Nippur), wider Near East regions (e.g. Bogazköy, Megiddo, Ugarit and Emar)	Mainly from the royal library of the Assyrian king Ashurbanipal I in Nineveh

period, ca. 2500 BC) in southern Mesopotamia and courts in the cities of Isin and Larsa during the early Old Babylonian period (ca. 19th century BC). Most likely these poems were performed orally, and might have been accompanied by musicians, for royal entertainment (George, 2003; Sallaberger, 2008).

The Old Babylonian (OB) period saw the emergence of an integrated story line in the Akkadian language, with new fragments still being published (George, 2009, 2019, George & Al-Rawi, 2020). The new epic known under its incipit *šūtur eli šarrī* incorporated many elements already present in the Sumerian poems, most prominently the account of the journey to the Cedar Wood and the fight against its monstrous guardian Huwawa. Further revisions and edits took place during Middle Babylonian (MB), when the epic spread over the whole of the Ancient Middle East; pertinent tablets translated into different languages are found in places such as the Hittite capital Bogazköy in modern Turkey and Megiddo in present-day Israel (George, 2007). Finally, towards the end of the second millennium BC, the Standard Babylonian version (SB) of the epic saw the light of the day; a scribe called *Sîn-leqe-unninnī* was credited as its author. Of this latter recension of the epic (with the Akkadian title *ša nagba īmur*), about 3000 lines of cuneiform text are extant in a standardised version consisting of 12 tablets. The evolution in the epic was accompanied by significant changes in the social, political, economic and religious conditions which took place between the emergence of the first stories revolving around Gilgamesh in the late third millennium BC and the consolidation of large empires, most notably the Neo-Assyrian Empire and the Neo-Babylonian Empire in the first millennium BC, when the SB version is best attested (Van de Mieroop, 2007). During the Sumerian circulation of the epic, local government was mostly in the hands of aristocratic families, sometimes tied to the royal household by means of marriage. Yet, these rulers had to rely on alliances and support by the local population, which to a large extent limited autocratic powers. Conversely, the Neo-Assyrian Empire at its climax reigned over large swathes of the Ancient Near East, from the Levantine coast in the west to parts of modern Egypt in the south and various small kingdoms in present-day Turkey, such as Phrygia and Lydia in the north

(Radner, 2015). These conquests were enabled by – and at the same time necessitated – a bureaucratic apparatus of hitherto unseen complexity and based on a stricter division of functions than in earlier periods. We are using the epic within this context to examine the structure and changes in personality descriptions across these discrete social-historical contexts in the early ages of large-scale human civilisation.

We use both adjectives and verbs as descriptors of personality to gain an insight into the implicit personality structure of these periods. By analysing both verbs and adjectives, we aim to circumvent the problem of the imperfect correspondence in word classes between languages, which is bound to come with any translation, especially in distant languages. For example, Akkadian has a grammatical category of the stative, a predicative construction of verbal adjectives (and sometimes nouns), which expresses the outcome of an action indicated by the verbal root, or a condition/state. Essentially, the stative constitutes a verbless clause and, depending on the word and the context, it may be translated with a (passive) verb, a noun or an adjective in English (e.g. 'I am sick' is rendered by a single form, *marṣāku*). Our analysis is thus aiming at a broader semantic representation and is less sensitive to artefacts of translation than a purely adjective-focused analysis, typical for most psycho-lexical research, would be. To examine the robustness of our approach, we will explore translation variability in Akkadian.

The Research Questions

The current study explores a number of questions about implicit personality structure based on character descriptions across different versions of the Epic of Gilgamesh. The circulation of the epic for more than two millennia and covering four discrete socio-historical periods provides a unique opportunity for studying both the possible stability and change in personality descriptions in the early stages of large-scale human civilisation. Our study is motivated by three overarching questions. First, how well do contemporary personality theories apply to person descriptors in such an ancient text? Second, can we use character descriptions in this epic to gain some understanding of the implicit personality

structures that may have guided our ancestors? Third, how stable or malleable are personality descriptions over this period? To make sense of possible changes over this time period, we are inspired by the niche construction hypothesis and will qualitatively interpret emerging structures against currently available information from historical research.

Using a mixture of natural language processing (NLP) and human coding, we answer these questions through a combination of analyses. First, we map the adjectives and verbs used to describe characters in the epic to contemporary psychological trait dictionaries, specifically a list of adjectives in which the Big Five has been identified (Goldberg, 1982), and two lists of verbs (De Raad, 1999). This provides us with a first descriptive snapshot of how relevant contemporary trait dictionaries may be for character descriptions in ancient texts, which is relevant for our first research question. Examining changes in the matching to contemporary dictionaries, we start addressing our third research question. Second, we examine whether the same terms (independent of their classification into modern dictionaries) were used in different versions of the epic, which provides a different perspective to help us address the third research question. To the extent that word frequency may reveal information about culturally shared meaning structures, we can uncover insights into the relative stability versus evolution of culturally shared implicit personality models. Third, we apply multidimensional scaling (MDS) on the term co-occurrence matrices across characters in each epoch to capture the implicit personality structures embedded in the different versions of the epic. We interpret these structures qualitatively using available evidence from archaeology and historical records and compare the emerging structures with contemporary personality concepts. This analysis allows us to directly address the second research question. Finally, we compare the MDS models of common terms across the different versions of the epic to explore the temporal stability of the implicit personality structure, which provides yet another perspective for addressing our third research question. In combination, these different analyses provide complementary answers to our three motivating questions. As our employed methods are relatively novel in personality science research, our study aims to provide some guidance and evidence for the usefulness of these analytical tools for studying personality in ancient texts.

Methods

Data sources

We used the English translation published in George's book for the Old, Middle and Standard Babylonian texts (George, 2003) as well as recently translated scripts (George, 2007, 2009, 2018, George & Al-Rawi, 2014, 2020). For the Sumerian poems, we used the English translation published by George (1999). The original Sumerian and Akkadian texts were written in cuneiform on clay tablets, which were deciphered and translated into English.

Sampling statement

All currently documented and publicly available texts were included. Each text found on an individual tablet was coded in the relevant temporal category using the standard archaeological conventions. Following current standards for the translation and interpretation of the epic, we included material across different tablets in cases where the texts were missing due to damages or were present but not deciphered (George, 1999).

We excluded the Middle Babylonian texts for two reasons. First, these texts were recorded in a number of different languages and therefore would have introduced additional variation into the analyses. Second and more importantly, these texts are very fragmentary and there were too few consistent occurrences and co-occurrences of terms in comparison to the other epochs to allow stable cross-temporal analyses (see Table 2).

Data processing

We pre-processed the text using a combination of computational and manual approaches. First, we used tokenisation and part-of-speech (PoS) tagging from the spaCy (version 2.1.3; Honnibal & Montani, 2017) NLP library in Python (Rossum & Drake, 2010; see Table 3). Tokenisation automatically split the texts into single word units, and PoS tagging annotated each word with word type such as noun, adjective and verb. We also used lemmatisation to remove the inflectional forms of the words (see Figure 1 for an overview of the pre-processing and data analytical steps and Table 3 for an example of the pre-processing).

With the texts broken down into single word units, the first author reviewed all the words that were automatically annotated as adjectives and verbs, and referenced the entities that these adjectives and verbs referred to based on human reading of the text (see Table 3 for an example). We used a manual coding process due to the fragmented nature of the tablets and the poem structure of these ancient texts, which made accurate and reliable automatic attribution of words to characters difficult. The automatically extracted entities included all the named entities, such as Gilgamesh, Enkidu, Humbaba, as well as any other single characters or groups of entities that are not named explicitly, such as *man*, *men*, *a person*, *mankind* and *gods*. A list of the adjectives that are used to describe these named entities is presented in Supplementary Table 1. The acting entity and the receiving entity of each action as captured by the verb were coded in the format of *actor:verb:receiver*. For example, 'Gilgamesh hugged him [Enkidu] tight' was coded as *Gilgamesh:hug:Enkidu*. Intransitive verbs such as in 'Gilgamesh began to weep' were coded as *Gilgamesh:weep:x*. Auxiliary verbs were excluded from further analysis.

When a single verb did not capture the specific action, verbal phrases were coded. For example, the verb *take* by itself is ambiguous in meaning in English and can refer to different actions when followed by different nouns. As a result, verbal phrases such as *take notice, take pity, take prisoner* were coded. Similarly, when two verbs were used together to specify one type of action, such as 'go weep', 'make grow', they were coded together as a verbal phrase. We calculated the total number of words, adjectives, verbs and verb phrases (after excluding auxiliary verbs) in each period's version of the texts (see Table 2).

Table 2. Summary of the Number of Total Words, Adjectives and Verbs Across Four Periods.

	Sumerian	Old Babylonian	Middle Babylonian	Standard Babylonian
Total words	14,612	5854	2,519	26,993
Total adjectives	563	246	130	819
Unique adjectives	157	122	83	273
Total verbs	2,058	868	328	2,882
Unique verbs	585	368	190	884

Note. Total counts may include repetitions of the same word; unique counts include counts of distinct adjectives or verbs.

Table 3. Example of the Text in the Coded Spreadsheet, Showing the Tokenisation, Lemmatisation, Part-of-Speech Tagging Using spaCy, and Human Coding of Adjective and Verb Attributions.

|	Purpose	Input	Return
Tokenisation	Automated splitting of text into single word units (tokens)	'Gilgamesh so tall, perfect and terrible, who opened passes in the mountains'	'Gilgamesh' 'so' 'tall' ',' 'perfect' 'and' 'terrible' ',' 'who' 'opened' 'passes' 'in' 'the' 'mountains'
Lemmatisation	Automated removal of inflectional forms and return of base form of words (tokens)	'Gilgamesh' 'so' 'tall' ',' 'perfect' 'and' 'terrible' ',' 'who' 'opened' 'passes' 'in' 'the' 'mountains'	'Gilgamesh' 'so' 'tall' ',' 'perfect' 'and' 'terrible', 'who' 'open' 'pass' 'in' 'the' 'mountain'
Part-of-speech tagging	Automated identification of word classes	'Gilgamesh' 'so' 'tall' ',' 'perfect' 'and' 'terrible' ',' 'who' 'open' 'pass' 'in' 'the' 'mountain'	PROPN (proper noun) ADV (adverb) ADJ (adjective) PUNCT (punctuation) ADJ (adjective) CCONJ (conjunction) ADJ (adjective) PUNCT (punctuation) PRON (pronoun) VERB (verb) NOUN (noun) ADP (adposition) DET (determiner) NOUN (noun)
Manual coding	Human coder assigns adjective and verb tokens to entities	'Gilgamesh' 'so' 'tall' ',' 'perfect' 'and' 'terrible' 'who' 'open' 'pass' 'in' 'the' 'mountain'	 Gilgamesh Gilgamesh Gilgamesh Gilgamesh:open:pass

Data availability, reproducible script and open material statement

The co-occurrence matrices at the character level and code are available on the OSF: https://osf.io/fzbgv/.

Description of dictionary and statistical analyses

Adjective dictionaries. Following the human coding of adjectives and verbs, our first step of analysis was to map all the adjectives from the Gilgamesh to existing modern psychological trait dictionaries (see Figure 1). We used Goldberg's (1982) set of 1710 English trait adjectives. This set of words traces its origins to the first comprehensive list

of 17,953 person-descriptive words (Allport & Odbert, 1936), narrowed down and supplemented by Norman (Norman, 1967) and further refined by Goldberg (1982). The list of 1710 words has been used in psycho-lexical research leading to the identification of the Big Five and the HEXACO models in English data (Ashton et al., 2004; Goldberg, 1990; see also Saucier & Iurino, 2020, for a recent re-analysis of this set of words). Previous work on text analysis of fiction literature (Fischer et al., 2020) compared the implicit personality structures emerging from the 1710 words to results using Allport and Odbert's complete list and Saucier's (1997) shorter yet conceptually broader list of 500 person-descriptive words. The results were comparable across the three lists, but the clearest
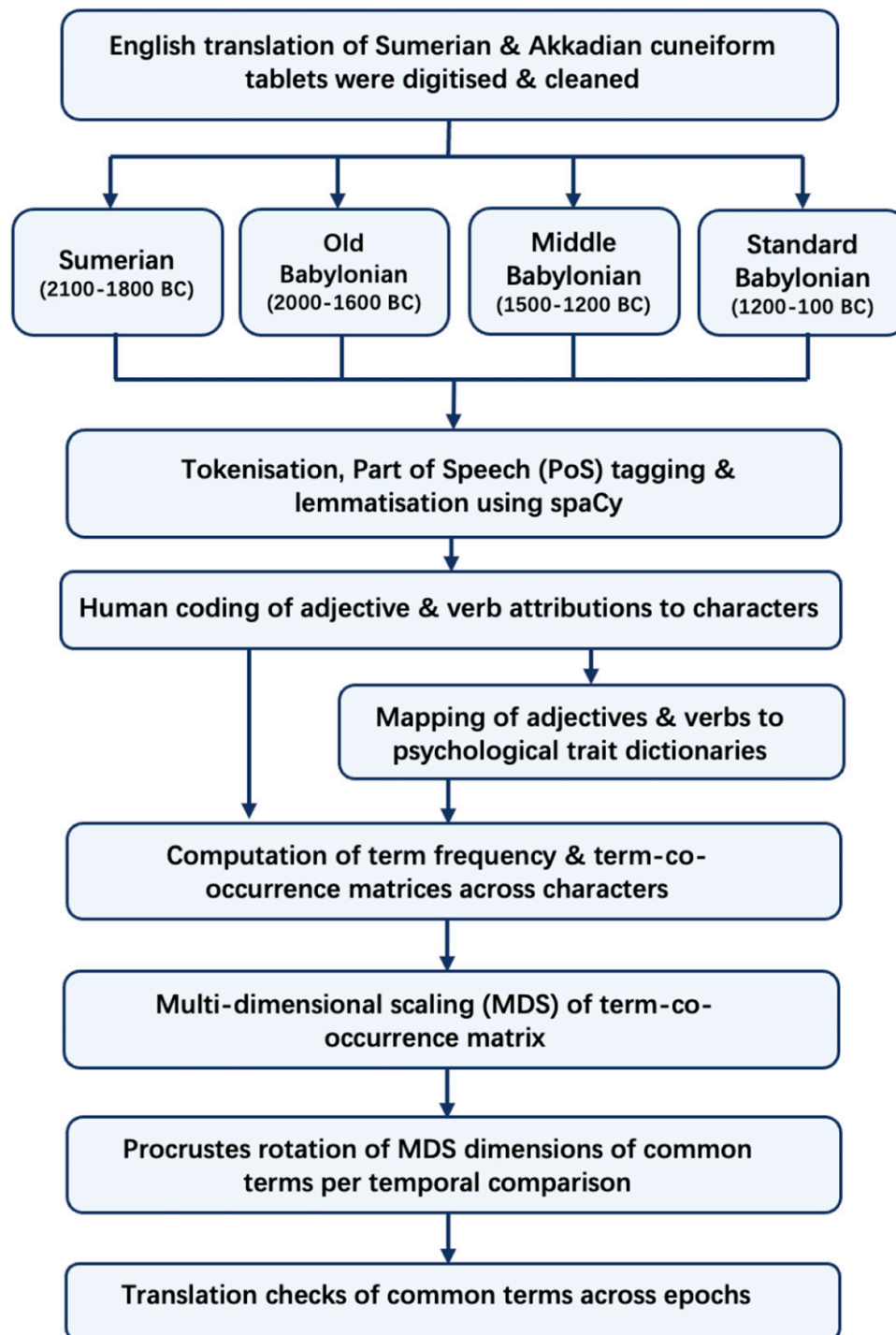
Figure 1. Flow chart detailing the data acquisition, transformation, pre-processing and statistical analysis.

interpretation was obtained using the 1710 list. This list was thus chosen for the present study. Because we were interested to relate our findings to the arguably most widely used current model of personality, the Big Five, we selected from the 1710 words only those terms that clearly and unambiguously loaded on a single trait factor in the original analysis by Ashton et al. (2004), with an absolute factor loading above .30 and no cross-loadings above a difference threshold of .20 ($k = 207$). Second, we used Goldberg's 100 trait markers for the Big Five factor structure (Goldberg, 1992). This list is commonly used as marker terms in lexical studies (e.g. Thalmayer et al., 2021). These two sets were used in the combined dictionary list analyses reported below.

Finally, we also manually coded adjectives from the Sumerian, Old Babylonian and Standard Babylonian version of the epic according to a synonym cluster approach (Wood et al., 2020). Three coders separately coded each adjective into one of 100 personality synonym clusters developed by Goldberg (1990). An adjective was assigned to a trait if two of the three coders agreed with each other about the clustering (e.g. *angry*, not a part of the 1710 list, was consistently coded as belonging to Agreeableness by all three coders);[1] the adjectives are presented in the supplement (Table S1). This procedure allowed us to use all adjectives that are present in the epic and derive a consensus synonym-clustering classification

in terms of the Big Five. At least two coders agreed on 95.2% of the coding. All three coders completely agreed on 46.4% of the coding. Fleiss' kappa was .48, which indicates moderate interrater agreement (Landis & Koch, 1977).

Verbs dictionaries. First, we included an interpersonal verbs dictionary developed by De Raad (1999), which was derived from principal component analysis of self- and peer-ratings of 303 interpersonal trait verbs. A two-factor circular structure showed Dominance and Nurturance factors, which were further divided into four segments: positive Dominance (e.g. *order, attack*), negative Dominance (e.g. *agree, permit*), positive Nurturance (e.g. *help, support*) and negative Nurturance (e.g. *betray, ignore*). We used a list of 164 verbs that had projections of .30 or higher on the circular structure in the study by De Raad (1999).

Second, we used a list of 160 interpersonal behaviour verbs (De Raad, 1999) which showed an eight-factor structure in a factor analysis based on similarity ratings. These factors were interpreted as: Love (e.g. *kiss, caress*), Hate (e.g. *overpower, insult*), Instructional Behaviours (e.g. *demonstrate, explain*), Repairing Behaviours (e.g. *give, provide*), Seeking Contact (e.g. *approach, listen*), Avoiding Contact (e.g. *flee from, avoid*), Emotional Demonstrative (e.g. *lament, cry*) and Behavioural Demonstrative behaviours (e.g. *justify, reveal*). We again matched any character-assigned verbs to these two dictionaries and counted the relative frequency across epochs.

Bottom-up dimensional analysis of person descriptors in each epoch. We treated each named character within the text as an individual and computed a Spearman rank-order correlation based on the term co-occurrences per named character (Fischer et al., 2020). These correlations were transformed into dissimilarities using Euclidean distances and we ran an ordinal MDS with the R package smacof (Mair et al., 2021).

Stability of the individual solutions was estimated using bootstraps with 250 replication samples. A new dissimilarity matrix was generated from each bootstrap sample with replacement; this dissimilarity matrix was submitted to an ordinal MDS as described above and the similarity with the original matrix was calculated after Procrustes rotation (Jacoby & Armstrong, 2014). Here we focus on the replicability of the solution, using a stability measure that compares the variability of each replication with the centroid configuration divided by the total variance across replications (Mair et al., 2022). Larger values imply better replicability. We also report the mean stress value across all bootstrap samples as well as the 95% confidence interval. Lower stress values with smaller confidence intervals suggest better representation of the data in the two-dimensional space.

For the structural comparison across epochs, individual MDS solutions of temporally matched descriptors were Procrustes-rotated to maximal similarity. We always used the older text as the target matrix. Similarity was judged using Tucker's Φ with values above .85 indicating similarity in structure (Fischer & Karl, 2019). Because there are questions about the applicability of overall congruence coefficients and coefficients of alienation a) when using a relatively large number of categories to represent in a multidimensional space and b) when there is the possibility of noise (both of which could be expected in our data), we focus in our interpretation on the similarity of each dimension separately (Borg & Groenen, 2005; Borg & Mair, 2022). We examined the similarity of the dimensions using correlations of the Procrustes-rotated dimensions.

Considering the choices available for NLP analyses, we chose this MDS analysis for two main reasons. First, it is a relatively assumption-free approach that examines the co-occurrence of terms in relation to specific fictional characters. In this sense, it is the equivalent of using data-reduction techniques on survey-based data with participants (e.g. we treat fictional characters as if they were participants in a survey). Using ordinal MDS with Procrustes rotation requires fewer assumptions about the data quality than typically employed data-reduction techniques such as exploratory factor analysis or principal component analysis (Fischer & Fontaine, 2011). It may be possible to use some variant of a latent semantic analysis (LSA) (Landauer & Dumais, 1997) which analyses the distribution of words within sets of documents to quantify the similarity of documents and/or to index documents more efficiently to optimise information retrieval. These techniques do considerably more than what we are interested in, while the applicability to the type of text that we are using has not been carefully evaluated. We used human coding to assign words as information tokens to fictional characters, therefore, preserving human-perceived meaning and not relying on statistical relationships in sparse text (Bender & Koller, 2020).

For all statistical analyses, we used two-sided tests. Our analyses focus on the magnitude of the correlation; we did not adjust significance values for multiple testing. Our research was purely exploratory and we clearly describe it as such.

Effects statement

We report basic descriptive statistics, effect sizes where applicable, exact p-values and 95% confidence (credible) intervals.

Results

Person descriptor matching to modern dictionaries

We mapped all the adjectives and verbs that were matched to characters in the epic onto contemporary psychological dictionaries (Ashton et al., 2004; De Raad, 1999; Goldberg, 1992). To test whether the dictionary matches increased over time, suggesting increasing similarity with modern psychological models of personality, we ran a general linear model to test temporal effects. We adjusted the dictionary matches by the overall frequency of adjectives and verbs, respectively, and included word class as a random effect. We found a significant effect of time period on overall term frequency: $b = 0.004$, SE $= 0.000$, $t = 11.02$ and $p < .005$ (see Figure 2 for graphical representation of the trends over time). The number of adjectives and verbs, relative to all
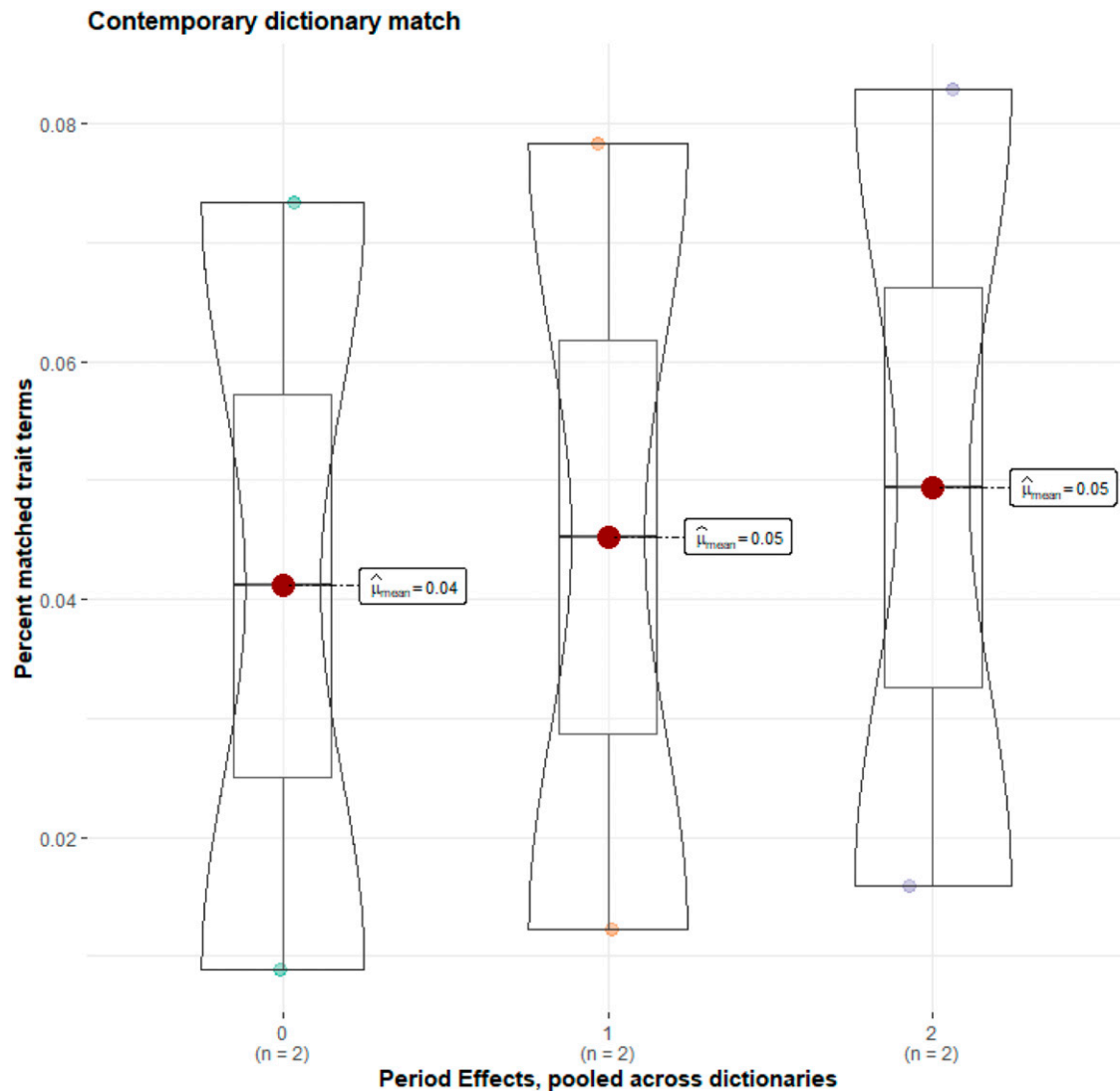
Figure 2. Period effects on dictionary mapping. The results show an increase in mapping of words to contemporary psychological dictionaries. Scores are adjusted for the relative frequency of word class in the texts of each epoch and analyses reported in the text are adjusted for type of word class. Sumerian is coded as 0, Old Babylonian as 1 and Standard Babylonian as 2.

adjectives and verbs in the epic, that could be clearly mapped to a contemporary personality factor increased over time. The total explained variance via the fixed time effect was .63% (Rights & Sterba, 2019).

We next analysed whether descriptions captured a broader range of personality factors over time, that is, whether the number of factors out of the Big Five (identified by adjectives) as well as the number of contemporary verb-based personality dimensions increased over time when describing characters in the epic. Considering the distribution of our count data, we ran a zero-inflated Poisson regression, with number of factors as the outcome, time period as the predictor and dictionary as a random effect. When focussing on all characters mentioned at least once in any version of the epic, we found no significant effect: $b = .216$, SE = .192, $z = 1.13$ and $p = .25$. However, different characters appeared over the different epochs. When focussing on the 10 most frequently mentioned characters across the versions to see whether the descriptions of the more common characters became more complex, we found a significant effect: $b = .466$, SE = .236, $z = 1.98$, $p = 0.048$ (see Figure 3). This suggests that the same characters over

time were described in more complex ways – that is, more of the modern Big Five and verb dimensions were employed by the scribes of the epic in later versions of the epic compared to earlier versions. Considering the different dictionaries and factor structures leading to a maximal possible number of 20 factors across all lists in our analysis, in the earlier versions, approximately two out of 20 possible factors were used to describe the key characters; in the last version of the epic, 4 out of the 20 possible factors were present. To extract explained variance estimates, we respecified the model as a multilevel model. The time effect was significant: $b = 1.125$, SE = 0.305, $t = 3.69$, $p = .008$ and the fixed slope time effect explained 20.5% of the variance. Over time, the more central characters were thus described in terms that captured a broader set of personality factors. This suggests that a greater number of terms can be matched to modern dictionaries in later periods of the epic and that descriptions of central characters in the epic capture a broader set of personality factors over time.

We also analysed the adjectives that we coded using the synonym cluster approach (Wood et al., 2020). A qualitative analysis of these trait representations across periods

Overall effects on trait factors
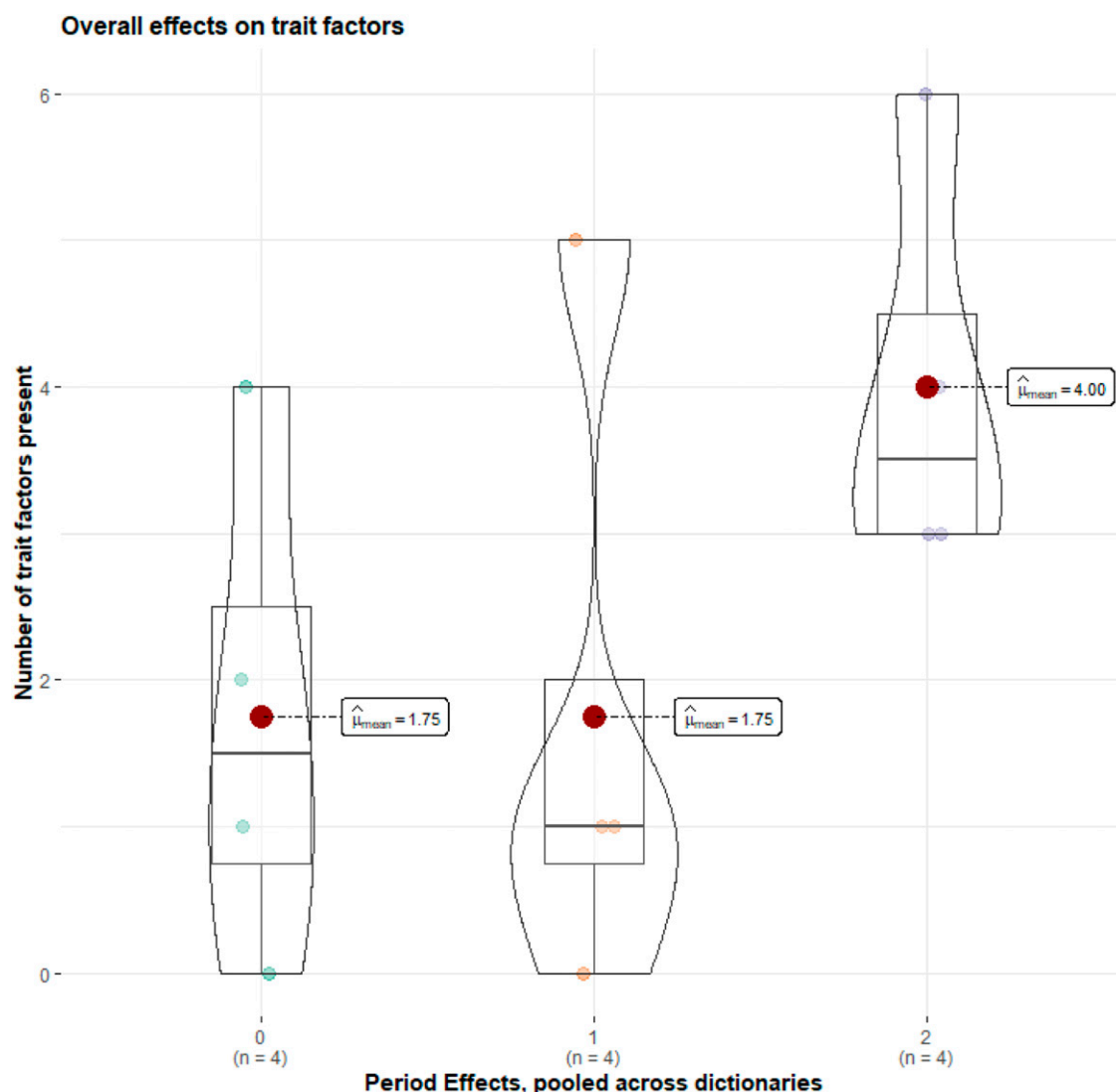
Figure 3. Period effects on psychological trait factor presence among the 10 most frequently mentioned characters. The results show an increase in the number of trait factors according to contemporary psychological models. Analyses reported in text are adjusted for dictionary. Sumerian is coded as 0, Old Babylonian as 1 and Standard Babylonian as 2.

Table 4. Unique Counts of Terms Classified Into the Big Five, According to Synonym Cluster Coding (Raw and Adjusted for Unique Count of Adjectives, Table 2).

Trait dimension	Sumerian	Old Babylonian	Standard Babylonian
Extraversion	7/.045	12/.098	24/.088
Agreeableness	4/.025	2/.016	10/.037
Neuroticism	3/.019	4/.033	6/.022
Conscientiousness	2/.013	1/.008	5/.018
Intellect/Openness	1/.006	1/.008	5/.018

Note. Terms that were not classifiable into the Big Five or where no consensus was reached are excluded.

similarly showed a greater convergence with modern trait classifications over time (see Table 4). Extraversion was the most frequently mentioned personality trait in all three periods. Agreeableness was the second most frequently mentioned trait dimension in Sumerian and SB (and third in OB). Neuroticism-related terms were the third most common trait terms in Sumerian and SB (and second in OB). Both intellect/openness and conscientiousness terms only

started to feature more centrally in the SB version of the epic; only intellect/openness increased linearly across time. When averaging across all five traits and adjusting for number of unique adjectives overall (see Table 2) as well as unique adjectives that were assigned to characters (Sumerian: 56; OB: 41; SB: 98), we observed a near perfect increase across the three epochs ($r_{unique\ adjectives} = .96$, $r_{character\text{-}unique\ adjectives} = .91$, $\rho_{unique\ adjectives} = 1.0$, $\rho_{character\text{-}unique}$

adjectives = 1.0, *p* < .01; CI cannot be computed with three observations only).

Temporal shifts in overall use of personality-descriptive terms

To examine the stability and differences across periods in word usage, we examined the shared frequency distributions of verbs and adjectives used across time epochs. This provides some estimation of stability of word usage over time, indicating whether more frequent terms in the older version continue to be more frequent in the subsequent period. To the extent that terms are more central, these terms should carry more consensually shared meaning and be used for similar purposes. Overall, we found a qualitatively greater similarity in word frequency across temporally closer periods: *r* = .72, 95% CI [.37; .89], *p* = .01 for the match between Sumerian and Old Babylonian texts that partially overlap in time, but differed in language, versus *r* = .40, 95% CI [.05; .66], *p* = .01, for the Old Babylonian and Standard Babylonian texts with a shared language; and *r* = .49 (95% CI [.17; .72], *p* = .01 for Sumerian with Standard Babylonian texts. When comparing these correlations with the time differences (using the earliest epoch dates from Table 1 as reference points), increasing temporal distance was negatively associated with word overlap: *r* = −.92, *p* = .25 (CI's cannot be computed with three observations). This means that word choices were qualitatively more similar in the temporally closer periods.

Co-occurrence analyses to reveal implicit cognitive structures

Next, we examined the structural properties of term co-occurrences at the character level over time periods (following an approach that has provided insights into more recent historical text data (Fischer et al., 2020, p. 202)); we transformed the co-occurrences to dissimilarities using Euclidean distances and applied multidimensional scaling (MDS) on these data. We ran a separate analysis for each epoch. Overall, the individual MDS solutions were quite stable and replicated well in the bootstrap analyses per epoch (see Table 5). The stability index was at .95 or above for each epoch, suggesting that the individual bootstrap solutions with random replacement recovered the same variability. The stress values and associated CIs also indicated that the two-dimensional representation provided an adequate representation of the available data per period.

The supplement (Table S2) shows keywords associated with the polar ends of the epoch-specific solutions. We identified the following core themes of person descriptions within the texts. In the Sumerian text, the first dimension separated emotions and sentiment from action (e.g.

heartstricken, *fine*, *sick*, *dead* vs *reach across*, *escape*, *go near*, *take pleasure*); the second dimension concerned social and moral norms as well as exchange of information within a dominance hierarchy (e.g. *bare lap*, *answer*, *respect*, *submit*, *help*, *captive* vs *receive reply*, *make answer*, *pay attention*, *come to*, *submit*, *wild*, *precious*). Connotations of 'wild' in particular indicated both high status and wealth, but also danger (Sallaberger, 2008). Dimension 1 in the Old Babylonian texts appeared to focus on vigour and age-related activities and states (e.g. *comely*, *young*, *happy*, *achieve*, *grown*, *hail*, *support*, *mighty*, *wise*), which were salient sacred- and social-status markers in ancient Mesopotamia. Dimension 2 differentiated external expression (acting out) of (often negative) emotional states that have an impact on others from internally focused (negative) emotional states (e.g. *enable*, *ferocious*, *make rage against*, *repel* vs *afraid*, *vexed*, *weary*, *pale*). This externalising-versus-internalising distinction is a salient dimension in contemporary clinical studies (Clark, 2005; Lo et al., 2017).

The Standard Babylonian Dimension 1 captured descriptions related to the coordination of actions among equals as well as fitting into a social hierarchy and exerting power over others (e.g. *comely*, *little*, *senior*, *wild*, *walk at side*, *great*, *welcome*, *brave*, *rule*, *prostrate*, *glorious*). This dimension shows some resemblance to the mastery dimension of virtues (Dahlsgaard et al., 2005; Partsch et al., 2022). Dimension 2 featured descriptions that capture salient moral boundaries in that period, especially concern with the transition from uncivilised to human life (Sallaberger, 2008; Tigay, 2002) (e.g. *take in scent*, *cloth*, *give bread*, *give beer*, *anoint*, *capture*, *spread clothing*, *delight*), but also combat-themes (e.g. *join combat*, *capture*, *join forces*, *bind arms*, *defeat*) at one end from the display of power (e.g. *free*, *angry*, *strong*, *able*, *precious*, *change*, *furious*) at the opposite end. These patterns are consistent with the historical conditions, where the emergence of urban centres as the seeds of civilisations was accompanied by military conflict with nomadic neighbours (Scott, 2017).

Exploring word frequency effects on co-occurrence structures

To examine whether word frequency within the text is related to these dimensional properties, we computed the correlations between the position of the individual terms in the two-dimensional solution and the frequency of word occurrences within the text in each epoch. Across the three epochs, positioning along the first dimension was associated with more frequent character descriptions in the epic (Sumerian: *r* = −.21, 95% CI [−.33, −.08], *p* < .01; OB: *r* = −. 18, 95% CI [−.37, .01], *p* = .06; *r* = −.17, 95% CI [−.27, −.07], *p* < .01). Only in the SB version of the epic did

Table 5. Stress and Replicability Statistics of the Multidimensional Scaling Results Per Period.

Statistics	Sumerian	Old Babylonian	Standard Babylonian
Stress I	.159	.093	.151
Stability	.960	.950	.970
Bootstrap stress (95% CI)	.037 (.005, .079)	.052 (.022, .089)	.081 (.056, .109)

we observe an association between structural features of the second dimension and word frequency ($r = .21$, 95% CI [.11, .31], $p < .01$), suggesting that issues of morality and power became more frequent during that period. Overall, this analysis suggests that word frequency had a small but significant effect on the positioning of person-descriptive words along the first dimension in each epoch.

Temporal stability of co-occurrence structures

So far, we primarily focused on structural analyses by epoch. The most intriguing information for evaluating cultural evolution in personality cognitions, however, is the extent of stability underlying the descriptions of characters within the epic across time. To directly examine temporal stability of the person descriptor structure over temporal epochs, we computed Procrustes rotations on the common adjectives and verbs for each pairwise temporal comparison and computed both overall congruence of the structures and the correlations of the main dimensions after rotation to maximal similarity. To preserve temporal progression, we always used the temporally older matrix as the target matrix and rotated the temporally more recent structure to be maximally similar to this historical target matrix.

The structural similarity of the Sumerian and Old Babylonian texts was .819, and the alienation coefficient .574 (common elements: 17). The similarity (or congruence) for Dimension 1 was $r = .68$, 95% CI [.30, .87], $p < .01$. For Dimension 2, the congruence was $r = .09$ (95% CI [−.41, .54], $p = .74$. Comparing Old Babylonian to Standard Babylonian scripts, the overall congruence coefficient was .788 (alienation coefficient = .616, 31 common elements). The similarity for Dimension 1 was relatively weak: $r = .12$, 95% CI [−.24, .46], $p = .51$, and somewhat higher for Dimension 2: $r = .54$, 95% CI [.23, .75], $p < .01$. Finally, comparing the most temporally distant matrices for Sumerian and Standard Babylonian, the overall congruence was .797 (alienation coefficient = .604, common elements = 31). Despite the overall similarity at the structural level, the correlation for Dimension 1 was $r = .18$, 95% CI [−.19, .50], $p = .34$, and for Dimension 2, $r = .30$, 95% CI [−.06, .59], $p = .10$ suggesting low structural similarity over the two millennia separating the two most distant versions.

We compared the cross-temporal correlations for each dimension with the temporal differences (the differences between earliest estimates for each epoch). For Dimension 1, we observed a qualitatively strong negative association ($r = −.98$, $p = .14$; CIs cannot be computed with three observations for this and the following analyses), whereas for Dimension 2, we found a positive association ($r = .78$, $p = .43$). We also examined the overall congruence (but see Borg & Mair, 2022 for some caution against using overall congruence coefficients) with the time difference and found a negative association: $r = −.92$, $p = .26$ (the association was proportionally inverse with the alienation coefficient: $r = .92$, $p = .25$). We note that these associations showed quite strong effect sizes (mostly $r > .90$) but due to the small number of observations, the effects were not statistically significant. As with the overall pattern of the word frequency analyses reported at the beginning, the similarity of the dimensions suggests that the temporally

closer periods (especially for Dimension 1 in the Sumerian and Old Babylonian versions, and Dimension 2 for the Old Babylonian with the Standard Babylonian version, but also the overall congruence) show higher similarity in implicit cognitive structures compared to the temporally more distant versions of the epic. However, these qualitative interpretations need to be considered in light of the observation that none of the congruence coefficients were reaching commonly accepted thresholds of .80 or higher and that all our analyses are based on only three observations, which significantly limit our ability to draw quantitative conclusions.

Robustness check regarding translation stability

To evaluate the extent to which our analysis using the English translation can be considered as a fair representation of the original Akkadian text, we reviewed a list of key person descriptors in the English text and their corresponding Akkadian roots. Person descriptors were considered key for this purpose if the English term occurred at least two times in both the Old and Standard Babylonian versions of the epic. The list of the terms is the following: comely, ferocious, glad, happy, mighty, sick, strange, wild, young. We found that 87.2% of the key English descriptors correspond to the same roots in the original Akkadian text. Therefore, the English translations for the Akkadian text at least seem to be consistent with the original Akkadian words.

Discussion

Using the Gilgamesh epic as a case study, we traced person descriptors to modern psychological dictionaries and observed evolutionary change suggesting a gradual approximation towards modern psychological personality structure by the end of the original Akkadian circulation of the epic. This epic is unique source material because it is one of the oldest known (and reasonably complete) epics, which moreover is extant in different versions, spread over nearly two millennia.

The first insight is that individual character-descriptive words can be related to contemporary personality dimensions, suggesting at least a basic level of diachronic psychic unity of humanity. From a human evolutionary perspective, the relative consistency and importance of Extraversion and Agreeableness in both modern person descriptions (Wood et al., 2020) as well as the epic of Gilgamesh is noteworthy. This finding suggests a universal concern for describing the social features of literary characters, which maps onto a universal human need to communicate social information to and about others. This observation obviously reinforces and explains the appeal of the epic over millennia, but it is not a trivial observation given recent claims about niche construction effects for personality characteristics. Our findings provide a more nuanced perspective on such niche construction theories and suggest that narrators and preserved scripts captured descriptions that may relate to biologically shaped predispositions which are nevertheless expressed variably across time and space.

Focussing on the personality dimensions that we extracted using MDS, probably the key finding is that the emergent dimensions and clusters could not easily be classified according to modern personality taxonomies. Instead, the main themes of these dimensions seem to be compatible with some of the major themes that have been discussed in the classical literature, such as issues of social status and hierarchy relations, in particular dominance themes related to social hierarchy, war and conquest as well as markers of youth-related hedonistic indulgence and vigour that reflect warrior or god-like qualities (George, 2003; Sallaberger, 2008; Tigay, 2002). Another theme that is meaningful within the historical context is the discussion of markers and actions that imply a transition from barbarian status to human civilised life. These have been central themes in historical texts on Sumerian and Akkadian culture (Selz, 2016) and our findings suggest that these can also be traced in the implicit personality concepts when describing characters in this epic.

Bringing these patterns back to the frequency distributions, it is important to consider our patterns in light of the relatively small number of words that can be clearly matched to contemporary dictionaries. We used a rather strict criterion (primary loading >.30 and no cross-loadings >.20) for deciding whether a word was clearly applicable to one of the Big Five in a classical analysis of this set of words (Ashton et al., 2004) (but see also Saucier & Iurino (2020) for a recent re-analysis of the same dataset). Using both such a strict criterion and a more lenient category classification, we did not encounter many words within the epic that would unambiguously be related to contemporary definitions. Although it may appear disappointing at a first sight, such a finding is in fact consistent with Galton's original speculations when first proposing the lexical hypothesis: it is possible to measure the core features of what we now call personality dimensions but the individual features are 'so intermixed that they are never singly in action' (Galton, 1949, p. 181). What our results nevertheless suggest is that over the two millennia of circulation of this epic, the number of words and the number of personality factors that can be unambiguously matched to contemporary models increased with greater societal complexity.

We interpret these patterns in line with a more nuanced niche construction process. As mentioned above, we observed that translations of the more recent stories contained a relatively higher proportion of terms that could be mapped onto modern trait dictionaries. We must presume that the translator was unaware of these trait dictionaries, and we hypothesise that this increased mapping suggests that the language within Gilgamesh gradually started to resemble more contemporary concepts that can be mapped onto relevant personality dictionaries. Furthermore, in terms of the temporal shifts more broadly, independent of modern dictionaries, both the word frequency analyses and the structural stability analyses of character-mapped word co-occurrences suggested that temporally closer periods tended to be more similar. This effect was clearest for the relationship between the Sumerian version and the Old Babylonian version, which showed the qualitatively highest correlation of words in common use and greatest overlap for

Dimension 1 of the MDS. The Old Babylonian and Standard Babylonian versions showed some convergence in the word frequency and also some similarity along Dimension 2 of the MDS. In contrast, the historically most distant Sumerian and Standard Babylonian versions showed moderate overlap in word frequency and weak overlap along the two dimensions of the MDS (max $r = .30$). Together with the overall dictionary matching results and overall congruence effects, these qualitative observations suggest a gradual psychological progression over the periods towards greater resemblance of modern personality concepts.

This temporal development coincides with a larger trend in ancient literature in the waning centuries of the second millennium BC (during the circulation of the SB version), when a 'greater profundity of reasoning (…) was not without its counterpart in a sobering of the Old Babylonian *joie de vivre*' (Lambert, 1996; p. 17). The second half of the second millennium BC also saw the intensification and regularisation of written exchange with wider regions beyond Mesopotamia, and the content of such exchanges pertained to not only the sphere of international diplomacy, but also ideas and literary *topoi* (themes). Military defeat and foreign rule that took place in Babylonia itself were characterised by a tension between cultural assimilation on part of the conquerors and significant changes in the social framework of the societies. The increasing distinction of internalising versus externalising psychological characteristics in the structural analyses also coincides with previously identified trends of greater individualisation (e.g. recognition of individuals as authors of texts, presence of ownership records and personal deeds mentioned instead of role descriptions; Selz, 2016). While a more in-depth social-historicist contextualisation of Babylonian literature is long overdue, we suggest reading the present results as tantalising evidence that the growing complexity of the authors' social and political environment may be associated with personality concepts that gradually start to resemble those found in modern-day societies; in other words, that social structure leads to a shaping of cognition in the form of implicit personality models.

Alternative interpretations of the historical significance of the data – for example, by focussing on developments in the language – would be less convincing. It is true that the SB version was significantly expanded, modernised and edited to reflect the changed cultural, religious and literary preferences; yet, it was still close to the OB one as regards its main plot (and some authors such as Tigay, 2002, actually consider the Old Babylonian version more innovative and revolutionary in its changed scope and focus, compared to the ritualistic changes made in the Standard Babylonian version). Overall, the poem's structure is rather stable over time, as evidenced by a simple count of the ratio of the word groups: adjectives make up about 4% in all three variants (Sumerian, OB and SB) analysed here, while verbs make up about 14–15% in both Sumerian and OB version and 11% in SB. Moreover, as observed by George (2003), alterations included both expansions, but mainly only by means of repetition (e.g. a passage of 22 lines occurs five times in the SB version; see George 2003, p. 46), and cuts and syncopations of the text. In an overwhelming majority of cases

of altered words, these were replaced by roughly synonymous terms, while other changes may have been introduced by concerns regarding poetic metre.

How do our findings match onto previous interpretations of the epic by Assyriologists? A number of scholars have discussed the evolution of the story (George, 1999; Sallaberger, 2008; Tigay, 2002), pointing out a shift in both content and style. Hence, the Sumerian stories are regarded as tales of a heroic warrior king, intended to provide entertainment at the royal court. In the Old Babylonian period, the casting of isolated stories into a coherent single narrative was accompanied by a shift in focus, though scholars differ in their interpretation. For Sallaberger (2008), the OB version centres on Gilgamesh as a human being, resulting in a moral tale about friendship and life in the knowledge of death's inevitability, while the SB work is grounded in 'wisdom literature' which was salient at the time of the Assyrian empire, adding not only a didactic facet but also a concern with one's afterlife. For George, on the other hand, the OB poem 'was a hymn to heroism and kingly might' while the SB version suggested a 'sombre meditation' that is also 'more introspective' (George, 2003, p. 33). Contrary to these previous structural analyses of form and content, our lexical analysis based on George's translation suggests greater continuity at a psychological level between the Sumerian and Old Babylonian texts. Despite the noted shifts in narrative and message, it appears that the story used relatively similar implicit personality concepts to cast the main heroes of the story. The psychological structure of actors and the frequency of terms used to describe characters were more similar across these two versions, even though they used a different language, in contrast to the Old and Standard Babylonian versions which both used Akkadian. Considering the closer proximity in time of the Sumerian and Old Babylonian sources, the psychological make-up of the audience may have remained similar and hence, we found greater continuity in relevant aspects in spite of a complete and creative retelling of the story. For this reason, we propose that a psychological analysis of the terms used in a text, in particular an analysis of the co-occurrence structures of character descriptions, provides an important tool for the analysis of ancient texts.

One central limitation of our analysis is that we used English translation of these texts. At the same time, the mapping onto modern dictionaries requires translation. It could be argued that it is safer to use a high-quality translation that preserves the textual integrity instead of working with the original cuneiform terms, which requires additional work in terms of both computational mapping and subsequent translation without having the textual context. We provided preliminary evidence that translator choices for original terms in the epic are relatively unlikely to have affected the results. Nevertheless, a bottom-up linguistic analysis in Sumerian and Akkadian may provide additional nuances and insights that are missed when using a translation.

Our analyses relied on mixed methods, combining human coding with computational tools from NLP. Considering recent developments in NLP, specifically various transformer models for word and sentence embeddings and their application to personality structure in contemporary corpora (Cutler & Condon, 2022), our approach may appear

rather simple, yet straightforward. We treated words as information tokens about linguistic choices to convey information about characters in the epic to an audience. This linguistic assignment of information tokens to characters allowed us to deduce some implicit cognitive models of personality. What terms are being used together to describe a specific character? Do these co-occurrences reveal some underlying dimensionality, similar to the insights that have been gained in personality research using self- or other-report surveys? We stayed as close as possible to the available text which did not require us to rely on additional assumptions about the language or modelling choices, while also preserving some conceptual similarity to canonical research practice in the field of personality research. Using a human coder to assign information tokens to fictional characters, we were also able to utilise all available information from the context in these highly fragmented and sparse texts.

We think our approach has merit in light of the available corpus. Large language models that have started to dominate text analyses in recent years are built and trained on large corpora of mostly contemporary texts (e.g. content from Web pages). To the best of our knowledge, the gold standard of contemporary transformer-based models has not been evaluated against highly fragmented documents which use a poetry structure with both elliptical and complex sentences and employ archaic language constructions.

Psychological analysis can often access the individuals that produced certain texts and obtain, for example, self-report data to cross-validate the textual analyses. This in itself has led to significant debates about the relative validity of survey responses versus textual analyses (Boyd & Pennebaker, 2017). Nevertheless, having access to data from independent methods pertaining to the same individuals allows at least in principle a disentangling of biases inherent in either method. When using historical data, this opportunity for cross-validation does not exist (at least not with the type of data that we are dealing with here), and new avenues for identifying the validity within the specific historical context in which the texts were generated need to be sought. Using large language models that have been pretrained on contemporary language and word usage would mean we may lose precision in tracking culturally and historically meaningful word associations, which represent nuances that may be considered insignificant within the statistical models that aim to predict the next likely word.

We suggest one interesting option for future research may be to pretrain such models on text corpora from specific time periods and then compare the resulting distance metrics from each temporal instance with each other. An interesting computational and statistical challenge is to develop algorithms that allow the identification of nuanced shifts and subtle connotations that change over time and are at the threshold of statistical detection within the distribution of terms in a larger temporal corpus. Diachronic word embeddings (Hamilton et al., 2016) are one line of work that seeks to address this challenge, but even with those methods important theoretical nuances may be pushed towards the boundaries of the trained language models due to the sparsity of the text at hand and the rare occurrence of individual terms. It needs to be investigated whether meaningful results can be produced given the very

small corpora available for specific temporal epochs, which may reduce the possibilities to reliably detect subtle changes that are not artefacts of the available source documents.

Current algorithms rely on statistical patterns to predict the next likely word, emphasising form over meaning (Bender & Koller, 2020). We used human coders to identify which information tokens (words) were associated with which character and then analysed the co-occurrences of information tokens to identify patterns of meaning across recurring associations of words with characters within the epic. In this sense our approach is more robust. Yet, even our analysis cannot overcome some general challenges of biased analyses, most importantly our inability to interact with the language users and analyse their 'lived' language experience against the power relations in place that are represented in the texts and are incorporated into the statistical models (Blodgett et al., 2020). Furthermore, the identified dimensions in the MDS need to be interpreted, which brings in contemporary bias. Together with our inability to verify the robustness of these dimensions beyond purely statistical replicability, we need to acknowledge that slight changes in the rotated spaces together with our contemporary intuitions about ancient societies may have led to biases.

In closing, we present a first linguistic analysis of the implicit personality structure in the Gilgamesh epic. We found an increasing usage of terms that can be mapped onto modern dictionaries in the later versions of the epic. The analysis of word frequencies and structural analyses suggested subtle temporal shifts in the implicit personality models that our ancestors had in mind when chiselling the epic onto their clay tablets. Examining the structures and changes of character descriptions, analyses of ancient texts using modern psychological and linguistic tools can provide new insights into both the psychology of periods long gone and cultural evolution dynamics.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Data accessibility statement

The co-occurrence matrices at the character level and code are available on the OSF: https://osf.io/fzbgv/.

ORCID iDs

Ronald Fischer https://orcid.org/0000-0002-3055-3955
Johannes A. Karl https://orcid.org/0000-0001-5166-0728

Supplemental Material

Supplemental material for this article is available online.

Note

1. Although angry hostility is a facet of Neuroticism, it has a pronounced double loading on Agreeableness (John et al., 2008; McCrae et al., 2005). More generally, coders' classification of some terms may have been affected by the interpretation of the terms in the context of the Gilgamesh epic.

References

Allport, G. W., & Odbert, H. S. (1936). Trait-names: A psycholexical study. *Psychological Monographs*, *47*(1), 1–171. https://doi.org/10.1037/h0093360

Al-Rawi, F. N. H., & George, A. R. (2014). Back to the cedar forest: The beginning and end of tablet V of the standard babylonian epic of gilgameš. *Journal of Cuneiform Studies*, *66*(1), 69–90. https://doi.org/10.5615/jcunestud.66.2014.0069

Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc*, *11*(2), 150–166. https://doi.org/10.1177/1088868306294907

Ashton, M. C., & Lee, K. (2020). Objections to the HEXACO model of personality structure—and why those objections fail. *European Journal of Personality*, *34*(4), 492–510. https://doi.org/10.1002/per.2242

Ashton, M. C., Lee, K., & Goldberg, L. R. (2004). A hierarchical analysis of 1,710 English personality-descriptive adjectives. *Journal of Personality and Social Psychology*, *87*(5), 707–721. https://doi.org/10.1037/0022-3514.87.5.707

Beckman, G., Foster, B. R., Hollis, S. T., Kottsieper, I., Pitard, W. T., & Rubio, G. (2011). In C. S. Ehrlich (Ed.), *From an antique land: An introduction to ancient near eastern literature*. Rowman and Littlefield Publishers.

Bender, E. M., & Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, *58*, 5185–5198 https://doi.org/10.18653/v1/2020.acl-main.463

Berry, J. W. (1989). Imposed etics-emics-derived etics: The operationalization of a compelling idea. *International Journal of Psychology*, *24*(2–6), 721–735. https://doi.org/10.1080/00207598908247841

Blodgett, S. L., Barocas, S., Daumé III, H., & Wallach, H. (2020). Language (technology) is power: A critical survey of "bias" in NLP. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, *58*, 5454–5476. https://doi.org/10.18653/v1/2020.acl-main.485

Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and applications* (p. 2). Springer.

Borg, I., & Mair, P. (2022). A note on procrustean fittings of noisy configurations. *Austrian Journal of Statistics*, *51*(4), 1–9. https://doi.org/10.17713/ajs.v51i4.1423

Boyd, R. L., & Pennebaker, J. W. (2017). Language-based personality: A new approach to personality in a digital world. *Current Opinion in Behavioral Sciences*, *18*, 63–68. https://doi.org/10.1016/j.cobeha.2017.07.017

Buss, D. M. (1991). Evolutionary personality psychology. *Annual Review of Psychology*, *42*(1), 459–491. https://doi.org/10.1146/annurev.ps.42.020191.002331

Buss, D. M. (1997). Evolutionary foundations of personality. In *Handbook of personality psychology* (pp. 317–344). Elsevier. https://doi.org/10.1016/B978-012134645-4/50014-7

Cheung, F., Fan, W., & To, C. (2008). The Chinese personality assessment inventory as a culturally relevant personality measure in applied settings: CPAI applications. *Social and Personality Psychology Compass*, *2*(1), 74–89. https://doi.org/10.1111/j.1751-9004.2007.00045.x

Clark, L. A. (2005). Temperament as a unifying basis for personality and psychopathology. *Journal of Abnormal Psychology*, *114*(4), 505–521. https://doi.org/10.1037/0021-843X.114.4.505

Cutler, A., & Condon, D. M. (2022). Deep lexical hypothesis: Identifying personality structure in natural language. In: Journal of Personality and social psychology, *No pagination specified-No pagination specified*. https://doi.org/10.1037/pspp0000443

Dahlsgaard, K., Peterson, C., & Seligman, M. E. P. (2005). Shared virtue: The convergence of valued human strengths across culture and history. *Review of General Psychology*, *9*(3), 203–213. https://doi.org/10.1037/1089-2680.9.3.203

Delnero, P. (2016). Literature and identity in Mesopotamia during the old babylonian period. *Problems of Canonicity and Identity Formation in Ancient Egypt and Mesopotamia*, *43*, 19–50.

De Raad, B. (1999). Interpersonal lexicon: Structural evidence from two independently constructed verb-based taxonomies. *European Journal of Psychological Assessment*, *15*(3), 181–195. https://doi.org/10.1027//1015-5759.15.3.181

De Raad, B., Barelds, D. P. H., Levert, E., Ostendorf, F., Mlacić, B., Di Blas, L., Hřebíčková, M., Szirmák, Z., Szarota, P., Perugini, M., Church, A. T., & Katigbak, M. S. (2010). Only three factors of personality description are fully replicable across languages: A comparison of 14 trait taxonomies. *Journal of Personality and Social Psychology*, *98*(1), 160–173. https://doi.org/10.1037/a0017184

De Raad, B., Barelds, D. P. H., Timmerman, M. E., De Roover, K., Mlačić, B., & Church, A. T. (2014). Towards A Pan–cultural personality structure: Input from 11 psycholexical studies. *European Journal of Personality*, *28*(5), 497–510. https://doi.org/10.1002/per.1953

Diggle, J., & Theophrastus. (2007). Theophrastus, characters (repr., digital print). Cambridge Univ Press.

Durkee, P. K., Lukaszewski, A. W., von Rueden, C. R., Gurven, M. D., Buss, D. M., & Tucker-Drob, E. M. (2022). Niche diversity predicts personality structure across 115 nations. *Psychological Science*, *33*(2), 285–298. https://doi.org/10.1177/09567976211031571

Fischer, R. (2017). *Personality, values, culture: An evolutionary approach*. Cambridge University Press. https://doi.org/10.1017/9781316091944

Fischer, R. (2021). Alternative four-factor structure of the Mini-IPIP in Thailand. *International Journal of Personality Psychology*, *7*, 35–42. https://doi.org/10.21827/ijpp.7.37978

Fischer, R., & Fontaine, J. R. J. (2011). Methods for investigating structural equivalence. In *Cross-cultural research methods in psychology* (pp. 179–215). Cambridge University Press.

Fischer, R., & Karl, J. A. (2021). *Niche diversity influences psychometric properties of a brief personality scale – evidence from representative samples during the COVID-19 pandemic* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/5jva9

Fischer, R., & Karl, J. A. (2019). A primer to (Cross-Cultural) multigroup invariance testing possibilities in R. *Frontiers in Psychology*, *10*, 1507. https://doi.org/10.3389/fpsyg.2019.01507

Fischer, R., Karl, J. A., Luczak–Roesch, M., Fetvadjiev, V. H., & Grener, A. (2020). Tracing personality structure in narratives: A computational bottom–up approach to unpack writers, characters, and personality in historical context. *European Journal of Personality*, *34*(5), 917–943. https://doi.org/10.1002/per.2270

Foster, B. (2009). Akkadian literature. In *From an antique land: An introduction to ancient near eastern literature*. Rowan and Littlefield.

Galton, F. (1949). The measurement of character. In *Readings in general psychology* (pp. 435–444). Prentice-Hall. https://doi.org/10.1037/11352-058

George, A. R. (1999). *The epic of Gilgamesh: The Babylonian epic poem and other texts in Akkadian and Sumerian*. Allen Lane The Penguin Press.

George, A. R. (2003). *The Babylonian Gilgamesh epic: Introduction, critical edition, and cuneiform texts*. Oxford University Press.

George, A. R. (2007). The gilgameš epic at ugarit. *Aula Orientalis*, *25*, 237–254.

George, A. R. (2009). *Babylonian literary texts in the schøyen collection*. CDL Press.

George, A. R. (2018). Enkidu and the harlot: Another fragment of Old babylonian gilgameš. *Zeitschrift Für Assyriologie Und Vorderasiatische Archäologie*, *108*(1), 10–21. https://doi.org/10.1515/za-2018-0002

George, A. R., & Al-Rawi, F. N. H. (2020). Gilgamesh dreams of Enkidu: An Old babylonian tablet of Gilgamesh in the suleimaniyah museum. *Revue d'assyriologie et d'archéologie Orientale*, *113*(1), 131–138. https://doi.org/10.3917/assy.113.0131

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. *Review of Personality and Social Psychology*, *2*, 141–165.

Goldberg, L. R. (1982). From Ace to Zombie: Some explorations in the language of personality. *Advances in Personality Assessment*, *1*, 203–234.

Goldberg, L. R. (1990). An alternative "description of personality": The Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216–1229. https://doi.org/10.1037/0022-3514.59.6.1216

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*(1), 26–42. https://doi.org/10.1037/1040-3590.4.1.26

Gurven, M., von Rueden, C., Massenkoff, M., Kaplan, H., & Lero Vie, M. (2013). How universal is the Big five? Testing the five-factor model of personality variation among forager–farmers in the Bolivian amazon. *Journal of Personality and Social Psychology*, *104*(2), 354–370. https://doi.org/10.1037/a0030841

Hamilton, P. (2003). *Historicism* (2nd ed). Routledge.

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2016). Diachronic word embeddings reveal statistical laws of semantic change. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, *1*(Long Papers), 1489–1501 https://doi.org/10.18653/v1/P16-1141

Honnibal, M., & Montani, I. (2017). *spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing*.

Jacoby, W. G., & Armstrong, D. A. (2014). Bootstrap confidence regions for multidimensional scaling solutions:

Bootstrapping MDS solutions. *American Journal of Political Science*, *58*(1), 264–278. https://doi.org/10.1111/ajps.12056

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 114–158). Guilford Press.

Konstantopoulos, G. (2017). Through the guts of a beggar: Power, authority, and the king in Old babylonian proverbs. *Kaskal*, *14*, 153–168.

Laland, K. N., Odling-Smee, J., & Feldman, M. W. (2000). Niche construction, biological evolution, and cultural change. *The Behavioral and Brain Sciences*, *23*(1), 131–146. https://doi.org/10.1017/S0140525X00002417

Lambert, W. G. (1996). *Babylonian wisdom literature*. Eisenbrauns.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174, https://doi.org/10.2307/2529310

Lo, M. T., Hinds, D. A., Tung, J. Y., Franz, C., Fan, C. C., Wang, Y., Smeland, O. B., Schork, A., Holland, D., Kauppi, K., Sanyal, N., Escott-Price, V., Smith, D. J., O'Donovan, M., Stefansson, H., Bjornsdottir, G., Thorgeirsson, T. E., Stefansson, K., McEvoy, L. K., & Chen, C. H. (2017). Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nature Genetics*, *49*(1), 152–156. https://doi.org/10.1038/ng.3736

Lukaszewski, A. W., Gurven, M., von Rueden, C. R., & Schmitt, D. P. (2017). What explains personality covariation? A test of the socioecological complexity hypothesis. *Social Psychological and Personality Science*, *8*(8), 943–952. https://doi.org/10.1177/1948550617697175

Mair, P., Groenen, P. J. F., & de Leeuw, J. (2022). More on multidimensional scaling and unfolding in R: Smacof version 2. *Journal of Statistical Software*, *102*(10), 10. https://doi.org/10.18637/jss.v102.i10

Mair, P., Leeuw, J. D., Groenen, P. J. F., & Borg, I. (2021). *smacof: Multidimensional scaling (2.1-3*. [Computer software] https://CRAN.R-project.org/package=smacof

Mayer, J. D., Lin, S. C., & Korogodsky, M. (2011). Exploring the universality of personality judgments: Evidence from the great transformation (1000 BCE–200 BCE). *Review of General Psychology*, *15*(1), 65–76. https://doi.org/10.1037/a0020711

McAdams, D. P., Anyidoho, N. A., Brown, C., Huang, Y. T., Kaplan, B., & Machado, M. A. (2004). Traits and stories: Links between dispositional and narrative features of personality. *Journal of Personality*, *72*(4), 761–784. https://doi.org/10.1111/j.0022-3506.2004.00279.x

McCrae, R. R., & Terracciano, A. (2005) 79 Members of the Personality Profiles of Cultures Project. Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology*, *89*, 407–425. https://doi.org/10.1037/0022-3514.89.3.407

Nel, J. A., Valchev, V. H., Rothmann, S., van de Vijver, F. J. R., Meiring, D., & de Bruin, G. P. (2012). Exploring the personality structure in the 11 languages of South Africa: Personality structure in South Africa. *Journal of Personality*, *80*(4), 915–948. https://doi.org/10.1111/j.1467-6494.2011.00751.x

Norman, W. T. (1967). *Personality trait descriptors–normative operating characteristics for a university population* (p. 2800). https://eric.ed.gov/?id=ED014738

Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (2003). *Niche construction: The neglected process in evolution*. Princeton University Press.

Partsch, M. V., Bluemke, M., & Lechner, C. M. (2022). Revisiting the hierarchical structure of the 24 VIA character strengths: Three global dimensions may suffice to capture their essence. *European Journal of Personality*, *36*(5), 825–845. https://doi.org/10.1177/08902070211017760

Passakos, C. G., & de Raad, B. (2009). Ancient personality: Trait attributions to characters in Homer's Iliad. *Ancient Narrative*, *7*, 75–95.

Porter, A. (2014). *Mobile pastoralism and the formation of near Eastern civilizations weaving together society*. https://www.vlebooks.com/vleweb/product/openreader?id=none&isbn=9780511895012

Radner, K. (2015). *Ancient assyria: a very short introduction* (First edition). Oxford University Press.

Rights, J. D., & Sterba, S. K. (2019). Quantifying explained variance in multilevel models: An integrative framework for defining R-squared measures. *Psychological Methods*, *24*(3), 309–338. https://doi.org/10.1037/met0000184

Rosenberg, S., & Jones, R. (1972). A method for investigating and representing a person's implicit theory of personality: Theodore Dreiser's view of people. *Journal of Personality and Social Psychology*, *22*(3), 372–386. https://doi.org/10.1037/h0032891

Rossum, G. V., & Drake, F. L. (2010). *The Python language reference*. Python Software Foundation. Release 3.0.1 [Repr.]).

Sallaberger, W. (2008). *Das gilgamesch-epos: Mythos, Werk und tradition (Orig.-Ausg)*. Beck.

Saucier, G. (1997). Effects of variable selection on the factor structure of person descriptors. *Journal of Personality and Social Psychology*, *73*(6), 1296–1312. https://doi.org/10.1037/0022-3514.73.6.1296

Saucier, G., & Goldberg, L. R. (2001). Lexical studies of indigenous personality factors: Premises, products, and prospects. *Journal of Personality*, *69*, 847–879. https://doi.org/10.1111/1467-6494.696167

Saucier, G., & Iurino, K. (2020). High-dimensionality personality structure in the natural language: Further analyses of classic sets of English-language trait-adjectives. *Journal of Personality and Social Psychology*, *119*(5), 1188–1219. https://doi.org/10.1037/pspp0000273

Saucier, G., Thalmayer, A. G., Payne, D. L., Carlson, R., Sanogo, L., Ole-Kotikash, L., Church, A. T., Katigbak, M. S., Somer, O., Szarota, P., Szirmák, Z., & Zhou, X. (2014). A basic bivariate structure of personality attributes evident across nine languages: Basic bivariate personality structure. *Journal of Personality*, *82*(1), 1–14. https://doi.org/10.1111/jopy.12028

Scott, J. C. (2017). *Against the grain: A deep history of the earliest states*. Yale University Press.

Selz, G. J. (2016). Sumerer und akkader: Geschichte - gesellschaft - kultur (3., aktualisierte auflage, originalausgabe). Verlag C.H. Beck.

Singh, J. K., Misra, G., & De Raad, B. (2013). Personality structure in the trait lexicon of Hindi, a major language spoken in India. *European Journal of Personality*, *27*(6), 605–620. https://doi.org/10.1002/per.1940

Smaldino, P. E., Lukaszewski, A., von Rueden, C., & Gurven, M. (2019). Niche diversity can explain cross-cultural differences in personality structure. *Nature Human Behaviour*, *3*(12), 1276–1283. https://doi.org/10.1038/s41562-019-0730-3

Thalmayer, A. G., Job, S., Shino, E. N., Robinson, S. L., & Saucier, G. (2021). ǂŪsigu: A mixed-method lexical study of character description in khoekhoegowab. *Journal of Personality and Social Psychology*, *121*(6), 1258–1283. https://doi.org/10.1037/pspp0000372

Tigay, J. H. (2002). *The evolution of the Gilgamesh epic*. Bolchazy-Carducci.

Tinney, S. (1999). On the curricular setting of sumerian literature. *Iraq*, *61*, 159–172. https://doi.org/10.2307/4200472

Van de Mieroop, M. (2007). *A history of the ancient Near East, ca* (2nd ed, pp. 3000–3323). Blackwell Pub.B.C.

West, M. L. (1999). *The East face of helicon: West asiatic elements in Greek poetry and myth*. Clarendon Press.

Wood, J. K., Gurven, M., & Goldberg, L. R. (2020). Ubiquitous personality–trait concepts in 13 diverse and isolated languages: A cluster–classification approach. *European Journal of Personality*, *34*(2), 164–179. https://doi.org/10.1002/per.2246

Zeinoun, P., Daouk-Öyry, L., Choueiri, L., & Van de Vijver, F. J. R. (2017). A mixed-methods study of personality conceptions in the levant: Jordan, Lebanon, Syria, and the west bank. *Journal of Personality and Social Psychology*, *113*(3), 453–465. https://doi.org/10.1037/pspp0000148

Zeinoun, P., Daouk-Öyry, L., Choueiri, L., & Van de Vijver, F. J. R. (2018). Arab-levantine personality structure: A psycholexical study of modern standard Arabic in Lebanon, Syria, Jordan, and the west bank. *Journal of Personality*, *86*(3), 397–421. https://doi.org/10.1111/jopy.12324