MASARYKOVA
UNIVERZITA

FILOZOFICKÁ FAKULTA

Functions of dance in nuptial rituals: test on ethnographic data, an evolutionary approach

Magisterská diplomová práce

DANIEL PECKA

Vedoucí práce: Mgr. et Mgr. Eva Kundtová Klocová, Ph.D.

Ústav religionistiky
FF N-RL_ Religionistika

Brno 2023

MUNI
ARTS

Bibliografický záznam

Autor: Daniel Pecka
Filozofická fakulta
Masarykova univerzita
Ústav religionistiky

Název práce: Functions of dance in nuptial rituals: test on ethnographic data, an evolutionary approach

Studijní program: FF N-RL_ Religionistika

Vedoucí práce: Mgr. et Mgr. Eva Kundtová Klocová, Ph.D.

Rok: 2023

Počet stran: 83

Klíčová slova: etnografické dáta, funkcie tanca, medzikultúrne databázy, medzikultúrny výskum, signalizácia, svadobné rituály, tanec

Bibliographic record

Author: Daniel Pecka
Faculty of Arts
Masaryk University
Department for the Study of Religions

Title of Thesis: Functions of dance in nuptial rituals: test on ethnographic data, an evolutionary approach

Degree Programme: FF N-RL_ Study of Religions - M.A. programme

Supervisor: Mgr. et Mgr. Eva Kundtová Klocová, Ph.D.

Year: 2023

Number of Pages: 83

Keywords: cross-cultural databases, cross-cultural research, dance, ethnographic data, functions of dance, nuptial rituals, signaling

Anotace

Táto diplomová práca skúma tanec ako nástroj signalizácie, ktorá je formou podvedomej neverbálnej komunikácie, v kontexte svadobných rituálov. Špecificky, štúdia skúma funkcie tanca v rámci týchto rituálov. Práca zároveň prispieva do evolučnej diskusie o tom, prečo sa behaviorálny prejav ako tanec objavil a akú rolu zohráva u súčasného človeka. Zvýšená pozornosť je venovaná medzikultúrnym databázam, ktoré sú zdrojom materiálu pre túto prácu, keďže ich vývoj a správne použitie sú predmetom aktuálnej vedeckej diskusie.

Abstract

This thesis investigates dance as an instrument for signaling, which is a kind of subconscious nonverbal communication, in the context of nuptial rituals. Specifically, this study examines the functions of dance within these rituals. At the same time, it contributes to the evolutionary discussion about why behavior like dance emerged and what its role is in contemporary humans. Increased attention is paid to cross-cultural databases that are the source of material for this work, as their development and proper use are the subjects of an ongoing scientific discussion.

Declaration

I hereby declare that the thesis titled **Functions of dance in nuptial rituals: test on ethnographic data, an evolutionary approach** that I have submitted for assessment is entirely my original work, and that no part of it has been taken from the work of others unless explicitly cited and acknowledged within the text of my thesis.

Brno May 16, 2023 .....

Daniel Pecka

Acknowledgements

I would like to express warm gratitude to my supervisor, Eva Kundtová Klocová, who guided me throughout this project. I would also like to thank my friends and family who supported me during the process. Without all of you, this work would not have been possible.

Table of Contents

List of Figures

List of Tables

List of Appendices

List of Supplements

Glossary

Abbreviations

d. o.	–	data omission
N/A	–	missing values
w/	–	with
w/o	–	without

Introduction

It is interesting that just being described as a dancer may increase an individual's perceived quality (Wade et al., 2015). What remains unclear, and is still subject to academic debate, are all possibilities that may lead to this change in perception. Is it that the dancer's physical quality is noticed more or recognized as better? Or is an individual seen as a more important, valuable part of the community, a person with whom it is worth improving relationships, spending time, and cooperating if they are a dancer? To contribute to answering similar questions, this research aims to engage in recently an actively discussed scientific dilemma on the functions of dance and articulation of its possible origin.

Dance may fulfill various functions. For example, through the release of substances (neurohormones, such as oxytocin, or endorphins) in the human brain, the performer of a dance may start to feel released, aroused, or excited while dancing (Tarr et al., 2014). Physical activity such as dancing may therefore alter a dancer's subsequent behavior. But dance may also function to alter the behavior of another organism by carrying a certain message (Maynard Smith & Harper, 2003). This study aims to improve knowledge about dance from the perspective of signaling theory, in other words, to test what kind of information is carried by the dance between individuals. This work will analyze functions of dance related to communication and will test hypotheses designed to indicate whether dance in nuptial rituals serves for courtship, specifically if it signals mate quality, or whether its functions are more relevant to social bonding or cooperation. The dance will be viewed as a signal or a cluster of signals whose presence in a specific set of conditions indicates its function.

This study will also contribute to the ongoing debate about the methodological challenges of using cross-cultural databases as sources of research material, especially if it includes narrative (unstructured) data such as ethnographies. This is important because these databases are becoming increasingly popular in research but there are still various issues related to how they are built, what kind of information they contain and how should those information be ideally processed. This work uses information from two databases which will be closer described later and comments on issues related to each of them. At the same time, some of the comments can be applied to cross-cultural ethnographic databases and research using them as a source of information in general.

In the first section of this work, I will introduce dance from the perspective of signaling theory, afterwards, the research regarding its functions will be discussed, and attention will be paid to dance as a part of a ritual. At the end of the section, individual hypotheses will be discussed and formulated. The second section will address the methods used in this work. The cross-cultural databases will be introduced, followed by challenges connected to using them as a material source. This section will be

concluded with a description of data collection and transformation. The next section, is the analysis. I will first describe and report the results of the source material analysis, to reflect its quality. Afterward comes the testing of the hypotheses, which is concluded by reflecting on the representability of the results. The last section contains a discussion of the results and issues connected to this study, including suggestions for improvement for future research.

1 Theory

1.1 Dance as a signal

From an evolutionary perspective, we cannot expect an organism to move if the advantages of moving do not outweigh the cost of that movement for that organism (Blythe et al., 1999). Humans, even though they can communicate verbally, transfer a large amount of information nonverbally, in the same way as the majority of known living species. This includes paralanguage, gestures, and body movements (Archer & Akert, 1977; Knapp, Hall, & Horgan, 2013; Wade et al., 2015). Dance is an energetically expensive body movement activity that is counted among the nonverbal communicative instruments. Specifically, it is viewed as a *signal* to which humans pay attention in many different contexts (Grammer et al., 2011; Hugill, Fink, & Neave, 2010).

From an evolutionary standpoint, dance is commonly approached through the lens of signaling theory. A signal is defined as a nonverbal method of communication. It is an action or a structure that has developed in one organism to alter the behavior of another organism and is successful because the recipient organism has developed a response to it (Maynard Smith and Harper, 2003, p. 15).

Dance, understood in such a way, appears both in humans as well as a multitude of other species. Among non-human life, typical examples include a male peacock showcasing his physical condition to a female, or an athletic "contest" between fruit flies (*drosophila subobscura*) of opposite genders, where males have to keep facing the females during the performance to be accepted (Johnstone, 1997; Laidre & Johnstone, 2013; Maynard Smith & Harper, 2003; Pecka, 2020). In humans, dance is an act, an expression, or a ritual that has a long history and a wide variety of forms and contexts in which it is performed around the world (Adshead-Lansdale & Layson, 2006).

Signals may not always be honest, but faking a signal always has its cost. The cost of a signal is what the signaler has to invest into the performance, to make the signal reliable, and therefore capable of eliciting the desired response in the receiver. There are three types of signals: indices, icons, and symbols. An index is an unfakeable signal, as its reliability is paid for during its development (Laidre & Johnstone, 2013; Maynard Smith & Harper, 2003, p. 45). The form of an index is determined by its content and its intensity is causally related to the quality being signaled (Maynard Smith & Harper, 1995). For example, a dancer's physical quality could be signaled by their visible body parts, and the better shape of the body, the better physical quality may be anticipated by the signal's receivers – the observers of the dance. Icon is a type of signal, whose form is similar to its content. Maynard Smith & Harper (1995) write that most icons used by non-human animals are used for cheating – to deceive the receiver or lower their fitness. However, I have failed to find any option for their usage in the case of human dance. Symbols are probably the most broad category of signals due to their

variability. It is because the symbol is only arbitrarily linked to the object it represents (Maynard Smith & Harper, 1995). An expensive dancer's costume may be a symbol of the material wealth of their family, and the color of this costume may be a symbol of allegiance to a tribe.

1.2 Anthropology of dance and its functions

One of the earliest anthropologists to study human dance, co-responsible for its formal establishment as a subdiscipline of anthropology, was Gertrude P. Kurath (1960). Her accounts of dance were largely ethnographic. She observed that dances of couples with physical contact are for the most part a behavior developed in the latest centuries, and largely of European origin. She also noticed that while there may be intimate contact in social dances while ritual dances often separate men and women into groups, in neither are genders described as separated altogether. Dance's choreography may also reflect various realities, such as hierarchies by economic or social status or the way people are grouped in society (p. 237). But what she argued for is that while dancers may carry out individual roles, dance mostly takes place in a group (p. 236).

Generally, dance movements and their combinations (choreography) were commonly subject to misinterpretations. They were often studied using linguistic approaches that tried to identify the meaning of every specific movement[1] and map these movements into a system that can be used similarly to the way we use language for communication (Kaeppler, 1978). Kaeppler (1978) criticized many of such research's early assumptions, but her critique was often aimed at attempting to find some universal language in choreographic movements. Her approach was also linguistic, however, in line with the current understanding of dance, she emphasized that since prehistoric times dance has been present in both courtship activities and rituals as a form of social communication (pp. 43-45).

Hanna (1987; 1988) asserts that dance may function to communicate different information, and throughout her work, she highlights its importance for courtship and courtship rituals. Her work is largely ethnographic, while she often makes her assumptions from the perspective of evolutionary psychology. Hugill, Fink, and Neave (2010) summarize Hanna:

> *[...] points out that dance may be an adaptive behavioral pattern in sexual selection, as it is a medium, which displays beauty, health, strength, and thus sexual attractiveness. It is possible, that the disposition to dance facilitates human mate selection, though the variation in dance patterns and styles is certainly mediated by culture. (p. 74)*

[1] These were usually defined differently by every researcher.

Her work could be described as involving a comprehensive analysis of dance from an evolutionary perspective. However, this explanation of dance was found insufficient by Hagen and Bryant (2003) who argue that music and dance (also) signal coalition quality. Their main argument is that a group with low coalition quality (e.g., newly formed groups) would not be able to create and perform complex music and dance as a group as their internal stability would be low. The group would have trouble synchronizing and cooperating on a larger scale (p. 30). Even though they brought empirical evidence only for the musical part of their argument, their work served as inspiration for more recent studies. The connection between dance and music, or dance and rhythm is hard to question, and there is definite strength to their argument, claiming that the sexual selection hypothesis cannot explain every single aspect connected to dance performances.

More recent studies commonly propose that dance may function as a way to increase group (social) cohesion and may signal group belonging, or signal commitment to a group. It is considered to be a by-product of imitative proficiency and its likely dominant function may be social bonding facilitated through synchronized movement (Laland et al., 2016; Reddish et al. 2013; Tarr et al., 2015; Wildermuth & Heath, 2009).

The proposition that dance functions to carry information about an individual's quality (physical or other), in other words, is relevant for courtship is also supported by several studies (e.g., Brown et al., 2005; Hugill et al., 2010; Hugill et al., 2011; Hugill et al., 2009; Röder et al., 2016). By some, sexual selection is considered to be one of the main reasons for the origin of dance (Fink et al., 2021; Fink & Shackelford, 2017).

1.2.1 Dance, physical quality and, courtship

Researchers showed that hiding bodily curves results in lower quality ratings of women by men, while more covered women also get less attention from men (Pazhoohi & Hosseinchari, 2014; Pazhoohi, Macedo, & Arantes, 2017). At the same time, Röder and colleagues (2016) show that men pay more attention to attractive women dancers and judge them more favorably. Unfortunately, the scope of research examining human female signaling to males[2] is rather limited, while the research employing the opposite perspective is considerably more extensive.

Sensation seeking (such as risk-taking), which is viewed as a costly behavioral trait that signals man's health and vigor, appears to be signaled through dance. Women rate men as more attractive if they score better in sensation seeking (Hugill et al., 2011). Hugill and colleagues (2009) also show that there is an association between women's rating of men's dancing ability and men's physical strength, in other words, women perceive men as better dancers if men are physically stronger. But this might

[2] In this work, while using the terms male and female in a human context I refer to biological, not social categorization. The term is used interchangeably with man and woman, respectively.

not work in the opposite direction as there seems to be no such association when men rate women's dance attractiveness. Brown et al. (2005) show how dance may reveal a perceived dancer's quality. In their study participants evaluated symmetrical[3] men as significantly better dancers compared to asymmetrical men. Similarly, symmetrical women dancers were evaluated as better dancers compared to asymmetrical women, which confirms that attractivity is rated by both genders in dance. However, symmetrical men were rated as significantly better dancers compared to symmetrical women. This research indicates that women may value the dancing capability of men more than it is the other way around. Such an interpretation would be in line with the evolutionary consensus saying mammal females are choosier in mate selection compared to males (Geary et al., 2004).

The research on quality signals in dance is currently not too robust, but when the evidence grows it may turn out that women observe more physical traits, or more specific traits in men than men observe in women while dancing (e.g. the case of Weege et al., 2015). If this assumption is correct, it would make sense for the presence of men's dance to be more strongly associated with indicators of physiological communicative capabilities of dance, compared to the presence of women's dance. However, the imbalance in the quantity of evidence for different sexes may also mean that these assumptions are rising due to the lack of research on female signaling. Recent research points out that patterns of female signaling among different species are complex (Wallner et al., 2019; Zweerus et al., 2022), and this complexity is not reflected in past studies. Future research must focus more on female signals observed by males to clear this doubt.

1.2.2 Dance, social bonding, and cooperation

According to Reddish and colleagues (2013), collective activities such as rituals, that include music and dance performances may play an important role in promoting synchrony and cooperative behaviors. The authors investigated the relationship between shared intentionality, synchrony, and cooperation. They show that synchrony combined with shared intentionality resulted in the highest level of cooperation. Tarr and colleagues (2015) investigated the bonding effects of exertion and synchrony. Their findings indicated that both physical exertion and movement synchrony had a significant positive effect on in-group bonding, implying that dance involving both exertive and synchronized movement can be an effective way to induce social bonding. According to the authors, endorphins, which are associated with pain relief and reward, may

[3] Researchers use an indicator of quality called the degree of fluctuating asymmetry "which is inversely correlated with the degree of developmental stability, which is an organism's ability to reach an adaptive endpoint despite ontogenetic perturbations" (Brown et al., 2005, p. 1148). Therefore, the more symmetrical an individual is, the healthier and better-developed physique they should have.

also play a role in social bonding during group dancing. Wiltermuth and Heath (2009) investigated how synchronous activities like marching, singing, and dancing can improve social bonding among group members. According to the authors, synchronous activity can weaken psychological boundaries between the self and the group, resulting in increased social attachment and cooperation. People who participated in the synchronous activity with others were found to be more likely to cooperate in subsequent group economic games, even when personal sacrifice was required. The authors claim that positive emotions are not required for synchrony to foster cooperation, implying that simply acting in unison with others can increase social attachment and cooperation among group members. However, Lang and colleagues (2017) show that the opioid system is activated during a cooperative exercise, while they expect β-endorphin release based on tests previously done on non-human animals.[4] This on the contrary indicates that positive emotions should affect the cooperation facilitated by the movement.

Laland and colleagues (2016) write that dance has symbolic properties, which are dependent on the dancer's ability to imitate specific people, animals, or events, as well as the audience's ability to recognize these correspondences. Authors suggest that dancing may only be possible for humans because it uses existing neural circuitry used in imitation. Imitation is important in learning to dance, and the acquisition of long sequences of choreographed movements is reliant on social learning, which implies that dance is rooted in social interaction. It appears that their definition of dance excludes instinctive movement that is typical for other animal species. The purpose of such movements is often clearly linked to courtship and reproduction. As noted above, we can observe these correspondences between many other animals, with humans currently being the only known species where dance appears to have several other (social) functions. Dance may have further evolved in humans (Fink et al., 2021), gaining different forms and variations in function which are linked to the context in which dance appears.

To conclude, synchronized movement plays a role in human dance and there is evidence that dance could be relevant for courtship as it is with other animal species. I find it important to note here, that the most often mentioned characteristic of dance – movement – is not the only aspect of its communicative capability, even though this research emphasizes it. For example, a dancer's costume design may symbolize their social status or role, their material wealth, or allegiance. All of the above-mentioned may be connected to a certain belief or a historical occasion. Therefore, dance may carry several different information as a single complex activity. Reproduction and social bonding are the functions that have been the topic of recent scientific debate,

[4] Exact measures of chemical agent levels released by the opioid system were never done on humans as it is currently considered infeasible.

intending to determine which is dominant and which should be considered to be found liable for the appearance of dance.

1.2.3 Dance in a ritual

For several functions mentioned in this chapter, rituals may be either the context in which the information is carried or the instrument which carries the information (Rappaport, 1999, pp. 24-25). In the latest review of research regarding dance from an evolutionary perspective, Fink and colleagues (2021) consider that its ritualization, or adaptation in human rituals was the point in evolutionary history that made dance gain more social importance. Until this occurrence, dance was an activity of individual characters whose main purpose was courtship. In other words, it may have evolved for purposes of sexual selection and after becoming a group activity, it gained several other functions, such as enhancement of group cooperation, or social cohesion. Importantly, the authors note that most studies of dance from the perspective of evolutionary psychology focused on individual dance performances including participants from only a few industrialized societies. It is also not easy to gain visual material from pre-industrialized societies and ethnographic texts often do not describe dances in sufficient detail. Therefore, it will take further effort to assess whether their results can be generalized across past time periods (p. 357). At the same time, it also shows that evidence pointing towards dance as a behavior generally important for sexual selection may not be as strong as it appears at first sight. This causes a major problem in evaluating whether courtship plays a significant role in dance after it became an activity performed by groups in a ritual context.

1.3 Hypotheses

This research aims to examine the functions of dance. As was described in the previous parts of this section, currently available research results can be clustered into two major propositions. First is that dance functions to transfer information relevant to courtship such as biological quality signaling, the second is that dance functions as a tool for the improvement of social connections, such as cooperation in a group or social bonding. Fink and colleagues (2021), summarize research done until recently and suggest that each of these propositions is right to a certain degree. They say that dance has likely evolved for courtship (same as in other animal species), but it changed in humans with ritualization and became also a social activity. In my research, I aim to examine whether data available about ritualized dance suggest that it functions for courtship or otherwise. To explore, whether courtship is at least partially important in ritualized dance, eventually indicating whether the proposition by Fink and colleagues may be

true or not. In this part of the text, I describe the reasoning behind all tested hypotheses.[5]

Men and women face different reproductive problems because of their biology, which led to the development of various sexual selection strategies. The biggest difference between men and women is seen in men's enormously greater reproductive potential. Men's sperm production is not fixed and replenishes at a pace of roughly twelve million per hour, whereas women create a fixed and non-renewable lifetime supply of about four hundred ova. This indicates that for men, one sexual act poses little danger of future investment, in contrast to women, who may become pregnant. Pregnancy (gestation) is an energy-intensive investment that temporarily limits alternative opportunities for reproduction (in humans at least 40 weeks, usually longer). It is followed by birth (as is the case with mammals in general), and a breastfeeding period that can last up to three or four years (Buss, 2003, pp. 19-20; Pecka, 2020, pp. 26-27). As a result of this pressure, women are perceived as the pickier of the sexes, whereas men are perceived as those who typically compete more for mates (Geary et al., 2004).

It could be expected that in polygynous societies, competition among men is higher as one man may be a partner to several women. This may limit access of other men to women, considering the sex ratio is not skewed in the population. Polygyny may, therefore, allow some men to achieve greater reproductive success than monogamy, while restricting other men's reproductive opportunities. This should result in a higher need to display men's eligibility to be partners since the competition over mates grows. Therefore, if ritual dance fulfills the function of signaling information relevant to sexual selection (such as intersexual quality signaling), men should take part in it more often in societies that have a higher degree of polygyny compared to societies where it has minimal presence, or none at all.

- Hypothesis 1: *In polygynous societies, the custom of men dancing is present in nuptial rituals more often than in non-polygynous societies.*

Existing research shows that there is an important factor to be observed and that is the difference between societies with higher and lower levels of social stratification. Material wealth is an important factor in mate choice as it also indicates an individual's quality (Buss, 2003). And according to Gaulin and Boster (1990) if a society is stratified we may expect wealthy men to outcompete poorer men, through their material resources (pp. 995-997). This may happen for various reasons even if the poorer man has an advantage in the scale of his other qualities. As the direct reproductive capacity of a woman is biologically greatly restricted, they may increase their reproductive success through their offspring substantially. Material wealth is easily transferable, and a woman's potential son may be able to provide for more mates to produce grandchildren. That is, if she marries a wealthier man, even though her husband's other qualities

[5] Exact hypotheses formulations are numbered (1-4) and *highlighted* below.

are lower compared to the rest of the possible partners. Therefore, in polygynous societies that are socially stratified, the custom of men dancing should be less frequent compared to non-stratified polygynous societies, as signaling of men's qualities through dance would have substantially reduced effectivity by differences in wealth among their co-competitors.

- Hypothesis 2: *In stratified polygynous societies, the custom of men dancing is less frequent in nuptial rituals compared to non-stratified polygynous societies.*

Dance adaptation in ritual circumstances where courting behavior is common may be anticipated. When different families participate in a ritual in an exogamous society the appearance of dance that conveys information about physical quality can be expected. This is because such an occurrence may present a valuable opportunity for courtship for persons living in an exogamous society. In the case of endogamy, it cannot be assumed that a similar condition will increase courtship possibilities because they have greater courtship opportunities without leaving the boundaries of their community (e.g., village). In an interfamilial context, dance should be more likely to convey information relevant to courtship, but in an intrafamilial scenario, it may be expected for dance in nuptial ritual to serve mainly for another purpose. Nonetheless, we may assume that with increasing exogamy, the likelihood of dance functioning to convey courtship-related information grows. If society's marital structure is exogamous, men should dance more in nuptial rituals.

- Hypothesis 3: *Men dance in nuptial rituals more when the degree of exogamy is higher in society.*

Furthermore, it is important to take the living environment into account. According to Gangestad and Simpson (2000), several shifts are going on when an environment is of harsher or more demanding nature. In such conditions the need for biparental care increases, monogamy is preferred, men make more efforts to invest in their offspring and women are less likely to engage in short-term and extra-pair mating. But as women's ovulation is concealed from men, and they are fertilized internally, men can never be certain of their paternity (Buss, 2003; Davis & Daly, 1997). Pazhoohi and colleagues (2017) suggest that harsher environments may result in the presence of stricter cultural provisions such as religious mate-guarding (e.g., religious veiling), which may also be enforced through religious institutions. Therefore, if dance is important for courtship, by for example signaling the quality of an individual, it should be expected that the tradition of women dancing should be less common in rituals of cultures based in harsher environments.

- Hypothesis 4: *The custom of women dancing as a part of nuptial rituals is less frequent in cultures located in harsher environments.*

1.3.1 Specific research context

Following the arguments of Fink and colleagues (2021), it should be considered that in the context of a ritual the function of dance shifts from courtship to other functions, such as social bonding. To test this, aside from the hypotheses outlined above, I chose a ritual setting of nuptial rituals. Marriage ceremonies take place in every human society (Carroll, 2016, p. 219) and despite their differences may be seen as a human universal (Burch, 2019; Pecka, 2020) and dance, in various forms, is commonly their part (Monger, 2004).

At the same time, the symbolic contents of nuptial rituals are predominantly related to reproduction, nuptials are perceived as rituals of clear reproductive nature (Buckle et al., 1996, Burch, 2019). Symbols of virginity or purity, such as the public display of a bloody handkerchief after the first couple's intercourse, are commonly observed, as are symbols of mate quality, which may take the form of numerous forms of displays, such as those of an individual's good physique, material wealth, or social status. Wearing a ring since betrothal or after marriage is a universal symbol of commitment to a mate, and there are many other practices similar to those mentioned (Buckle et al., 1996; Burch, 2019; Pecka, 2020). The context of nuptial rituals should therefore increase the chance that the results will support the relevance of dance for courtship. At the same time, this decision gives the research more control over unpredictable influences that could be present if the tests were run on dance in general, as it is an activity that takes place in many situations and contexts.

Testing hypotheses in the nuptial ritual context could also help with the interpretation of results. The evolutionary research of dance is still growing and there are multiple methodological issues that the scientific community has yet to resolve. These issues may create doubts regarding the reliability of research results and will be discussed in more detail in the following section. Using data from a context that is intensely focused on reproduction may help more broadly answer one of the research questions which is whether dance functions for courtship or some other social function. Potential results supporting the courtship position would be relevant for this specific context. At the same time, potential results supporting the position that dance functions for other purposes may indicate that dance functions as such even outside of the nuptial context as this context is strongly connected to reproduction.

2 Methods

To examine previously discussed questions, I use narrative ethnographic information coded by myself and two other students (fully crossed coding), and data pre-coded by other researchers. Hypotheses tested along with their exact formulations were described above, while remarks on model fitting and selection are part of the analysis. This section introduces the data sources and discusses several methodological issues related to the method of cross-cultural research. It describes the data acquisition, transformation, and overall preparation of the data for hypothesis testing.

All calculations, graphs, and tables included in the text are done using the R programming language in the RStudio environment (RStudio Team, 2020; R Core Team, 2021). R packages used are always referenced in the relevant part of the text.

2.1 Cross-cultural databases: An introduction

To shed more light on the role of courtship in the group dance, I use cross-cultural data gathered from Human Relations Area Files (eHRAF, electronic version) and Database of Places, Language, Culture and Environment (D-PLACE; Kirby et al., 2016). In this chapter, I will introduce the databases and their advantages and in the following chapter, I will discuss challenges related to their usage.

eHRAF is a cross-cultural database that provides access to information on cultures all over the world. The main advantage of eHRAF is that it provides researchers with an easily searchable collection of ethnographic materials. The database is organized into a series of ethnographies, each of which focuses on a particular culture or cultures. The database contains a large amount of unstructured data, such as text descriptions of the culture's history, social structure, economics, and religious beliefs, which may be supplemented with visual materials such as illustrations, photographs or maps.

eHRAF allows users to search for information about specific cultures, cultural traits, and topics of interest. In the advanced search, users may select the cultures they intend to study. Cultures are marked by an identifier called Outline of World Cultures[6] (OWC; e.g., FO04 Mbuti)[7] which makes it easier to link cultures between samples and different databases. This is especially important as one culture may be listed multiple times under the same name while the material refers to different times and locations (Ember, 2007). Cultural traits are categorized using the Outline of Cultural Materials[8] (OCM; e.g., 585 Nuptials) and may also be selected by the user in the advanced search.

[6] Originally developed by Murdock (1969).

[7] FO04 is the OWC identifier for culture Mbuti.

[8] Originally developed by Murdock in the 1940s (https://hraf.yale.edu/resources/reference/outline-of-cultural-materials/; Murdock, 1961).

The search also allows users to enter keywords regarding their topic of interest that will be returned from the texts corresponding to the previous search criteria. Selection may be further specified using Boolean operators. Search results are returned in the form of paragraphs from relevant documents organized by culture. They may be subsequently filtered according to culture's subsistence type, region, date when the material was published, culture level samples and document level samples, document type, and series.

In contrast with eHRAF, D-PLACE is based on structured data. It provides information on over 1400 societies from around the world coded into datasets and variables by a large number of researchers. Information typically found in D-PLACE are related to demography, language, subsistence, social organization, beliefs and values, material culture, or environment in which societies live. It is also worth noting that, aside from tabulated data, D-PLACE provides several phylogenetic trees.

Societies in D-PLACE are described using a set of standardized variables. Aside from society's name and identification code, there is also glottocode, language, language family, latitude and longitude, region, and societyset[9], which may be used to compile datasets for various analytical purposes. D-PLACE societies are also directly linked to eHRAF cultures using the OWC identifier. This versatility makes it a valuable resource for conducting cross-cultural quantitative research. It may decrease the resources required for research as it lowers the necessity for manual coding of information that was coded and published by others before.

2.2 Challenges and issues of cross-cultural databases

Along with (quantitative) ethnographic research, there is a discussion about how much information we have, how biased it may be, and how much misinterpretation coding of narrative materials may cause. This chapter will address specific problems concerning my research.

The overarching question is what scholars and travelers were able to record about the cultures' lives, traditions, and behaviors, and how much of this material accurately represents their lives. It is critical to recognize that cultures may not want to reveal everything about themselves. They may purposefully limit outsiders' access to some of their customs. It could be for protection or a completely different reason, such as collector's gender and taboo on some topics. Male ethnographers were most likely denied access to some rituals performed privately by women as these rituals are inaccessible

[9] Users may use filtering, or pair datasets based on previously created cross-cultural sets: Binford Hunter-Gatherer (Binford, 2001), Ethnographic Atlas (orig. publication Murdock, 1967a; rev. Murdock et al., 1999), Standard Cross-Cultural Sample (Murdock & White, 1969) or Western North American Indian database (Jorgensen, 1980). Ethnographic Atlas and Standard Cross-Cultural Sample are also among the sampling options in eHRAF.

to men even from their society. This is a crucial point as the majority of material collectors were men. We may have a large corpus of narrative descriptions of cultures, but we don't know how comprehensive or exhaustive it is, and the level of quality varies from document to document.

In a chapter of his *History of Religions*, Jonathan Z. Smith (1971) lists several topics from Herodotus' Ionian ethnography accompanying it with an observation. He writes:

> *What is being collected under these headings for some fifty cultures is basically a set of traveler's impressions. Something other has been encountered, and it is surprising either in its similarity or dissimilarity to what is familiar "back home". (p. 73)*

Smith precedes this text with a discussion on how it is natural for humans to understand the world through comparisons as they need to give it some order. We especially use some scale of likeliness to ourselves (as individuals and as groups) (pp. 67-71). To some degree, this way of interpreting the world affected ethnographers at all time periods, no matter if they were trained to collect information in the most unbiased way achievable. This inherently present dichotomy of self and other in human thinking is not reflected in many databases in any standardized way.

It is up to researchers working with those records to assess whether any given material is acceptable as evidence. This step may be difficult as materials vary in length from a few pages up to hundreds. Materials are coming from authors with all kinds of backgrounds or roles in the cultural group (e.g., missionaries' records, anthropologists' records) and metadata often does not assist the researcher sufficiently to reflect material quality effectively (Watts et al., 2022). It may take a lot of effort to determine to what extent the ethnographer used the scope of their cultural background to interpret what they were observing, and how much it could have offset their record from the observed reality. It is, therefore, no surprise that one of the commonly raised criticisms of research using narrative data is the accusation of ethnocentrism. Even more so, when these kinds of records are consecutively coded by another person or people who might need to further infer their meaning. For now, eHRAF is the only database that provides metadata that could help determine material quality effectively. For each document, they list the author's background, or more exactly their work field, and provide the document with an evaluation score in the range of 0 to 5. This may be of help, however, there is still a lot of work to be done and some clarifications need to be made from the side of HRAF management. They provide users with information on who rated individual documents, however, it is unclear what were the criteria of their evaluation.

2.2.1 Validity of ethnographic materials

In an attempt to assess the validity of ethnographic materials, Bahrami-rad, Becker, and Henrich (2021) contrasted the Ethnographic Atlas with the Standard Demographic

and Health Surveys (DHS; collected since 1984). They have discovered positive associations between the DHS's self-reported individual-level data and the ethnographic data from the Atlas. They claim their research shows that ethnographers' accounts of pre-industrial society life as portrayed in the Ethnographic Atlas are informative across a range of dimensions. At the same time, they found a significant association in most dimensions they examined.

The authors did not cover a large number of topics; rather, they focused on a few general cultural traits. Those are twelve domains that were equivalently represented in both, the DHS and the Atlas: (1) patrilocality, (2) matrilocality, (3) polygyny, (4) reliance on animal husbandry, (5) reliance on agriculture, (6) length of postpartum abstinence, (7) breastfeeding duration, (8) insistence on virginity, (9) a preference for sons, (10) prevalence of domestic violence (11) age difference between husband and wife, and (12) geographical location.

Some of the topics covered appear rather straightforward to measure or record, however, it may be more complicated to gain reliable information regarding points 6, 7, 8, and 10. These are topics people might not wish to speak about, information may be kept actively hidden from the public or outsiders, people may have a motivation to be dishonest regarding these topics, or access to information may be problematic to gain due to gender taboos. It is not the case with DHS because there are publicly available materials that may help assess the reliability of data and in case some methods are found unclear, it is possible to contact the collectors and account for the issue in subsequent analysis. However, this is problematic in the case of narrative ethnographic records as there was no standard for data collection and we lack information on how it was done.

The fact that we lack the means to assess the accuracy of ethnographic records creates doubt about whether some questions can be answered using narrative ethnographic data such as those available in eHRAF. Some practices may be more difficult to understand or more prone to misinterpretation and researchers must be careful how they formulate questions that should help explain the behaviors. In the case of rituals, it may certainly be easy to record the ritual frequency, gender, or age of ritual specialists or the age of participants, for example, in the case of transition rituals. But it is way more problematic to research the ritual meaning as it may be different for each individual and the ethnographer most likely did not ask a representable sample for the meaning. The records often rely on one or a few informants which may cause exaggeration or understatement of some behaviors such as the above-mentioned topic of domestic violence. It may be hard to quantify such behavior based on narrative descriptions if the descriptions are not too detailed or are based on a single or a few cases. If the frequency or brutality of such behavior is not properly reflected in the record it deepens the issue. And it is hard to tell whether a record based on direct observation or an informant's description is more reliable as it may vary from topic to topic and

from one informant to another. The more sensitive the topic, the harder it may be to trust the record's accuracy.

To come back to the mentioned research there is one other issue aside from the reliability of the narrative material. If the observed culture had significant contact with Western cultures following the ethnographer's visit, some shift toward Western interpretation of their own culture likely occurred. Moreover, the culture could have had contact with Western cultures before the ethnographer's visit. The Western perspective may have an impact on both initial ethnographer recordings and self-reports in the DHS. Such influence may have resulted in better associations between the samples than in a case where either any significant Western influence on self-reports would have been confidently dismissed, or the initial record could have been accepted as having sufficiently small bias. It is up for discussion (and future research) how much the self-reporters' understanding of themselves may become Westernized in the period between the ethnographers' recordings and their self-reports. Especially, if the variables are of a kind that requires some interpretation to be measured.

2.2.2 Using precoded variables carelessly

Another issue may arise when variables are taken from previously published research without sufficient analysis of the methods and theory used by that research. Lightner, Bendixen, and Purzycki (2022) describe how a variable of moralizing *high gods* was used in several papers as a proxy measure of the presence of *moralizing gods* and led to false outcomes. First, they discuss the coding criteria for the high god variable, which determines whether or not a god is the creator or director of the universe, but does not account for the power or omniscience which are considered features of moralizing gods. The authors then illustrate how these criteria, which have no relevance to the question of whether gods are moral or punishing, have caused researchers to consistently produce false negatives, which lead them to conclude that moralizing gods are not present in societies because moralizing high gods are not present. They also found that researchers are more likely to infer false negatives when studying social complexity at lower levels using datasets that include both moralizing gods and moralizing high gods. This resulted in producing illusory positive correlations between variables of social complexity and the presence of moralizing gods.

Every time researchers use data from someone else's work, or sources other than their research, it is important to examine their background properly. Especially when it comes to data that were not quantitative when originally collected, because such data require further inference for conversion into a quantitative type. Slingerland and colleagues (2020) expect some data from qualitative sources such as narratives can be coded with a large certainty and require little inference. To avoid above described mistakes, I followed their argumentation, and for my precoded variables used ones that

could have been coded with high certainty and could be expected to require little inference.

2.2.3 Time foci problem

Another issue connected to cross-cultural databases is the problem of time foci. Researchers note that differences between the times of records may increase random error, which almost always leads to a reduction in correlations (Divale, 1975; Ember & Ember, 2001; Ember et al., 1991). Records in eHRAF have two relevant timestamps, which are *field date* and *date coverage*. The former refers to the time the ethnographer was in the field and the latter refers to what time period is described in the material. In the eHRAF data for my research (see *suppl. 1*) metadata about field date is missing for 20 and metadata about date coverage is missing for 102 out of 247 paragraphs, while the dates are sometimes only approximate (e.g., *ca. 1660*), or cover a very large period (e.g., *ca. 1800-1950* or *ca. 1500-1800*). This may cause distortion when it is unclear which time period is in question if, for example, an informant tells the ethnographer about past events or customs of their ancestors. Such records complicate the research process as they may contain useful information, but may also cause noise in the analysis. Omitting such records from the analysis may be problematic as there is a large proportion of such material. A hint of a solution is again provided by already mentioned research by Bahrami-rad, Becker, and Henrich (2021) whose comparison shows that information from the Ethnographic Atlas may still be useful despite generational time differences. However, as they only focused on a few general cultural traits, it is impossible to take their results for granted. Similar research would be necessary for a wide range of variables to specifically address which of them may be understood as transferable through time periods, and what would be the period extent.

Due to the amount of metadata missing, this research resigned on attempting to control for time periods and instead accounts for possible error when interpreting the results. Limitations caused due to the problem of time foci are further addressed in the discussion.

2.2.4 What data may be of use?

Even though Fink and colleagues (2021) are right about ethnographic texts being often insufficiently detailed in the way they describe dance performances and suggest that future studies take an approach that is highly dependent on the capture of visual material (pp. 357-358), I think ethnographic texts may still be used to indicate whether there is a significance to courtship signals in ritual dance. The often relatively vague description of dance contains some information that may serve to gain better insight into the existing knowledge regarding the strategies of sexual behavior.

Information on whether dance is present in certain contexts may be valuable. For example, if it is present in a ritual that is shared by different exogamous families or

villages, it can be understood as an opportunity for courtship. Similarly, knowing whether men or women engage in the dance (if there is one) and whether both genders are present and may be viewed as witnesses of the dance are details that may be useful and are possible to be gathered even from the less detailed descriptions of dance in ethnographies. In texts, I've worked with (see *suppl. 1*), ethnographers frequently refer to the gender of the dancers and those present, even though it is not always entirely outright. If they do not mention genders explicitly, ethnographers usually give information such as "all the villagers take part in the dance", which implies that both sexes are taking part, or "the entire village participates in the festivities", which implies that members of both genders can be regarded as observers.

2.3 Data collection and transformation

As I have mentioned previously, data for this work were collected from two cross-cultural databases, eHRAF and D-PLACE. From eHRAF I have collected unstructured data that were coded specifically for this research, while from D-PLACE I have used pre-coded variables to enrich the dataset and gain the ability to analyze the problem in more depth.

To gain data whose results may be generalized across societies worldwide and are at the same time compatible with already existing datasets, the paragraphs from the eHRAF search were filtered according to the Standard Cross-Cultural Sample (SCCS; Murdock & White, 1969). Ethnographic Atlas (EA; Murdock et al., 1999) describes practices of more than 1250 societies, SCCS is a sample drawn from the EA's societies that had the fullest ethnographic coverage to make sure no major cultural variant[10] was overlooked. The well-described societies were divided into clusters, which are groups of societies with cultures so similar that no global sample should include a single one of them. The similarity may be due to diffusion or a recent common ancestry (Murdock & White, 1969).

Filter was further combined with Probability Sample Files (PSF; Naroll, 1967). The PSF was developed to provide a sample of usually pre-industrial, randomly selected cases from the 60 world-culture areas.[11] To minimize the effects of historical

[10] "Cultural variant" is a term used by Murdock and White (1969), it is not entirely clear what it refers to, but based on Murdock's previous work I assume it refers to the categorization of behaviors in the Outline of Cultural Materials mentioned above. Such as: no major *cultural variant* of *Nuptials* should be overlooked.

[11] These areas are different from the clusters mentioned with SCCS but they overlap and share many societies making them pairable.

relatedness (Galton's Problem),[12] only one case was chosen from each area, and the selected case had to meet certain data-quality control standards (Ember, 1997).

2.3.1 Self-coded variables: eHRAF

To search for relevant materials, that is all mentions of dance in nuptial rituals that are contained in the above-specified sample, I have used advanced search options in eHRAF. I have searched among *all cultures*, under the subject *585 Nuptials* connected by the boolean operator *AND* with the keyword *danc** (dance). As I wanted eHRAF to return every result that fit the rest of the criteria I applied *Any of these Cultures* setting to the culture tab. Similarly, I have used specified settings *Any of these Subjects* and *Any of these Keywords* in the other criteria as I have searched only among one subject and one keyword. The keyword *dance* was written with the asterisk[13] in place of the last character to get results containing all modulations of the word, such as danc[ed], danc[ing], or danc[ers]. The result in eHRAF is generated in the form of paragraphs corresponding to the specified search.

All paragraphs that described cultures present in both SCCS and PSF samples were extracted along with other relevant data and metadata. Culture's name and eHRAF number of paragraph[14], author's name, publication name, the section of the text (such as book chapter or similar), page(s)[15], the publication's name, field date, and date coverage, author's background, document evaluation, comment, and the paragraph text itself.

As there were often paragraphs from the same ethnography following one after the other or with some text separating them, I have always extracted this passage and noted it in a comment or directly in the text using square brackets. As I found out during the data collection process, the way eHRAF separates texts into paragraphs is not always ideal and in several cases, I had to add surrounding paragraphs to provide context for better coding decisions regarding dance. This was also noted in the comment per each relevant paragraph. All mentioned information is compiled in the data collection table (see *suppl. 1*).

As in some of the datasets from D-PLACE the Ojibwa[16] culture is divided into several regions, I have looked up in which regions were the paragraphs listed in eHRAF

[12] This was also attempted in the creation of SCCS, but PSF sampling is more strict in the comparison of cultural similarities, resulting in a smaller set of cultures (60). SCCS's "sampling provinces" created based on linguistic and cultural evidence for the same purpose have more connections among themselves resulting in a sample of 186 societies.

[13] Asterisk functions as a wildcard character that matches any number of characters in a word.

[14] This was relevant for referencing in the coding process.

[15] If the text appears on the page break or multiple pages, the range is listed. At the time of data collection, eHRAF only listed the page where the text starts.

[16] In eHRAF it is uniformly listed as Ojibwa, with no regard to the region.

recorded and added the specification into the data. Originally eHRAF displayed all of them as "Ojibwa", but this culture is regionally very diverse and is present in quite different locations. The four Ojibwa mentions in eHRAF were changed and recoded in the rest of the data as (1) Minnesota Ojibwa (even though the description in eHRAF mentions that it was also recorded in Wisconsin and Michigan, according to the book contents the vast majority of material came from Minnesota), (2) Eastern Ojibwa (according to the D-PLACE society map there are Eastern Ojibwa only in Michigan, where this was recorded according to eHRAF description), (3) Round Lake Ojibwa (a clear reference to Round Lake in the eHRAF description of the material), (4) North Albany Ojibwa (note in eHRAF refers to northeastern Ontario where D-PLACE localizes Northern Albany Ojibwa).

After the data was collected, it was coded by myself and two bachelor's student coders. Coding was done using coding sheets containing the culture's name and paragraph number, the comment made during data collection and the text to be coded. For this purpose, a coding manual (see *suppl. 2*) was created where it was defined based on what information coding decisions should be made and what values are attributed to them. Definitions were accompanied by general references to some situations that occurred in the text and could complicate decision-making if a coder who is not familiar with the research had to make decisions about them.

Because the same ethnographic material was fully coded three times by three different coders,[17] a decision had to be made about which coded values would be used for analysis if the coders disagreed. The method chosen was to use code chosen by the majority of coders (2/3). If there was no overlap, I made a decision based on the coder's comments. If there were no comments, the paragraph was reconsidered, and the final decision was made by myself (see *suppl. 6* for comments on all records where any of the decisions occurred and *appx. A* for a detailed description of the process).

After the decisions were done some paragraphs which contained information about the same ritual were merged (see *suppl. 7*). This was done to limit duplicates in the data, specifically paragraphs that followed directly one after another or were separated by a short text[18] were merged into a single coded value. In cases where paragraphs appear to refer to the same ritual but carry different codes, paragraphs were kept in their original form as they may refer to two different rituals that are part of a single ceremony and this may not have been clearly distinguished by the author of the material. There was, however, one exception to this, and that was cases when paragraphs differed only by a missing value (or values). In this case, paragraphs were also

[17] See *suppl. 3*, *4*, and *5*.

[18] As mentioned above, if there is the separating text specified in the comment or brackets, it means coders were provided this information in their blank coding sheets.

merged, as the paragraph with missing value is perceived as supplementary information for the complete paragraph.[19]

Lastly, the merged coding sheet was recoded into binary values (see *appx. B* and *suppl. 8*) creating a dataset that was enriched by variables of region and subregion of cultures (eHRAF classification). Data from D-PLACE were later added and a final dataset for hypothesis testing was created (see *suppl. 13*).

2.3.2 Pre-coded variables: D-PLACE

From D-PLACE I have used five variables, which are all compatible with the SCCS. To gain data about the presence of polygyny in cultures I have used a pre-coded variable *Standard Polygamy Code: Whyte, Murdock and Wilson, Murdock Atlas [SCCS861]*[20] that originated in White, Burton, and Whiting (1986). This categorical variable was recoded from the original five values ranging from "*Polyandry*" to "*Full polygyny (20% or more of married males)*" to three values "*non-polygynous*", "*minimal polygyny*" and "*polygynous*" to make the result interpretation clearer relative to the hypothesis formulation (for details regarding data transformation see *appx. C* and *suppl. 9*).

For information about social stratification, I have used the variable *Social Stratification [SCCS1751]*[21] originating in Lang (1998).[22] Both of these variables are, according to D-PLACE, based on the SCCS, and their compatibility was double-checked by inspection of whether all needed cultures in variable entries are linked to the right OWC code. There was no conflict found. Similarly to the previous variable, this one was also recoded from four values ranging from "*no differences in access to economic resources, political power, and/or status*" to "*complex stratification into more than two classes*" into

[19] When paragraphs were merged, the "Paragraph number" column was merged too in the form of a range of numbers of merged paragraphs.

[20] This represents the *Variable name* followed by *[variable ID]*, which together function as a variable identifier in D-PLACE and can be used to search among the D-PLACE variables. Link to the variable in D-PLACE: https://d-place.org/parameters/SCCS861

[21] Link to the variable in D-PLACE: https://d-place.org/parameters/SCCS1751#2/11.5/149.6

[22] Neither of the two studies – White, Burton, and Whiting (1986), Lang (1998) –is accessible anywhere online and I did not have the opportunity to read them. Moreover, the first one is cited wrong in D-PLACE, where only the first author is listed. It would be ideal to manually prepare these variables using, for example, Ethnographic Atlas as a source, but I was unable to do so due to the limited resources for a diploma thesis (especially time). For the first variable, such an approach was likely used as D. R. White is one of the forward researchers related to HRAF and the EA. The way both of the variables are structured resembles the EA, especially information in *Column 14* in Murdock (1967b) pages 155-156. However, it cannot be said with certainty what source was used for the code, especially in the case of Lang (1998).

three values "*not stratified*", "*partly stratified*" and "*highly stratified*" (for details see *appx. D* and *suppl. 10*).

For information about marriage organization, I have used a variable from the Ethnographic Atlas dataset[23] in D-PLACE.[24] As the data is originally from a different sample from the rest (SCCS, PSF) I did a review on whether this variable may be used in combination with the rest of my data and how much overlap is present. This was done by checking the OWC against the cultures' names using a table by Ember (2007, *Table 1*) which contains an overview of overlap among eHRAF, SCCS, and EA. As many cultures in the SCCS are also part of the EA, the overlap turned out to be very good, there were only 3 conflicts out of 32 cultures. Moreover, two of these conflicts are presumably due to a typo or a printing mistake as there is a clear pattern in them. "O" was replaced by "0" (zero) for cultures Mbuti (FO04) and Somali (MO04) and there are no cultures with OWC codes F004 and M004. These two conflicts were accepted as valid data. The last conflict was for culture Saramaka (SR15) which is present in the EA only in pair with OWC SR08, representing a different time or localization of the records. Therefore, records for Saramaka were replaced with missing data in the variable. All of the pairs are according to the table part of the PSF sample too. This categorical variable was also recoded from its original six-value range from "*Demes, i.e., communities revealing a marked tendency toward local endogamy but not segmented into clan-barrios*" to "C*lan-communities, each consisting essentially of a single localized exogamous kin group or clan*". The new code has three values that enable clearer result interpretation: "*Tendency towards endogamy or absence of an indication or tendency towards exogamy*", "*Agamous communities without localized clans or any marked tendency toward either local exogamy or local endogamy*" and "*Exogamous, or tendency towards exogamy*" (for details see *appx. E* and *suppl. 11*).

For the final two variables, I used a D-PLACE dataset that was derived from the ecoClimate database by Lima-Ribeiro and colleagues (2015), which is a collection of climate information from various past, present, and future predicting models created for research on biodiversity. Variables in D-PLACE use Colwell's (1974) information-theoretic index, which indicates the extent to which climate patterns are predictable due to either constancy or contingency. Specifically, I use the variables of predictability of temperature (*Temperature predictability [TemperaturePredictability]*[25]) and

[23] Ethnographic Atlas is referenced for this variable in D-PLACE as a whole. I have confirmed that the code is taken directly from EA and was not subject to further data transformation. See *Column 19* in Murdock (1967a) pages 48-49 for comparison.

[24] Link to the variable in D-PLACE: https://d-place.org/parameters/EA015#2/19.3/150.5

[25] Link to the variable in D-PLACE: https://d-place.org/parameters/TemperaturePredictability#1/30/152

predictability of precipitation (*Precipitation predictability [PrecipitationPredictability]*[26]) as indicators of environmental harshness in the regions where sampled cultures are present. The choice of these specific variables is based on the work of Martin et al. (2020) who studied the effects of living conditions on changes in alloparental care. They conducted principal component analysis and found temperature predictability and precipitation predictability the main components influencing alloparenting behavior among human societies (pp. 4-5). This indicates that climate predictability may facilitate significant changes in cultural practices. An issue with this variable was that for most societies, there are at least two recordings listed. According to the variable maps they represent different locations where the measures were taken belonging to an area where one society resides. Martin et al. (2020) always use a single of these values, but they do not mention how it was chosen in their study, and there is no mention of it in the supplements. It is not clear whether their choice was random or approximated relative to culture localization in the SCCS. I do not consider the latter approach ideal, as the localization in SCCS is not exact and cannot be paired with *WGS 84*[27] coordinates used for the D-PLACE maps. As there are no large differences between the values which range from 0 to 1, mostly up to 0.01, and few up to 0.03, with a single large exception having a -0.148 difference, I have decided to use the averaged values of the measures as I intend to better represent communities in different parts of the area where the settlements are located. After the changes (for details see *suppl. 12*), this data was added to the dataset from the previous subchapter (see *appx. B* and *suppl. 8*) finalizing the data for hypothesis testing (see *suppl. 13*).

[26] Link to the variable in D-PLACE: https://d-place.org/parameters/PrecipitationPredictability#1/30/152

[27] WGS 84 is the latest version of The World Geodetic System, which was established in 1984 as the complete standard for georeferencing by the U.S. Department of Defense. See Boucher & Altamimi (2001) for details.

3 Analysis

This section of the thesis will first analyze the material collected from eHRAF. This is done to gain more insight into the quality of the procedure and the reliability of the material itself. Afterward, it will contain the testing of the hypotheses and will be concluded with a reflection of the geographical coverage of individual models in comparison to the sample that was described in the data collection.

3.1 Research material analysis

3.1.1 Inter-rater reliability

To get a better insight into the quality of data for my analysis I have performed an Inter-rater reliability (IRR) analysis. This analysis shows the degree to which coders consistently assign values to subjects of study (Hallgren, 2012). In this specific case, it shows how much agreement was between coders while assigning categorical values to individual paragraphs. Analysis was performed using a dataset created by joining the coding sheets of all three coders and labeling every variable by the coder's number.

I have chosen Cohen's kappa (κ) (Cohen, 1960) as the most appropriate IRR index for my data. Cohen's kappa is designed for nominal data, however, its mathematical foundation is only fit for two simultaneous coders. As my research design is fully-crossed, meaning every coder coded all of the observations, it was appropriate to use an approach for three or more coders offered by Light (1971). He suggests that an arithmetic mean of Cohen's kappa calculated for all coder pairs is taken as an overall index of agreement. Each Cohen's kappa was calculated as unweighted, as the variables have no ordinal scale (Cohen, 1968).

In the initial dataset, I have replaced missing values with the number 99 so the *irr* package (Gamer et al., 2019) in R can process them (see *suppl. 14*).[28] Hallgren (2012) suggests the removal of missing values before the calculation of the kappa, but I also needed to assess how often coders agreed that the information they are looking for is lacking in the coded text. I have, therefore, kept the missing values and calculated the kappa twice, including (see *suppl. 15*) and excluding (see *suppl. 16*) missing values.

Results were similar for both calculations. In either case, with (w/) or without (w/o) missing values (N/A), mean index values fell in the same category of agreement. For the variable of dance presence, the mean of all coder pairs was κ≈.429 (w/ N/A) and κ≈.464 (w/o N/A) which according to Landis and Koch (1977) is a moderate agreement between coders (0.41-0.60). Results for the means were slightly higher for the

[28] Package *dplyr* (Wickham et al., 2022) was used to transform data for this analysis.

variable of dancers' gender κ≈.492 (w/ N/A) and κ≈.540 (w/o N/A). There was, however, a much smaller agreement for variables observers' gender (κ≈.328 w/ N/A; κ≈.243 w/o N/A) and the presence of families at the dance[29] (κ≈.361 w/ N/A; κ≈.326 w/o N/A), which should be understood as fair agreement (0.21-0.40) only according to Landis and Koch (1977). Kappa for both, means and all coder pairs were always below 0.67, which tells us that the results of the following analysis should be discounted according to Krippendorff (1980). Moreover, none of the coder pairs' kappa's p-values were over 0.05, which means that we can reject the null hypothesis that found agreement between coders was due to chance.

I have mentioned above that information about genders in the dance may have been stated in the coded texts implicitly. I have included this in the coding manual and had the coders assign different codes if the information about the presence of both genders in the dance was stated implicitly or explicitly. I have calculated additional Cohen's kappa for data where the distinction between implicit and explicit mentions in gender variables was removed (see *suppl. 17*, *18*, and *19*) and there was a noticeable difference. Using the interpretation by Landis and Koch (1977), the kappa means for the variable dancer's gender showed a moderate agreement (κ≈.603) including N/As and substantial agreement (0.61-0.80) after their exclusion (κ≈.738). There was also moderate agreement (0.41-60) for the variable of observer's gender when N/As were included (κ≈.506) and substantial agreement after their exclusion (κ≈.802). Same as before, none of the p-values were over 0.05, so it can be assumed that calculated agreements are not due to chance. Information about who participates or observes the dance was expected to be straightforward in coding but it turned out that more inference was required than originally anticipated. This might also indicate that some reservations should be taken regarding the informative value of the material. As the agreement was much higher when the implicit and explicit codes were merged, the same interpretative approach will be used in the hypothesis testing. However, I must note that the results will depend on implicitly stated information in the original texts.

3.1.2 Inter-rater reliability: Interpretation

There are several possible reasons why the agreement between coders is generally not too high. It might be that the data required much more inference than previously expected. This could mean that there might have been a mistake made in the preparation of the coding manual. The instructions for coders may not have been clear enough, they might have lacked sufficient examples or the manual failed to notice some important factor that could be of influence in the decision-making.[30]

[29] Variable *location*, refers to the location of families when the dance takes place (together or apart).

[30] It would be possible to conduct additional analysis where I would ask coders why they have chosen specific codes. Its results, however, would be questionable as some time has passed since the

Results for the variable of dance presence are surprising as there was only moderate agreement between coders found regardless of whether missing values were included or not. In the light of how the eHRAF search was conducted and how were the codes defined in the manual,[31] the agreement was expected to be substantial or higher. Here it appears that there were several cases where both dance and nuptials are mentioned in a paragraph, but with no clear relation to each other. And it was necessary to infer whether the dance mentioned in the paragraph was or was not part of a nuptial ritual. It is not easy to predict how much this may have influenced other results. It should become a subject of further methodological research, whether it is acceptable to use such data or how to best interpret such results. Results indicate that the tagging of paragraphs by cultural traits mentioned in the text (OCM) may not be as helpful as anticipated. That may be caused by paragraphs not being tagged with enough care, or because such categorization may be too broad to use the way it was used in my research and more search specifications are necessary.

Also, results from the original IRR dataset show that in the case of kappa's calculated including missing values, the agreement between the first and third coder was consistently higher than agreements with the second coder. This may be partially explained by the shared background between first and third coders. They were both students of the Study of Religions, so their interpretations may be closer to each other. The second coder was from a different program. However, their background was also focused on cultural practices and human behavior, so they were considered appropriately equipped for the work of a coder. It is therefore unlikely that the difference did rise because of the lack of experience or understanding, necessary for proper interpretation of the material. After the removal of missing values, this pattern could no longer be observed in the results, the same as it was not present after merging implicit and explicit mentions for the gender variables. I may only speculate whether this indicates that the coder's background influence was not too strong.

3.1.3 Further material analysis

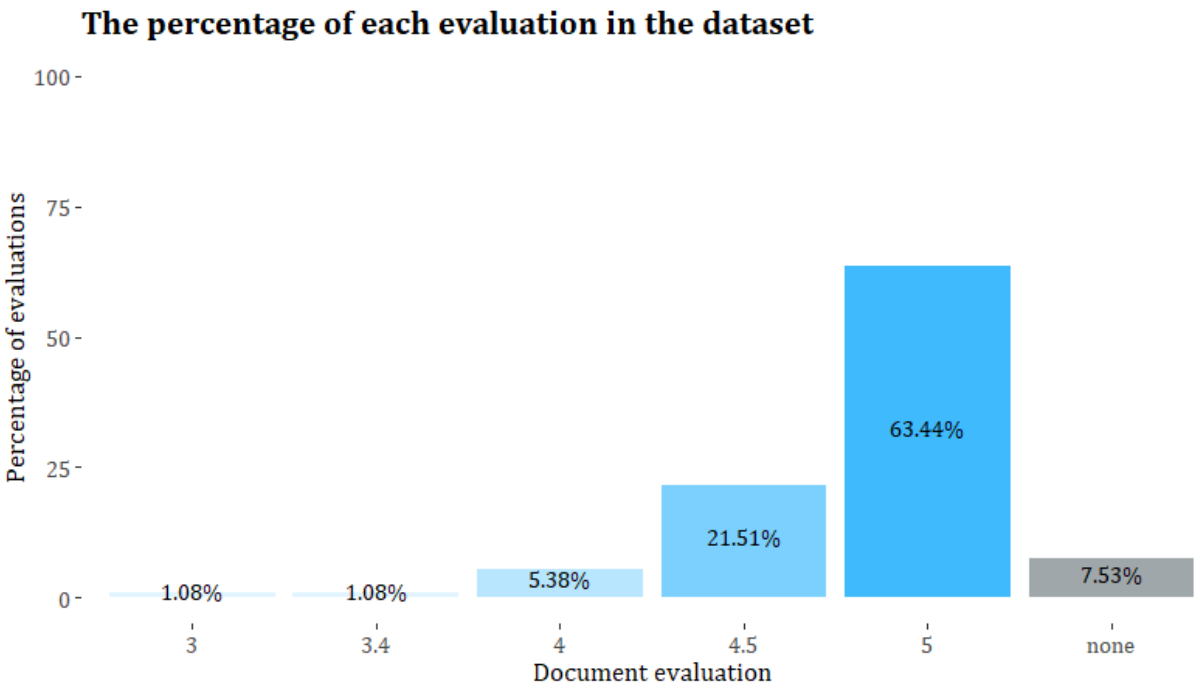
The relatively low IRR results may also indicate that the quality of the material may not be very high, or insufficiently high for this kind of research approach, because descriptions are not informative enough or confusing. As I have noted before, eHRAF offers the user the material collector's background in pair with an evaluation score of the document in question. We currently do not have a way of measuring the degree to which the collector's background could have influenced the material when it was collected, which is also the reason why the meaning of dance was not coded in any form.

coding was done and there is no telling whether coders would remember any of their choices and the original reasoning behind them.

[31] 0: Dance appears in a nuptial ritual; 1: Dance does not appear in a nuptial ritual; N/A: It cannot be said whether dance appears in a nuptial ritual; (for details see *suppl. 2*)¨.

Such a variable would be strongly prone to bias from the side of the material collector as it may require a lot of inference on their end.[32] Nonetheless, it seems unlikely for the collector's background to have influenced the IRR results.

Figure 1



There is the possibility to look at the evaluation score to get a better picture of the material quality. In my sample the majority of documents are rated as having high quality.[33] More than 63% of documents are rated as 5/5 and more than 21% are rated as 4.5/5 (Fig. 1). If there is a correlation between the document evaluation and the amount of agreement between the coders, it would be possible to claim that the source material is confusing or not sufficiently informative. To show whether there is any correlation between the evaluations and agreement between the coders I have created a dataset where the degree of agreement between the coders is coded on an ordinal scale

[32] More than 60% of material collectors in my sample have an ethnologist background and many others have backgrounds that imply the capability of collecting information in a relatively unbiased way (see table in *suppl. 24*). However, I find it problematic to use this information for document evaluation as it paints the picture about the data collector very poorly.

[33] See *supplements 1*, *23*, or *24*.

(0-2; no agreement, partial agreement, full agreement) and compared it with the evaluations which were also coded on an ordinal scale.[34]

I have decided to use Kendall's Tau (τ) (Kendall, 1948) as a statistical measure of correlation as it is best fitting for the calculation of correlation coefficients on ordinal data (Schaeffer & Levitt, 1956). For its calculation, I have used the *stats* package in R (see *suppl. 22*)[35]. Tau estimates were very low for every variable (dance presence, τ≈-.029; dancer's gender, τ≈.009; observer's gender; τ≈-.029, location, τ≈-.065), which shows that there should not be any correlation between tested variables. P-values were very high (≈.649, ≈.887, .652, ≈.309 respectively), so this result has no statistical significance. Either way, the null hypothesis that true Tau is equal to 0 is kept and there does not seem to be any correlation between the document evaluations and the degree of agreement among the coders. There should be no dependency between the variables.

From the results, it appears that the mistake was made on the side of the research design as noted above. The results may, however, be distorted due to the small number of cases among the lower ratings, a larger sample would have provided a more reliable conclusion. Another problem is that the criteria of evaluation for individual documents are unclear. I did not find any public statement from eHRAF on the criteria under which the evaluations are made. Therefore, some bias may be expected until this doubt is cleared from the side of eHRAF management. Lastly, in my sample, there were also 7.53% of materials (Fig. 1) that have no evaluation whatsoever and were omitted from the test as missing data, which may have also clouded the results.

3.2 Hypotheses testing

Since there are several groups of recordings by individual cultures that bear a certain degree of variability that needs to be accounted for in the statistical tests, I have chosen linear mixed effects models as the best approach to the development of my models. In contrast to simple or multiple linear regression, these models are capable of reflecting the variability in individual groups of observations in a variable (random effects). As I work with categorical (often binary) variables, I have specifically chosen generalized linear mixed effect models (GLMM) as the best-fitting mixed effect model variant (Lee & Grimm, 2018).

[34] Evaluations in my sample had values of: 3, 3.4, 4, 4.5, and 5. As the two smallest ratings were of one case only, they were coded together with the five 4-point ratings as "average". For details see *appx. F*, and *suppl. 20* and *21*.

[35] Package *stats* is part of the base R and is, therefore, not cited separately.

All models are structured as the following example (exact formulations for individual hypotheses can be found in *supplement 25*):

```
model <- glmer(response_variable ~ fixed_effect1
                                 + fixed_effect2
                                 + (1|grouping_variable),
               data = dataset,
               family = binomial(link = ["logit" or "probit"]),
               na.action = na.omit)
```

`model` is the object in R to which the specified model is assigned. `glmer` is the function from the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017) in R that is used for the calculation of model results when there are categorical variables present. Its arguments can be read as follows: response (dependent) variable is affected (~) by the fixed effect(s)[36] and random effects (grouping variable; this is variable *culture* for all models); data from the data frame *dataset* are used; we inform the function that the response variable should be treated as binary (both of my response variables are binary), while the link (either "logit" or "probit") references the distribution of predictor variables;[37] and rows with missing values (only from the variables specified in the model) should be omitted.

3.2.1 Hypotheses 1 and 2

This subchapter looks at the first two hypotheses together as they are theoretically more related compared to the third and fourth hypotheses. For these two hypotheses, records (rows) where women do not observe men dancing were removed as women are the courtship signal receivers. If men are not observed by women while dancing, it can be immediately assumed that such dance serves other purposes than courtship, as there is no one to court.[38] There are three models discussed for these hypotheses (Tab. 1). *Model 1* contains polygyny as a fixed effect, which is relevant mainly for the first hypothesis, *model 2* contains social stratification as a single fixed effect, which is relevant for the second hypothesis and *model 3* includes both of these variables as fixed effects simultaneously.

[36] There may be one or multiple fixed effects at the same time.

[37] Argument *link = "logit"* is used for logistic regression, in other words, when both the response variable and predictors are categorical (hypotheses 1, 2, 3), while *link = "probit"* is used when the predictor variables are normally distributed (hypothesis 4). Shapiro-Wilk test was performed to confirm the normal distribution of the fourth hypothesis fixed effects (see *suppl. 25*).

[38] This study does not account for the possibility of intrasexual ("homosexual") courtship as the theoretical basis of this work cannot be used to explain it. Moreover, the sources would most likely not provide sufficient or reliable evidence for such tests due to their historical background.

Table 1

Effects of polygyny (FE) and social stratification (FE) on whether men dance (culture = RE, grouping var.)

	Model 1: men dance			*Model 2: men dance*			*Model 3: men dance*		
Predictors	*Estimate*	*Conf. Int.*	*p*	*Estimate*	*Conf. Int.*	*p*	*Estimate*	*Conf. Int.*	*p*
(Intercept)	74.33 (121.18)	3.04 – 1815.43	**0.008**	139.55 (236.60)	5.03 – 3872.07	**0.004**	127.30 (227.57)	3.83 – 4231.44	**0.007**
polygyny	0.45 (0.36)	0.10 – 2.17	0.322				1.11 (0.83)	0.26 – 4.77	0.891
social stratification				0.13 (0.13)	0.02 – 0.85	**0.033**	0.12 (0.13)	0.02 – 0.98	**0.048**
Random Effects									
σ^2	3.29			3.29			3.29		
τ_{00}	2.26_{culture}			0.77_{culture}			0.73_{culture}		
ICC	0.41			0.19			0.18		
N	24_{culture}			17_{culture}			17_{culture}		
Observations	140			115			115		
Marginal R^2 / Conditional R^2	0.055 / 0.440			0.201 / 0.352			0.204 / 0.349		

The confidence intervals (CI) indicate that we should be 95% sure that the true population estimates fall into the calculated confidence range and in model 1 they are in accord with the first hypothesis as the range is in positive numbers for both intercept and the predictor. As the result for the intercept is statistically significant (p = 0.008), we can trust the CI even if the estimate calculation indicates the possibility of an opposite effect. Nonetheless, the effect of the predictor (polygyny) on the response variable (men dance) is statistically not significant. This means that we cannot reject the null hypothesis that the custom of men dancing in nuptial rituals is present in the same amount regardless of whether the society is polygynous or not. This result indicates that dance might function for other purposes than courtship.

Model 1 contains 140 observations of 24 cultures (N) out of the original 31 before the removal of missing data. The conditional r-squared (R^2), shows that the proportion of variance in the intercept explained by the fixed and random effects is 44%, and the intraclass correlation coefficient (ICC) suggests that a large part of the variation in the response variable is caused by the grouping variable (*culture*). The marginal r-squared (R^2) tells us that the proportion of variance explained by the fixed effects is low in this model (5.5%). This means that polygyny has a low impact on the model results with no regard to the statistical significance of its effect on the intercept. The value of sigma-squared (σ^2) suggests that there is a considerable variation in the response variable that is not explained by the fixed effects and is instead attributable to random error. This would indicate that the model might perform better and provide more reliable results if we controlled for other possible influences that were not accounted for in the

model (and in the study), such as the time foci. This statistic is the same for all models for the first three hypotheses (Tab. 1, Tab. 2, $\sigma^2 = 3.29$) and it can be considered noteworthy also for the fourth hypothesis models (Tab. 3, Tab. 4, $\sigma^2 = 1$). Tau double-zero (τ_{00}) shows the estimated variance of the random effect for the intercept across the different levels of the grouping variable (culture). It seems that the variance on the cultural level is quite high, especially in comparison with τ_{00} for *models 2* and *3*. This means that there are noticeable differences in how variance in individual cultures influences the statistic for whether men dance in a nuptial ritual. It tells us that paying closer attention to unspecified subgroups of variable *culture* (e.g., some local subregions, village clusters) represented in the model may help us better understand the response variable as their effects may vary. It needs to be kept in mind that cultures whose data are missing could have caused a significant shift in the results in both directions.

As the effect of polygyny was found not significant, I have proceeded to test the effects of social stratification as an individual fixed effect. In the second model (Tab. 1, *Model 2*) we can see a statistically significant effect of social stratification (*p = 0.033*) on whether men dance in nuptial rituals. The positive confidence intervals indicate that with the rise of social stratification, the appearance of men dancing in the ritual should grow. This might be contrary to the formulation of the second hypothesis, where the rise of social stratification was expected to reduce the presence of men's dance in nuptial rituals in polygynous societies. But as polygyny is not included in this model, it cannot be concluded that the model suggests that dance functions for a different purpose than courtship. The conditional R^2 shows that the proportion of variance explained by both the fixed and the random effects is roughly 35% in this model, and the marginal R^2 shows that the proportion of variance explained by the fixed effects is about 20%, which is substantial. ICC shows that the variation in the response variable caused by the random effect is also present in a noteworthy amount. This indicates that social stratification plays a noticeable role in whether men dance in nuptial rituals or not across different cultures

Model 3 includes both variables, polygyny, and social stratification, as fixed effects. Since there is more than one fixed effect in the model, the variance inflation factor (VIF) was calculated to make sure there is no multicollinearity between the predictors (Mansfield & Helms, 1982). A high correlation between predictor variables may cause the regression model to fail to converge or produce unexpected results. VIF for each of the fixed effects in the third model is 1.2518,[39] which means there is some collinearity present (VIF > 1) but as the value is below 1.5 it is generally accepted as non-problematic. We further observe a scenario very similar to the previous two models. Regarding both, the confidence intervals and statistical significance of fixed effects. This model contains 115 observations of 17 cultures (N) out of the original 31, which raises the

[39] *car* package (Fox & Weisberg, 2019) in R was used to calculate VIF (see *suppl. 25*).

question of how representative the model is on a worldwide scale. This issue was present with both, *model 1* where 24 cultures were represented, and *model 2* which has a very similar representative extent as this one. τ_{00} is 0.77 for *model 2* and 0.73 for *model 3*. It shows that the variance across different levels of the grouping variable is considerably smaller compared to *model 1*. This is likely caused by fewer cultures being represented due to the omission of missing data, as in the third model missing data for variables from both previous models were omitted. Values of ICC, conditional R^2, and marginal R^2 are only slightly different in comparison to *model 2*. Generally, the differences between *model 2* and *model 3* are only minor, reflecting the inclusion of the polygyny variable. Following the results of *model 1* and *model 3*,[40] it can be expected for dance to function for a different purpose than courtship in nuptial rituals.

3.2.2 Hypothesis 3

For the third hypothesis, only a single model was built (Tab. 2, *Model 4*), showing the effect of marriage organization (or marital structure) in a society on whether men dance in a nuptial ritual.

Table 2

Effects of marriage organization (FE) on whether men dance (culture = RE, grouping var.)

	Model 4: men dance		
Predictors	*Estimate*	*Conf. Int.*	*p*
(Intercept)	27.99 (37.06)	2.09 – 375.07	**0.012**
marriage organization	1.12 (0.86)	0.25 – 5.05	0.883
Random Effects			
σ^2	3.29		
$\tau_{00\ culture}$	3.36		
ICC	0.51		
$N_{\ culture}$	24		
Observations	145		
Marginal R^2 / Conditional R^2	0.001 / 0.506		

Even though there is a large standard error for the intercept estimate, the confidence intervals reflect the hypothesis expectation. They show that with the rise of the predictor, the intercept also rises, in other words with the rising presence of exogamy, men's dance should appear more in the nuptial ritual. Similarly to most of the models, there is again the question of worldwide representability, which I am going to address in more detail later. The result for the intercept is statistically significant so we can trust its calculated estimate and CI, however, the result for the marriage organization predictor is statistically not significant. τ_{00} shows large variance across grouping variable levels, showing strong differences in how the presence of men's dance is affected by the lower cultural level. The marginal R^2 shows that the variance in the intercept explained by the fixed

[40] *Model 2* helped clarify the impact of social stratification in the third model as described above.

effect is only 1%. This is very low and shows that from a worldwide perspective, the marital structure has little explanatory power for whether men dance in nuptial rituals. According to ICC and conditional R^2, most of the variance is explained by the random effect. We can safely conclude that marriage organization has little effect on the intercept, the same as it has little effect in this model generally. Therefore, we keep the null hypothesis that in nuptial rituals the occurrence of men's dance is not influenced by the degree of exogamy.

3.2.3 Hypothesis 4

One theoretical challenge in answering this hypothesis has surfaced in the data. There were a few observations where women danced while men were not observing them in the nuptial ritual context. Similarly to the first two hypotheses, it could be assumed that if men do not observe women during the dance it occurs for a different reason than courtship, as men are the signal receivers of the courtship signal. However, this might not be necessarily true in this case. The first two hypotheses were based on direct reproductive signaling between men and women. But in the fourth hypothesis mate-guarding is involved and behaviors connected to it may be strongly socialized.

A good example of this is the wearing of wedding rings. They are typical cross-culturally present marital markers (Burch, 2019; Monger, 2004) and their wearing can be considered a mate-guarding method (Pecka, 2020), especially if it is required by a partner. However, individuals usually do not put their wedding rings off when they do not interact with the opposite gender, but wear them constantly. Similarly, behaviors that are motivated by mate-guarding strategies may persist even in situations where they are not necessarily required as they are mandatory elsewhere. There might be different institutional pressures (not just mate-guarding) that limit women's dance in harsher environments, but current research has not identified them yet. For this reason, models for the fourth hypothesis were calculated twice, once with and once without the omission of cases where men are not observing the dance (Tab. 3, *models 5-7* with data omission [w/ d. o.], and Tab. 4, *models 8-10* without data omission [w/o d. o.]).

As discussed before, climate predictability is expected to bear a significant influence on change in cultural practices. Two models (w/ and w/o d. o.) were built as outlined in the introduction to this chapter, including an interaction between the predictability variables[41] to model climate predictability. As there are two fixed effects in each model, VIF was calculated to indicate possible multicollinearity. VIF values were very high (precipitation predictability, ≈142 w/ d. o., ≈100 w/o d. o.; temperature predictability, ≈45 w/ d. o., ≈44 w/o d. o.; interaction, ≈301 w/ d. o., ≈224 w/o d. o.) indicating

[41] In the formula asterisk (*) was used between the fixed effects instead of plus (+), for details see *suppl. 25*.

strong multicollinearity. The estimates were critically affected by the issue (extreme estimate inflation), and the model was deemed unstable and inaccurate.[42]

Table 3

Effects of precipitation and temperature predictability (FEs) on whether women dance (culture = RE, grouping var.; with data omission)

	Model 5: women dance (w/ d. o.)			*Model 6: women dance (w/ d. o.)*			*Model 7: women dance (w/ d. o.)*		
Predictors	*Estimate*	*Conf. Int.*	*p*	*Estimate*	*Conf. Int.*	*p*	*Estimate*	*Conf. Int.*	*p*
(Intercept)	40.60 (85.63)	0.65 – 2533.92	0.079	5.64 (7.59)	0.40 – 78.84	0.199	39.48 (79.66)	0.76 – 2059.59	0.068
precipitation predictability	2.52 (6.52)	0.02 – 399.15	0.720	0.57 (1.23)	0.01 – 40.40	0.794			
temperature predictability	0.02 (0.07)	0.00 – 23.08	0.271				0.04 (0.11)	0.00 – 7.90	0.233
Random Effects									
σ^2	1.00			1.00			1.00		
τ_{00}	0.30_{culture}			0.41_{culture}			0.28_{culture}		
ICC	0.23			0.29			0.22		
N	26_{culture}			26_{culture}			26_{culture}		
Observations	143			143			143		
Marginal R^2 / Conditional R^2	0.060 / 0.276			0.003 / 0.290			0.052 / 0.261		

To improve the fit, I have removed the interaction for both models. VIF for models without interaction between the fixed effects was slightly over (*Model 5*, w/ d. o., ≈1.68) and slightly below (*Model 8*, w/o d. o., ≈1.47) the accepted value of 1.5. I have calculated Pearson's product-moment correlation (Puth, Neuhäuser, & Ruxton, 2014), which was substantial and statistically significant for both *models 5* and *8* respectively ($p < 0.001$, $p < 0.001$; Est.: ≈0.55, Est.: ≈0.49; CI: ≈0.45-1, CI: ≈0.40-1), confirming the presence of multicollinearity for both models. Nonetheless, models without interaction did not have estimation problems and are included in the study, as suggested by Tay (2017).

As there was multicollinearity present in the models with both climate predictability predictors (*Models 5* and *8*) I have proceeded to build models with single fixed effects. These are *models 6* and *9* (w/ and w/o d. o., respectively) where precipitation predictability is the fixed effect, and *models 7* and *10* (w/ and w/o d. o., respectively), where temperature predictability is the single fixed effect predictor.

[42] Due to this issue this model was rejected and is not reported in the tables below along with the rest of the models for this hypothesis. Concrete results can be checked in respective summaries and sjPlot in the *suppl. 25*.

Confidence intervals for all six models indicate that with the rise of predictability of either precipitation or temperature the occurrence of women dancing in the nuptial rituals should grow. This is following the fourth hypothesis which expects a decrease in women's dance in harsher environments. ICC and conditional R^2 show that a noticeable portion of the variance is explained by the random effect for each model. From the single fixed effect models' results, it is clear that temperature predictability is responsible for the majority of variance caused by fixed effects in the models. What is unexpected is that the marginal R^2 for models with a single fixed effect of precipitation predictability is very small, showing that this predictor is responsible for only 0.3% of the variance in *model 6* and 1% of the variance in *model 9*. This is substantially lower than that of the predictability of temperature, showing that models that are best capable of answering the hypothesis would probably be those using temperature predictability as a single fixed effect. Nonetheless, none of the results is statistically significant, so we must keep the null hypothesis that the harshness of an environment does not affect the frequency of women dancing in nuptial rituals.

Table 4

Effects of precipitation and temperature predictability (FEs) on whether women dance (culture = RE, grouping var.; without data omission)

	Model 8: women dance (w/o d. o.)			*Model 9: women dance (w/o d. o.)*			*Model 10: women dance (w/o d. o.)*		
Predictors	*Estimate*	*Conf. Int.*	*p*	*Estimate*	*Conf. Int.*	*p*	*Estimate*	*Conf. Int.*	*p*
(Intercept)	20.98 (35.13)	0.79 – 558.95	0.069	6.11 (6.72)	0.71 – 52.87	0.100	20.86 (35.11)	0.77 – 564.53	0.071
precipitation predictability	0.89 (1.82)	0.02 – 48.32	0.955	0.36 (0.65)	0.01 – 11.89	0.570			
temperature predictability	0.08 (0.23)	0.00 – 18.42	0.368				0.08 (0.18)	0.00 – 6.81	0.263
Random Effects									
σ^2	1.00			1.00			1.00		
τ_{00}	$0.15_{culture}$			$0.20_{culture}$			$0.15_{culture}$		
ICC	0.13			0.16			0.13		
N	$26_{culture}$			$26_{culture}$			$26_{culture}$		
Observations	149			149			149		
Marginal R^2 / Conditional R^2	0.036 / 0.163			0.010 / 0.173			0.036 / 0.165		

Multiple estimates and confidence interval ranges in the models are quite high, while it is not the case for others. *Models 3*, *6*, and *9*, do not seem to be affected by the issue. It is possible that the inflated estimates were partially caused by the time foci issue since the random error is known to grow when the time foci are not properly controlled for (Divale, 1975; Ember & Ember, 2001; Ember et al., 1991). It is possible

that in some models the observations with non-compatible time foci were removed, either along with missing data or among observations omitted based on the theory. This explanation sounds plausible as the issue is not present in every model. However, there must be another unknown problem present as the intervals are unusually wide for some of the fourth hypothesis models, and for some, they seem standard, even if the models share the same population.

3.2.4 Representability of results

Initially, this research was designed as a worldwide cross-cultural study. However, due to the amount of missing data, the results for individual models were calculated from populations that do not represent all of the cultures that should be present for the results to be generalizable worldwide. This subchapter will review which regions and subregions were represented in the individual models and to what extent they were present.

In the complete sample, there are 34 cultures distributed among 8 regions that are divided into 17 subregions (Tab. 5). Cultures of Africa cover about 32.4%,[43] cultures of Asia 23.5%, cultures of America 29.3% (South America 11.7%, Middle America and the Caribbean 2.9%, North America 14.7%), Europe covers 2.9%, the Middle East region also covers 2.9%, and Oceania covers 8.8% of the sample. The following bar chart (Fig. 2) describes the number of cultures in each subregion in the sample and how many percent of total cultures in the sample are covered by the cultures in each subregion. The rest of the bar charts in this subchapter (Fig. 3-8) will display this information for models used in the analysis. Since the percentage distribution of total cultures in the subregions is

Table 5

Locations of represented cultures

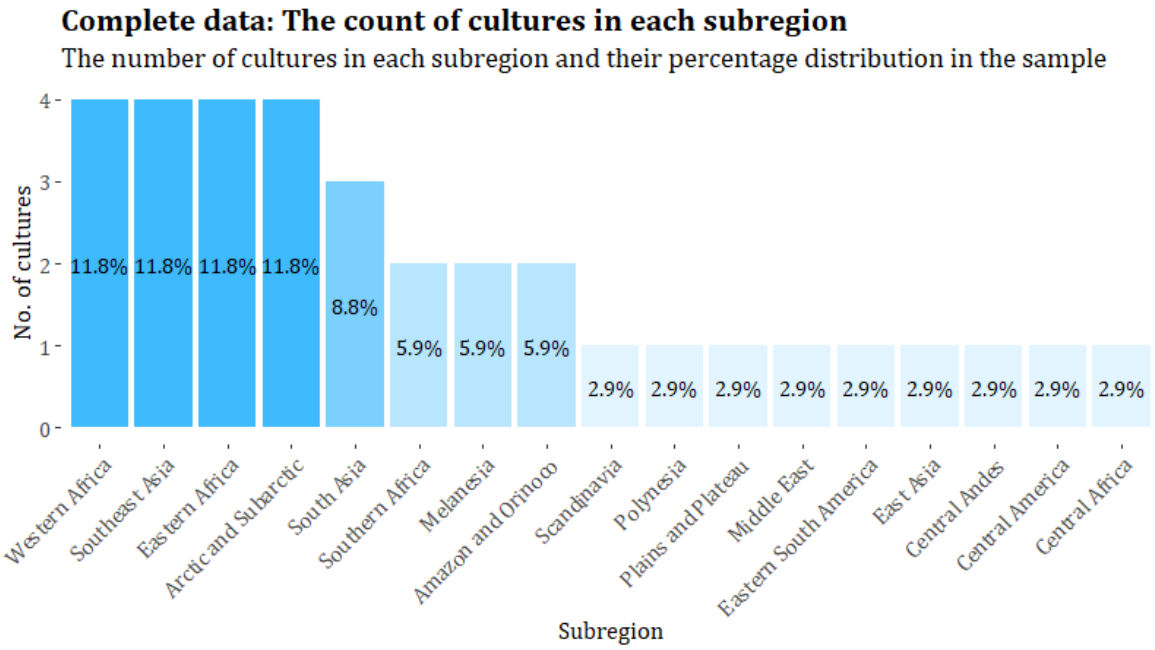
Regions and subregions in the original dataset

Region	Subregion
Africa	Central Africa
Africa	Eastern Africa
Africa	Southern Africa
Africa	Western Africa
Asia	East Asia
Asia	South Asia
Asia	Southeast Asia
Europe	Scandinavia
Middle America and the Caribbean	Central America
Middle East	Middle East
North America	Arctic and Subarctic
North America	Plains and Plateau
Oceania	Melanesia
Oceania	Polynesia
South America	Amazon and Orinoco
South America	Central Andes
South America	Eastern South America

[43] All of the percentages in this subchapter are approximate, as the calculations were rounded to one decimal (see *suppl. 27*).

displayed, it needs to be kept in mind that even if the distribution of cultures in subregions is the same or similar between models or in comparison to the complete data, the combination and number of represented cultures might differ from the original sample. The percentage is displayed to improve the readability of the balance between regions in the model populations compared to the balance in the original sample in the resulting graphs.

Figure 2



In *models 1* and *4* (Fig. 3 and 4) 24 out of 34 originally sampled cultures are represented. These two models have the same count and percentage distribution in the subregions. The African cultures cover 45.9% of the model samples, with 10 out of 11 cultures present in the original sample. Because 10 cultures are missing from the model data, the increased percentage indicates that some other (sub)regions are likely to be underrepresented. Asian cultures cover 25% of the model samples, which is a similar amount to the distribution in the full dataset, both South and Southeast Asia subregions are missing one culture, weakening the representation of Asia. Cultures of America represent only 16.8% of the model samples. This is slightly over half of the original sample. The subregion of Amazon and Orinoco (South America) is missing 1 out of 2 cultures. At the same time, the cultures from North American subregions are entirely missing in the sample causing the majority of the percentage distribution drop. The Middle East (Middle East) and Europe (Scandinavia) are represented by the single culture that represents them in the original sample, here accounting for 4.2% instead of 2.9% each. Oceania is represented with 4.2% (1 culture in Polynesia), while the Melanesia subregion is missing in the model samples. Even though some subregions are

weaker in the number of cultures present in models *1* and *4*, the distribution of cultures is not too far from the original sample. With a reserve, it can be considered that the results are generalizable worldwide except for North America and Melanesia which are not represented at all.

Figure 3

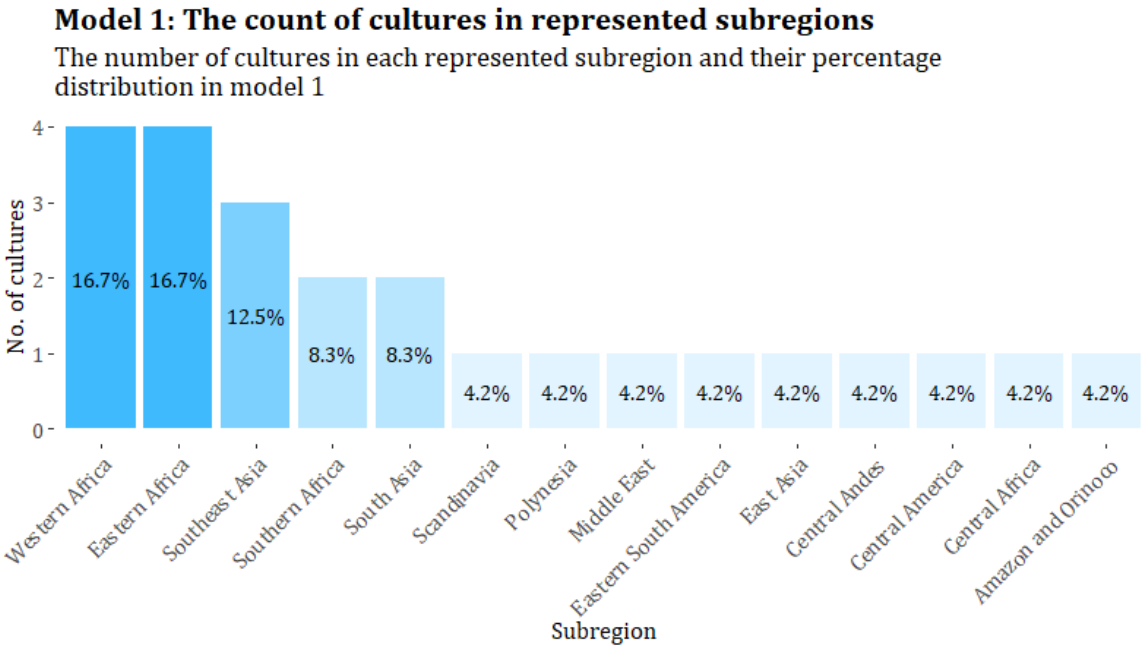
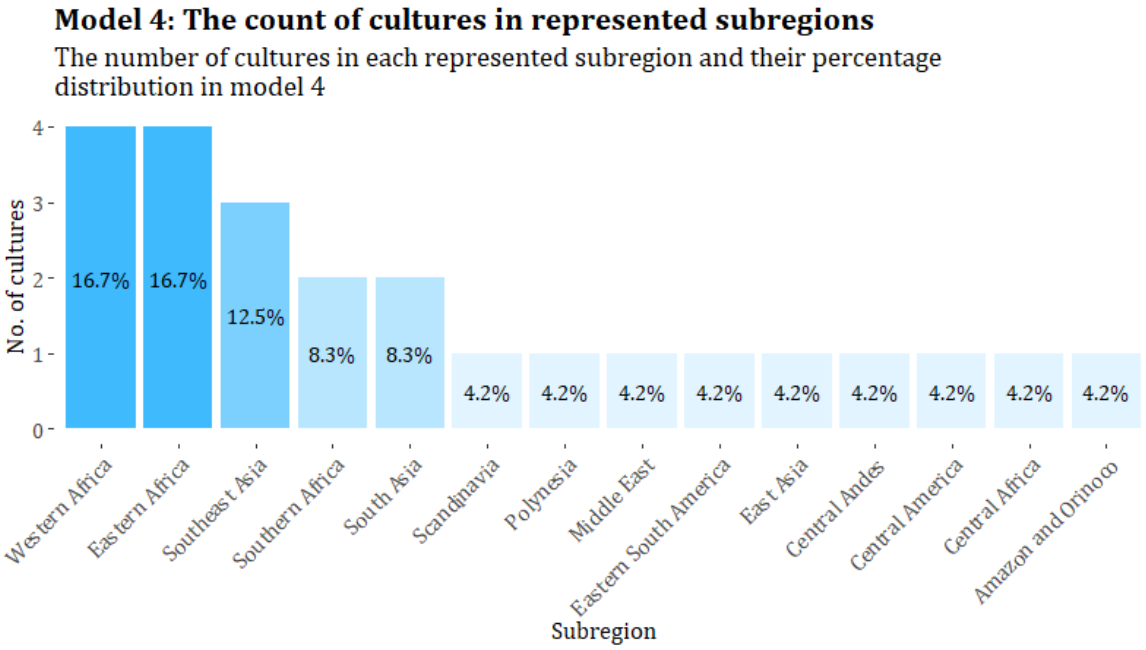


Figure 4



In *models 2* and *3* (Fig. 5 and 6, respectively) the count and percentage of cultures in subregions are the same. This is not surprising considering how the models were built but highlights that the data for social stratification had substantially more missing values than the data for polygyny. This is important to mention as the only statistically significant fixed effect found in the analysis was that of social stratification.

Figure 5

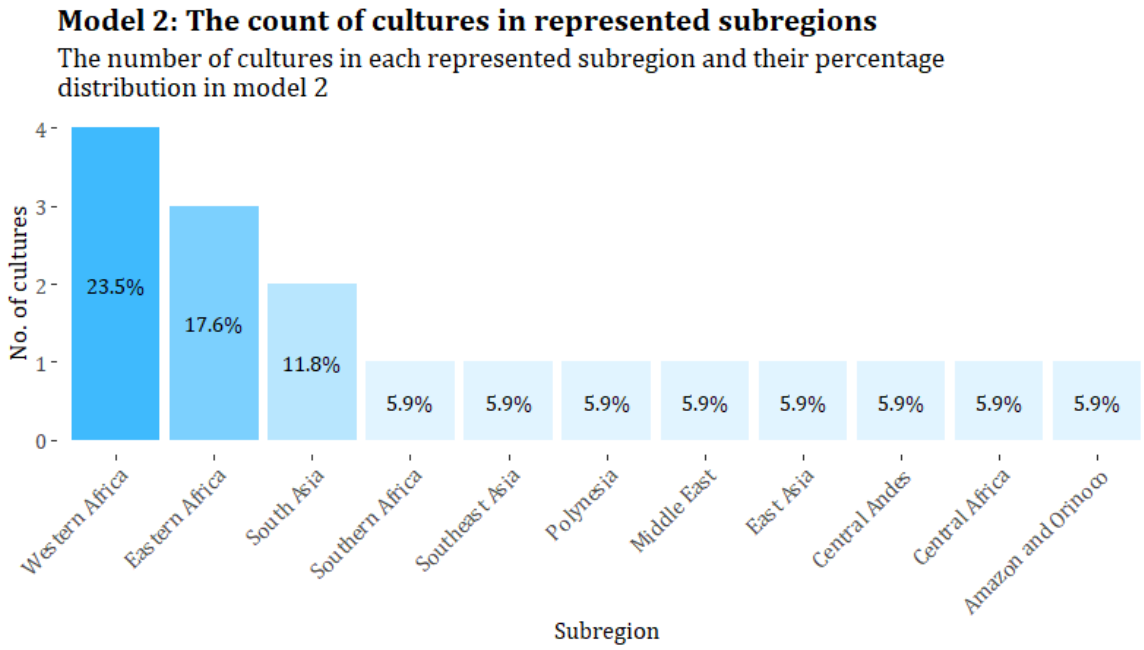
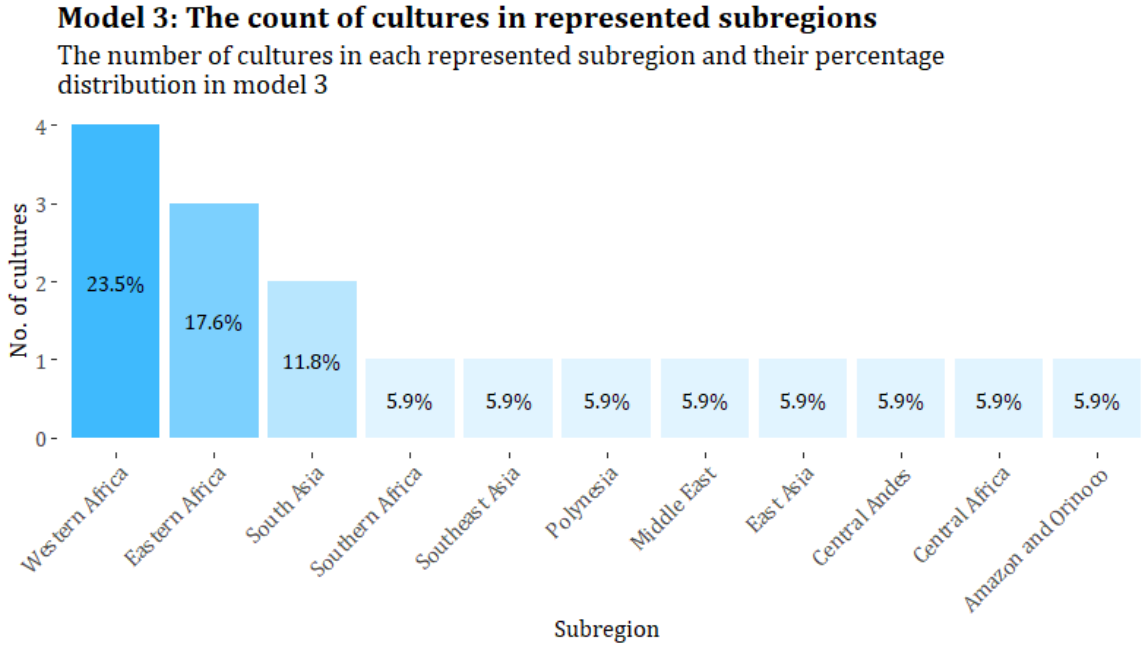
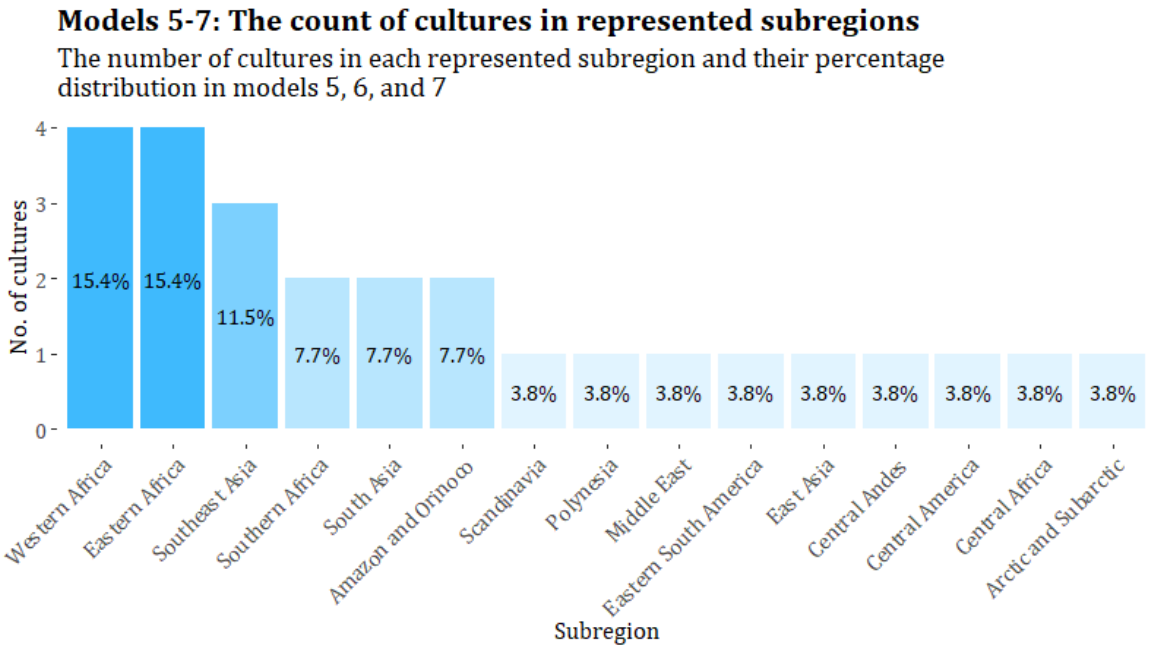


Figure 6



African regions in models *2* and *3* cover 52.9% of the samples, showing that Africa is well represented in the models. The percentage might seem large but it is caused by fewer cultures being represented in the model overall. The percentage accounts for 9 out of 11 cultures in the full sample. Asian regions cover 23.6% of the sample populations which is almost the same percentage as in the complete sample, but the culture count is only 4 out of 8, with weaker representation of Southeast and South Asia. America is heavily underrepresented with only 11.8% coverage from two out of three South American cultures. Data for Middle America and the Caribbean, and North America are completely missing in these models. And so are missing any data for Europe. Data for Melanesia are missing too, leaving Oceania underrepresented in this model with only one Polynesian culture. It can be concluded that the results for the second hypothesis might be generalizable for Africa. However, representation of other regions is often weak or none, therefore the results are not generalizable worldwide.

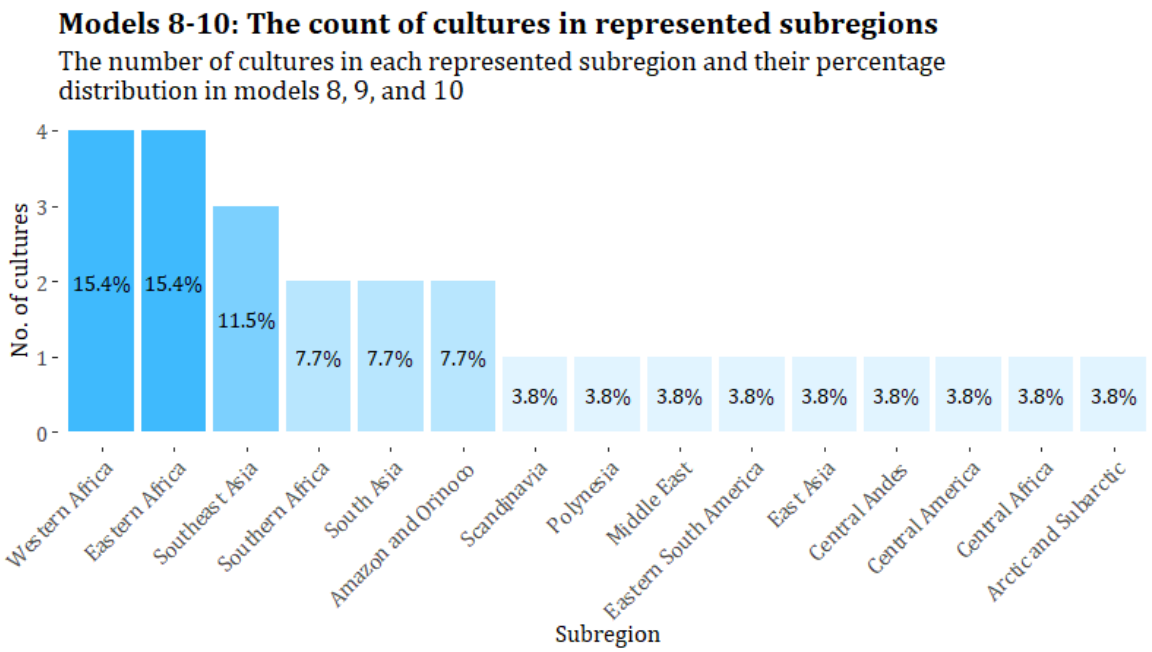
Figure 7



Models 5-10 (Fig. 7, 8) have the same count and percentage distribution of cultures in the subregions. Africa is represented by 42.3% of cultures in the models, which is all of the cultures from the original sample. The percentage is again increased indicating missing data elsewhere. Region of Asia is represented by 23% of cultures in the models, while missing two cultures, one from South and one from Southeast Asia. America is represented by 22.9% of cultures with full representation in South America and Middle America and the Caribbean regions. North America is represented very weakly, as there is data for only 1 out of 4 cultures for the Arctic and Subarctic subregion in the sample, and data for the subregion Plains and Plateau is completely missing. Europe

and the Middle East are fully represented in these models, while their cultures cover 3.8% of the model populations each. Oceania is again missing all data for the Melanesia subregion, being left with only one culture in Polynesia covering 3.8% of the model population. Considering both, the cultural count and distribution among regions and subregions a similar conclusion can be made as was with *models 1* and *4*. Highlighting the issue of poor representation of the North American subregion and no representation of Melanesia, the results may be generalized worldwide with a reserve.

Figure 8



4 Discussion

This research aimed to contribute to answering what are the main functions of dance in humans from the evolutionary perspective, specifically using the framework of signaling theory. The study used this approach to find out whether dance functions for courtship or other purposes such as social bonding and cooperation. In this section, I am first going to pay additional attention to the results of the analysis, and afterward, I will discuss study limitations, together with recommendations for future research.

4.1 Summary of the results

Some of the model confidence intervals were in line with the hypotheses' expectations, but these effects were not statistically significant. The only statistically significant effect was of the degree of social stratification on whether men dance in nuptial rituals, but here the intervals suggested an opposite effect as the one expected by the second hypothesis. Therefore, the overall outcome of the analysis is that results suggest dance functions for other purposes than courtship.

The statistically significant effect of the degree of social stratification was found in the model where it was a single fixed effect and persisted in the model where the degree of polygyny was also included as the second fixed effect. However, it needs to be highlighted that the models including the social stratification had very poor representation in subregions outside of Africa. Therefore, the statistically significant results cannot be generalized worldwide. Nonetheless, the confidence intervals indicate that with the rise of social stratification, the presence of dance in nuptial rituals rises. At the same time, the effects of polygyny that are not statistically significant suggest that intrasexual competition is not increased in the model with both fixed effects. The result can be interpreted as dance functioning for some other purposes than courtship on a stronger basis than it is in the case of the rest of the models from which the conclusion is made based on keeping the null hypothesis. At the same time, the results of models including social stratification can be claimed as representable only for Africa. This issue, however, does not impact the overall conclusion. Dance may function as a facilitator of social bonding or cooperation at the group level. As the narrative materials used were collected in the latest centuries, it is possible to conclude that the results support the suggestion made by Fink and colleagues (2021). Dance might have evolved for courtship (as in other species) that takes place between individuals, but in humans, it has transformed into a behavior that takes place in groups, with possible functions such as social bonding, or cooperation enhancive effects.

Hypotheses were tested in the context of nuptial rituals that are strongly associated with reproduction. As the results from this background suggest dance in nuptial ritual functions for other purposes than courtship, it may be speculated that they apply

to dance more broadly. Future research might consider testing specific social functions and, at the same time, using data about dance in general, without contextual restriction to gain more universal results.

4.2 The other functions of dance

Other than for courtship, social bonding, or cooperation, researchers suggest that dance may function to communicate (transfer) religious "knowledge", or historical knowledge (Fink & Shackelford, 2017; Grammer, et al., 2011). Its choreography may tell a specific story, such as myth, legend, or reference historical occurrences. The location where the activity takes place may similarly bear a certain meaning. Dancers may further promote their wealth, the wealth of their family, or a group they are affiliated with, by using an expensive costume for the dance. The usage of dance for cultural transfer is hardly explainable by the theory for its other functions discussed in this study (courtship, bonding, or cooperation enhancive effects of dance) and it should also be taken into account in future research. As it was not explored, it cannot be said that such behavior is either aligned with or in conflict with the results of this study. But it seems unique to humans and is certainly of great interest.

An individual may be the sole performer of a dance who is observed by many and also may be the sole performer observed by an individual. Dance that carries information about historical or religious events or beliefs received much less attention in the past research and there might be cases where it happens outside and inside of a group. Due to this uncertainty, I find it necessary to note that dance outside of a group (between two individuals) may not necessarily carry reproductive implications, at least not in humans. A qualitative approach is encouraged for research to explore whether such situations occur, how exactly they look, in what circumstances they happen, and how often they appear in the ethnographies. In this sense, dance still needs to receive a lot more attention.

4.3 Limitations and recommendations

This chapter will address several theoretical and methodological issues encountered during this research along with their possible solutions. It will also suggest some adjustments for cross-cultural databases that might help improve the research process for future studies.

4.3.1 Female signaling

In this work, the approach to sexual signaling was rather forthright. Cultural behaviors were understood as a non-dynamic consequence of biological processes and needs.

Meaning, that biological changes such as hormonal fluctuation, together with cultural changes stemming from them that were previously not reflected in much detail. This, dare I say classic approach, is more likely to work for studying human male sexual signaling as it seems rather stable, but it may not be so for signaling in females, where the hormonal fluctuation is regular and significant due to the menstrual cycle. At the same time, age or experience may play a role in the adjustment of individuals' signaling patterns.

I have mentioned previously, that there is a general lack (not just in humans) of research on female signaling and the relevant studies show us that signaling in females may adjust throughout their lifetime. In tobacco budworms (*Chloridea [Heliothis] virescens*) females are both choosers and signalers and their calling behavior, mating latency, and mate choice change during their lives. When compared to mated females, virgin females that were equally ready to mate signaled longer and showed weaker mate preference. Compared to virgin females, mated females may improve their fitness by being pickier about who they mate with and they may utilize their signaling behavior to do so (Zweerus et al., 2022). At the same time, the signaling of women may be dependent on their menstrual cycle. Human females may for example wear different clothes or make-up during different parts of the cycle. To gain a proper understanding of female sexuality a greater insight into culture-biology interactions is necessary (Wallner et al., 2019). The same goes for the related signaling as in comparison to men, female signaling is likely to be more dynamic and harder to identify.

Female signaling patterns need to be examined in a large number of species to improve the current knowledge of female signaling as it seems that the past research underestimated the degree to which females signal. It applies especially to humans, where a complex culture has evolved. Here, this remark is relevant especially for the fourth hypothesis, as understanding female signaling may influence the way we construct our questions regarding behaviors such as mate-guarding. If we better understand female signaling, we may improve our insight into circumstances in which mate-guarding appears and how it may change depending on age, and possibly other factors such as social status. At the same time, this knowledge may also influence how we perceive males and their role in the sexual selection of many species. Namely, if it is found out males may be choosers in the reproductive dynamic significantly more often than previously understood. Improved knowledge of the topic of female signaling may greatly affect how we construct our hypotheses regarding mate selection.

It also needs to be pointed out, that narrative materials from the past two centuries, such as the material in eHRAF, will likely contain limited information that could be useful for research on female signaling similar to that cited above. There is often a lack of detail in the ethnographic descriptions that is caused by the nature of the records. It is unlikely that the collectors would have access to accurate information about topics that are subject to taboo in the society where they are guests. At the same time, these materials often do not go too much into detail in their descriptions. The

menstrual cycle is a great example of why there may be a lack of data. Topics such as menstrual seclusion (e.g., menstrual huts) may be described in the material, approximate times women spend in seclusion too as these are information that can be easily observed. But that is probably as far as the information would go in most cases since the topic is unlikely to be discussed willingly by informants. When it comes to topics such as this one, the gender of the collector is important. Here the issue especially applies if the collector is a man, as the majority of collectors in these periods were.

4.3.2 Polygyny, culture, and the random effects

One of the challenges is connected to the first and second hypotheses that deal with polygyny. Polygyny should allow men to have more partners, which should raise competition in the given society according to the theory presented. That, however, may not be necessarily true in all cases. If the population ratio is skewed and there are more women than men living in *a group*, the competition may not rise as expected. At the same time, other behaviors could take place if there is a lower number of women in *a group*. In such a scenario the most competent men could outcompete less competent men to the degree where it makes more sense for them to invest in their relative's offspring instead of directly reproducing themselves. The latter is a bit extreme scenario but is useful to illustrate the problem as the population ratio was not controlled for in this study. This issue might have impacted the results of this study since they indicate that polygyny did not increase intrasexual competition as was expected based on past research. Future research should take this variable into account. Considering the high influence of the random effects in my study, future research might want to reconsider what should be the scale of *a group* in the grouping variable. The way human societies are currently grouped into *cultures* might not be the most accurate way of categorization achievable and such units of analysis might be too broad. If, for practical reasons, we keep the cultures defined as they are, we need to pay a lot of attention to possible inner differences. This may especially apply to cultures that are spread in a larger geographical area. Accounting for this spread properly may help improve the accuracy of results. The way this can be done was demonstrated in the case of the culture Ojibwa in this study, however, if such a process is to be done for every culture a deep dive into the process of collection of the original narrative data may be necessary. This is a challenge that might be tackled by the developers of cross-cultural databases in the future.

Generally, the large influence of random effects (*culture*) in the models indicates that the analyzed behavior is strongly associated with the cultural group or a geographical location. The τ_{00} statistics for all models further indicate that there may be substantial differences in rules or customs regarding dance on the level of regions smaller than those inhabited by *cultures* as categorized within the SCCS. This makes complete sense as the cultural names should only be understood as categories under which we gather what we evaluate as similar based on many attributes (such as

geographical location, language, etc.) whose values may hardly be perfectly aligned. In some cases, these values carry more differences than in others. In cross-cultural studies, especially worldwide, this reminds us that even a small unit of analysis such as *culture* (by population, geographical coverage) may still be too large and not properly account for the diversity of human customs. I must agree with Watts and colleagues (2022) that the units of analysis should ideally be selected specifically for individual research questions. And also, cross-cultural databases should be as clear and explicit as possible in the way they code these units, so researchers can choose the best fit for their studies properly.

The substantial influence of random effects in all of the models posed a question of how much importance accounting for them bears in cross-cultural research. There is no way of knowing how much influence on the outcome may be caused by these effects in simpler models such as linear regression or generalized linear models. As the generalized linear models (GLM) are the appropriate alternative for my data that does not account for the random effects, I have used them to recreate each of the models from the analysis.[44, 45] In most cases this led to a noteworthy increase in the statistical significance of both, intercept and predictors, while intercept statistics in several cases gained statistical significance. No additional predictors (compared to the models in the analysis) gained statistical significance even though their p-values decreased in the majority of cases. Therefore, the conclusion regarding any of the hypotheses would not change if I did not control for the random effects of *culture*. However, the overall insight into models for my data shows that controlling for the random effects might be of utmost importance in cross-cultural research as their influence can be expected to be strong. And it makes sense to expect the growing influence of random effects along with rising population or geographic coverage of the random effect variable. Future studies should always consider whether they can access data of better accuracy (with lower coverage) as *culture* and calculate the random effects of the variable they use to get a more detailed description of their data.

4.3.3 The marriage organization

Another challenge in this work is connected to the third hypothesis. Due to the nature of the data I had no way of controlling whether both families that are mentioned have the same rules regarding marriage. The data about marriage organization was paired

[44] See *models 1.1, 2.1, 3.1, 4.1, 5.1, 6.1, 7.1, 8.1, 9.1*, and *10.1* in the *suppl. 26*. Code for the reporting table containing the corresponding *model 1-10* is provided in the supplement.

[45] This is mentioned here and not in the analysis deliberately. GLM is mentioned to indicate how may some results change when the random effects of the unit of analysis such as culture are not accounted for, and the relevant code is provided for an interested reader. The purpose of this paragraph is to emphasize the importance of random effects in cross-cultural research, not to do an analytical comparison of the GLM and GLMM results.

with the rest of the dataset by culture and it was also originally coded this way. It is possible that the material collector described the occurrence only from the perspective of the culture where he was a guest. This should be the culture that is used as a parameter in the database. However, there is a possibility that the second family might have had different customs and rules regarding marriage compared to the culture primarily described in the material. The original data for marriage organization were coded on a six-degree scale and were recoded into a three-degree scale based on the degree of exogamy for analysis.[46]

However, it would be presumptuous to say that the differences between the individual degrees on the original scale are so large that they make it impossible for a marriage to happen between two groups categorized in adjacent scale points. It is possible that the "other" family in the record did not have the same rules, but conformed to the customs of the family described for economic, political, or other reasons. This might have caused some inaccuracy in the results and should be considered by future research using similar variables based on the narrative data. In my data, paragraphs often describe the marriage of a pair from neighboring villages. In such a case this problem is unlikely to surface in a noteworthy amount, but for results precision, it would be ideal to code whether this is always the case and account for it in the analysis.

4.3.4 Women and mate-guarding methods

The results for the fourth hypothesis may have been impacted by two factors that were not accounted for during the research preparation. First is that the research did not incorporate the status of women in question. Mate-guarding behavior of men may differ towards women who are already married and who are engaged or brides to be. It is possible that if research tested only for already married women, the results would support the alternative hypothesis, especially in monogamous societies, as being cuckolded in such conditions would be very costly for men. The impact of a harsher environment might be significantly different in such a situation. However, information this detailed is not available in the ethnographic materials, therefore, this research was unable to control for the variable of women's marital status.

The other factor might be coded from the ethnographic materials in the future. In the fourth hypothesis, the cases where women do not dance were considered as results of mate-guarding pressure. This study, however, did not control for other methods of mate-guarding. It is possible that in societies where other mate-guarding methods are practiced, there is no need to forbid women from dancing. For example, if a woman is sufficiently covered, she may dance even if the environmental conditions are harsher in the region, as there is little opportunity to signal her quality. Since, circumcised women frequently have trouble reaching orgasm, lack sexual desire, and are less satisfied by sex in general (El-Defrawi et al., 2011), circumcision may lower the chances

[46] See *appx. E* and *suppl. 11*.

of them being adulterous. Therefore, if a society in a harsher environment practices female circumcision, women may be allowed to dance as they may already be considered unlikely to cheat on their husbands. The mate-guarding practices present in individual societies should be taken into account in future research.

4.3.5 Coding

One of the limitations about which there is no clear scientific consensus is the ideal method of coding the narrative material. Materials for studies are most commonly coded in three different ways: by a single individual, by several people while not all of them cover every item creating partial overlap between coders, and by several people while each coder fully codes the material. This study belongs to the last category.

The results of IRR did show that the agreement between coders in my study was often not too large, indicating some results should be discounted[47] as the reliability of the material might not be ideal or that there are imperfections in the coding manual that confused the coders. However, there is no clear consensus among researchers on what amount of agreement should be accepted as reliable, and this applies generally, not just in cross-cultural research based on narrative material. I must also say that the interpretative power of calculation of agreement in case there are only some overlaps between the coders drops as only the overlaps are represented in such calculation. Moreover, if there is only a single coder for the entire material there is no agreement to calculate at all. A discussion should be carried out on whether such a coding process should not mean the results need to be discounted by their very nature, as may be problematic to evaluate the quality of the coding. The only lead to evaluate the coding quality in such a situation is the analysis of the coding manual. However, even if the manual is greatly detailed, some bias may still be reflected in the data from the side of the coder. At the same time, a coding manual that is too descriptive may be difficult to follow in practice.

Decisions based on the coder's comments and authoritative material revisions described in the methods were taken to account for the disagreement in this study. I must argue that such an approach carries the least bias among the coding methods mentioned, even if there is a noteworthy amount of disagreement between the coders. This kind of procedure gives the study better control over the quality of the coded data. It is up for future debate to properly assess whether, or to what degree should results be discounted if such methods are used in cross-cultural research based on narrative material.

The possibility of mistakes or imperfections in the coding manual was already mentioned, but some specific issues must be discussed. The coding manual did not account for situations where an individual might have been dancing strictly in the presence of their partner, who is of the opposite sex. This might have caused some

[47] According to Krippendorff (1980).

distortion in the results, as dancing in some contexts may be allowed between already established partners, even if neither of them would be able to dance with other people of the opposite sex. In the context of the fourth hypothesis, where mate-guarding is involved, it might be unacceptable for women to dance in front of men in some cultures present in harsher environments. The way data were coded, the situation of partners dancing together in seclusion would influence the results towards the outcome indicating that it is acceptable for women to dance even if the environment is harsh in the region where the culture lives. Therefore, the estimates would be moved in a direction contrary to the true state of the population. As the narrative descriptions often contain information from informants and not only from direct observations, such situations can be described in the materials, even if it would have been impossible for the collector to directly observe them. This problem might have affected the outcome of the analysis and should be properly controlled for in the future.

The other problem is similar and is related to the seclusion of genders. In the analysis, it was automatically assumed that if only men or only women dance and there are no observers of the gender opposite to the dancers', the separation is perfect. However, this may not always properly reflect reality. Due to the nature of the descriptions it usually cannot be confidently claimed that genders were entirely secluded. Ethnographers often do not mention whether the spaces where such dance takes place were completely separated from anyone outside. It might not be possible to confirm whether there have been some random bypassers who could have acted as possible signal receivers. At the same time, it was assumed that there was no one to receive the signal. In this research, it was not coded whether there is an explicit mention of complete space separation. If future research incorporates this into the coding manual, it may help assess the extent to which undescribed contact may occur, which may cause some degree of error in the results.

4.3.6 Time foci metadata

This study did not properly account for cultural change over time, owing to the previously mentioned poor quality of date metadata. This may have resulted in an offset in results because the tested observations were grouped by culture without regard for cultural development over time. Future research should determine the best sampling strategy, specifically how to deal with observations that lack metadata about the time when the researcher was present in the field and the time period they describe in their work. If there were no missing metadata, I would recommend categorizing culture's observations based on their time focus while still accounting for some relationship between a specific culture in the earlier and later periods. In practice, if the future analysis was done using the generalized linear mixed effect models, creating a time period variable and using it as a random effect could indicate how much time periods affect the results. However, a much deeper understanding of the data might be gained using

time series analysis. It is specifically designed to account for development in time and could even predict possible development in individual cultures based on multiple observations of the same culture from different periods (e.g., years).

It is, however, difficult to deal with missing metadata because we lack them for a large portion of observations. Excluding observations that lack time metadata from samples entirely may result in misleading results, as may giving the observations less weight in statistical analysis. These are also values that may be difficult to properly fill using predictive modeling if we do not acquire a substantial amount of information about the collectors. Moreover, it is likely that if research into the lives of the collectors was conducted the information about the time foci of their visits would have been often found even without predictions. If the database creators can provide at least an approximation of the missing time periods, it could be extremely helpful. Unfortunately, as mentioned, a lot of additional research into the data collectors' lives would be necessary, which means such data will likely not be available soon, and it is possible that in some cases no information may be available after all.

As a lot of metadata is missing it is impossible to determine how much random error was created by not properly matching time periods in the data (Divale, 1975; Ember & Ember, 2001; Ember et al., 1991). Right now, it may be assumed that from the perspective of the time foci issue the results' quality may be best assessed for tests using variables from the ecoClimate database and polygyny. For the former, we have high-quality metadata regarding the time of data recording, and the latter was subjected to the test by Bahrami-Rad, Becker, and Henrich (2021). As discussed in the theoretical part, their test is not methodologically perfect. It was done for only a few variables and needs further development to be widely usable. However, right now it is probably the most reliable validity test done on this topic and at the same time the best published approach focused directly on cross-cultural research using ethnographic material.

4.3.7 Cross-cultural databases: an outlook to the future

To improve the entire research process that uses cross-cultural databases more might be done than just filling in the already-mentioned holes. Previously, I have especially discussed the time foci metadata, but for assessing the reliability of data a few other pieces of information might prove very useful and some are even crucial.

I have mentioned above that I came upon a work from which the data in D-PLACE were derived that was not accessible online. This is a major problem, as one of the purposes of these databases is to make quantitative cross-cultural research more effective time and resource-wise. The study can likely be found in some physical archives, but that is a large complication in the research process. Without it, the way data were gathered or transformed remains unknown and that causes reliability issues. In practice, there may be legal obstacles in hosting such studies directly in the database,

but D-PLACE should explore possibilities of providing some summary information that would help clear up the processes of data collection and transformation. Having this kind of summary would be useful for all variables and datasets in D-PLACE. Along with this, listing studies that were cited in the source works, might not be a legal issue and might be of great use.

In the analysis, I mentioned that eHRAF provides information on the background of material collectors, but the way it is provided makes it confusing. For example, there are authors with backgrounds such as "Ethnologist and Government Official", "Natural Scientist, Ethnologist" or "Soldier, Government Official" each pair of quotes for a single person. In assessing the reliability of the information they provide, it would be much clearer if the material collectors' *background* was divided into their primary *education* and *profession*. Considering the time periods most of these narrative materials are from, it might also be important to know the country of origin and gender[48] of the collector. The information about the collectors would be clearer and more ready for analysis. Same as is with the time foci data, there are cases when the background information is unknown. It would be helpful to provide information on whether this data is actually missing, or whether the database management did not yet go through the existing materials to provide the information. A large amount of the information mentioned is already in eHRAF but in a very impractical form. Lastly, eHRAF must publish clear guidelines regarding the document evaluation metadata, if there are any. Researchers cannot rely on a name of a single specialist who evaluated the document to consider its reliability. Without proper rating standards, we can never be sure about the presence of possible biases in the evaluation.

In the future, it would be helpful if eHRAF's paragraph search could use an algorithm to determine whether some paragraphs surrounding the search results are relevant for providing context for coding. It would greatly improve the effectiveness of this kind of research as I had to do it manually. I know this is a difficult thing to request, but considering the recent growth of artificial intelligence it is not such an ambitious outlook anymore as it would be a few months ago.

[48] This can be assumed from collectors' names, but gender-neutral names may be a complication. Furthermore, having the variable would be ideal so the information is ready to use along with the rest of the metadata.

Conclusion

This work focused on researching the functions of dance from an evolutionary point of view, using the framework of signaling theory. It was centered around dance as a behavior that has evolved to communicate information between individuals and attempted to explain what are the purposes of dance in connection to nuptial rituals.

By analyzing dance in nuptials I reached the conclusion that dance in this specific context functions for different purposes than courtship. According to the analysis results, none of the alternative hypotheses could have been accepted and results for testing of the second hypothesis indicate effects opposite to the hypothesis expectation. The hypotheses were built to show whether signaling of mate quality, which is essential in courtship, is a fundamental part of contemporary[49] human dancing behavior. For nuptial rituals, results indicate that mate quality signaling does not bear much significance. Future research should study dance in a context that is more outspread. Since nuptials are rituals strongly associated with reproduction, it can be expected that a broader study of dance might uphold the conclusions of this work.

At the same time, replication of this work on a larger dataset could help improve both the precision and generalizability of the results. Initially, this research was designed as a worldwide cross-cultural study. But as there turned out to be a lot of missing data, some regions are not represented as well as they should be and some are not represented at all. This varies for each model and starting with a larger sample might help avoid the issue. Nonetheless, the results are in all cases sufficiently representative for Africa.

This study paid extensive attention to methods. Methods used in this research were lately being challenged in scientific discussions, and I cannot avoid saying that certain parts of the research process still need further development. Among them are the main tools used for such research: cross-cultural databases, especially in terms of the metadata they provide.

[49] Even though some of the materials are more than 100 years old, from the evolutionary point of view, this term is appropriate.

Bibliography

1. Adshead-Lansdale, J., & Layson, J. (2006). *Dance History: An introduction*. Taylor & Francis e-Library.
2. Archer, D., & Akert, R. M. (1977). Words and everything else: Verbal and nonverbal cues in social interpretation. *Journal of Personality and Social Psychology, 35*(6), 443-449. https://doi.org/10.1037/0022-3514.35.6.443
3. Bahrami-Rad, D., Becker, A., & Henrich, J. (2021). Tabulated nonsense? Testing the validity of the Ethnographic Atlas. *Economics Letters, 204*: 109880. https://doi.org/10.1016/j.econlet.2021.109880
4. Binford, L. (2001). *Constructing Frames of Reference: An Analytical Method for Archaeological Theory Building Using Hunter-gatherer and Environmental Data Sets*. University of California Press.
5. Blythe, P. W., Todd, P. M., & Miller, G. F. (1999). How Motion Reveals Intention. In G. Gigerenzer, P. M. Todd, & the ABC Research Group. *Simple heuristics that make us smart* (pp. 257-85). Oxford University Press.
6. Boucher, C., & Altamimi, Z. (2001). ITRS, PZ-90 and WGS 84: current realizations and the related transformation parameters. *Journal of Geodesy, 75*, 613-619.
7. Brown, W., Cronk, L., Grochow, K., Jacobson, A., Karen Liu, C., Popović, Z., & Trivers, R. (2005). Dance reveals symmetry especially in young men. *Nature, 438*, 1148–1150. https://doi.org/10.1038/nature04344
8. Buckle, L., Gallup, G. G., Jr., & Rodd, Z. A. (1996). Marriage as a Reproductive Contract: Patterns of Marriage, Divorce, and Remarriage. *Ethology and Sociobiology, 17*, 363-377. https://doi.org/10.1016/S0162-3095(96)00075-1
9. Burch, R. L. (2019). The Wedding as a Reproductive Ritual. *Review of General Psychology, 23*(3), 1-17. https://doi.org/10.1177/1089268019832848
10. Buss, D. M. (2003). *The Evolution of Desire: Strategies of Human Mating*. Basic Books.
11. Carroll, J. L. (2016). *Sexuality Now: Embracing Diversity* (5th ed.). Cengage Learning.
12. Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104
13. Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin, 70*(4), 213–220. https://doi.org/10.1037/h0026256
14. Colwell, R. K. (1974). Predictability, Constancy, and Contingency of Periodic Phenomena. *Ecology*, *55*(5), 1148–1153. https://doi.org/10.2307/1940366

15. Davis, J. N., & Daly, M. (1997). Evolutionary Theory and the Human Family. *The Quarterly Review of Biology, 72*(4), 407-435. https://doi.org/10.1086/419953
16. El-Defrawi, M. H., Lotfy, G., Dandash, K. F., Refaat, A. H., & Eyada, M. (2011). Female Genital Mutilation and its Psychosexual Impact. *Journal of Sex & Marital Therapy, 27*(5), 465-473. https://doi.org/10.1080/713846810
17. Divale, W. T. (1975). Temporal focus and random error in cross-cultural hypothesis tests. *Behavior Science Research, 10*, 19-36. https://doi.org/10.1177/106939717501000102
18. Ember, C. R. (2007). Using the HRAF Collection of Ethnography in Conjunction With the Standard Cross-Cultural Sample and the Ethnographic Atlas. *Cross-Cultural Research, 41*(4), 396–427. https://doi.org/10.1177/1069397107306593
19. Ember, C. R., & Ember, M. (2001). *Cross-cultural research methods*. AltaMira Press.
20. Ember, C. R., Ross, M. H., Burton, M. L., & Bradley, C. (1991). Problems of Measurement in Cross-Cultural Research Using Secondary Data. *Cross-Cultural Research, 25*, 187-216. https://doi.org/10.1177/106939719102500109
21. Ember, M. (1997). Evolution of the human relations area files. *Cross-Cultural Research, 31*(1), 3–15. https://doi.org/10.1177/106939719703100101
22. Fink, B., & Shackelford, T. K. (2017). Why Did Dance Evolve? A Comment on Laland, Wilkins, and Clayton (2016). *Evolutionary Psychological Science, 3*, 147-148. https://doi.org/10.1007/s40806-016-0075-3
23. Fink, B., Bläsing, B., Ravignani, A., & Shackelford, T. K. (2021). Evolution and functions of human dance. *Evolution and Human Behavior, 42*(4), 351-360. https://doi.org/10.1016/j.evolhumbehav.2021.01.003
24. Fox, J., & Weisberg, S. (2019). An {R} Companion to Applied Regression [R package version 3.0-13, Third Edition]. Sage Publications. https://socialsciences.mcmaster.ca/jfox/Books/Companion/
25. Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement* [R package version 0.84.1]. https://CRAN.R-project.org/package=irr
26. Gangestad, S. W., & Simpson, J. A. (2000). The evolution of human mating: Trade-offs and strategic pluralism. *Behavioral and Brain Sciences, 23*(4), 573-644. https://doi.org/10.1017/s0140525x0000337x
27. Gaulin, S. J. C., & Boster, J. S. (1990). Dowry as Female Competition. *American Anthropologist, 92*(4), 994-1005. https://doi.org/10.1525/aa.1990.92.4.02a00080
28. Geary, D. C., Vigil, J., & Byrd-Craven, J. (2004). Evolution of Human Mate Choice. *Perspectives on Human Sexuality, 41*(1), 27-42. https://doi.org/10.1080/00224490409552211

29. Grammer, K., Oberzaucher, E., Holzleitner, I., & Atmaca, S. (2011). Dance: the Human Body as a Dynamic Motion System. In W. Tschacher & C. Bergomi (Eds.), *The implications of embodiment: Cognition and communication* (pp. 173-192). Imprint Academic.
30. Hagen, E. H., & Bryant, G. A. (2003). Music and dance as a coalition signaling system. *Human Nature, 14*, 21–51. https://doi.org/10.1007/s12110-003-1015-z
31. Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23-34. https://doi.org/10.20982/tqmp.08.1.p023
32. Hanna, J. L. (1987). *To dance is human: A theory of nonverbal communication*. University of Chicago Press.
33. Hanna, J. L. (1988). *Dance, sex and gender: Signs of identity, dominance, defiance, and desire*. University of Chicago Press.
34. Hugill, N., Fink, B., & Neave, N. (2010). The Role of Human Body Movements in Mate Selection. *Evolutionary Psychology, 8*(1), 66-89. https://doi.org/10.1177/147470491000800107
35. Hugill, N., Fink, B., Neave, N., Besson, A., & Bunse, L. (2011). Women's perception of men's sensation seeking propensity from their dance movements. *Personality and Individual Differences, 51*(4), 483-487. https://doi.org/10.1016/j.paid.2011.05.002
36. Hugill, N., Fink, B., Neave, N., & Seydel, H. (2009). Men's physical strength is associated with women's perceptions of their dancing ability. *Personality and Individual Differences, 47*(5), 527-530. https://doi.org/10.1016/j.paid.2009.04.009
37. Iannone, R., Cheng, J. Schloerke, B., Hughes, E., Lauer, A., & Seo, J. (2023). *gt: Easily Create Presentation-Ready Display Tables* [R package version 0.9.0]. https://CRAN.R-project.org/package=gt
38. Johnstone, R. A. (1997). The Evolution of Animal Signals. In J. R. Krebs & N. B. Davies (Eds.), *Behavioural Ecology: An Evolutionary Approach* (4th ed.) (pp. 155-178). Oxford University Press.
39. Jorgensen, J. G. (1980). *Western Indians: Comparative Environments, Languages, and Cultures of 172 Western American Indian Tribes*. W. H. Freeman and Company.
40. Kaeppler, A. L. (1978). Dance in anthropological perspective. Annual Review of *Anthropology, 7*, 31-49.
41. Kendall, M. G. (1948). *Rank correlation methods*. Griffin.
42. Kirby, K. R., Gray, R. D., Greenhill, S. J., Jordan, F. M., Gomes-Ng, S., Bibiko, H-J, Blasi, D. E., Botero, C. A., Bowern, C., Ember, C. R., Leehr, D., Low, B. S., McCarter, J., Divale, W., & Gavin, M. C. (2016). D-PLACE: A Global Database of

Cultural, Linguistic and Environmental Diversity. *PLoS ONE, 11*(7): e0158391. https://doi.org/10.1371/journal.pone.0158391

43. Knapp, M., Hall, J., & Horgan, T. (2013). *Nonverbal communication in human interaction*. Cengage Learning.
44. Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Sage Publications.
45. Kurath, G. P. (1960). Panorama of dance ethnology. *Current Anthropology, 1*, 233-254.
46. Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models [R package version 3.1-3]. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13
47. Laidre, M. E., & Johnstone, R. A. (2013). Animal Signals. *Current Biology, 23*(18), R829-R833. https://doi.org/10.1016/j.cub.2013.07.070
48. Laland, K., Wilkins, C., & Clayton, N. (2016). The evolution of dance. *Current Biology, 26*(1), R5-R9. https://doi.org/10.1016/j.cub.2015.11.031
49. Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.
50. Lang, H. (1998). CONAN: An electronic code-text data-base for cross-cultural studies. *World Cultures, 9*(2), 13-56.
51. Lang, M., Bahna, V., Shaver, J. H., Reddish, P., & Xygalatas, D. (2017). Sync to link: Endorphin-mediated synchrony effects on cooperation. *Biological Psychology, 127*, 191-197. https://doi.org/10.1016/j.biopsycho.2017.06.001
52. Lee, W., & Grimm, K. J. (2018). Generalized Linear Mixed-Effects Modeling Programs in R for Binary Outcomes. *Software Review, 25*(5), 824-828. https://doi.org/10.1080/10705511.2018.1500141
53. Light, R. J. (1971). Measures of response agreement for qualitative data: Some generalizations and alternatives. *Psychological Bulletin, 76*(5), 365–377. https://doi.org/10.1037/h0031643
54. Lightner, A. D., Bendixen, T., & Purzycki, B. G. (2022). Moralistic supernatural punishment is probably not associated with social complexity. *Evolution and Human Behavior*, [In press, corrected proof]. https://doi.org/10.1016/j.evolhumbehav.2022.10.006
55. Lima-Ribeiro, M. S., Varela, S., González-Hernández, J., de Oliveira, G., Diniz-Filho, J. A., & Terribile, L. C. (2015). EcoClimate: a database of climate data from multiple models for past, present, and future for macroecologists and biogeographers. *Biodiversity Informatics, 10*, 1-21. https://doi.org/10.17161/bi.v10i0.4955
56. Lüdecke, D. (2021). sjPlot: Data Visualization for Statistics in Social Science [R package version 2.8.10]. https://CRAN.R-project.org/package=sjPlot
57. Mansfield, E. R., & Helms, B. P. (1982). Detecting Multicollinearity. *The American Statistician, 36*(3), 158-160. https://doi.org/10.2307/2683167

58. Martin, J. S., Ringen, E. J., Duda, P., & Jaeggi, A. V. (2020). Harsh environments promote alloparental care across human societies. *Proceedings of the Royal Society B, 287*(1933): 20200758. http://doi.org/10.1098/rspb.2020.0758
59. Maynard Smith, J., & Harper, D. G. C. (1995). Animal Signals: Models and Terminology. *Journal of Theoretical Biology, 177*(3), 305-311. https://doi.org/10.1006/jtbi.1995.0248
60. Maynard Smith, J., & Harper, D. G. C. (2003). *Animal Signals*. Oxford University Press.
61. Monger, G. P. (2004). *Marriage Customs of the World: From Henna to Honeymoons*. ABC-CLIO.
62. Murdock, G. P. (1961). *Outline of Cultural Materials*. Human Relations Area Files.
63. Murdock, G. P. (1967a). *Ethnographic Atlas*. University of Pittsburgh Press.
64. Murdock, G. P. (1967b). Ethnographic Atlas: A Summary. *Ethnology, 6*(2), 109–236. https://doi.org/10.2307/3772751
65. Murdock, G. P. (1969). *Outline of World Cultures*. Human Relations Area Files.
66. Murdock, G. P., Textor, R., Barry, H. III, White, D. R., Gray, J. P., & Divale, W. T. (1999). Ethnographic Atlas. *World Cultures, 10*(1), 24-136.
67. Murdock, G. P., & White, D. R. (1969). Standard Cross-Cultural Sample. *Ethnology, 8*(4), 329–369. https://doi.org/10.2307/3772907
68. Naroll, R. (1967). The Proposed HRAF Probability Sample. *Behavior Science Notes, 2*(2), 70-80. https://doi.org/10.1177/1069397167002002
69. Pazhoohi, F., & Hosseinchari, M. (2014). Effects of Religious Veiling on Muslim Men's Attractiveness Ratings of Muslim Women. *Archives of Sexual Behavior, 43*(6), 1083-1086. https://doi.org/10.1007/s10508-014-0259-5
70. Pazhoohi, F., Lang, M., Xygalatas, D., & Grammer, K. (2017). Religious Veiling as a Mate- Guarding Strategy: Effects of Environmental Pressures on Cultural Practices. *Evolutionary Psychological Science, 3*, 118-124. https://doi.org/10.1007/s40806-016-0079-z
71. Pazhoohi, F., Macedo, A. F., & Arantes, J. (2017). The Effect of Religious Clothing on Gaze Behavior: An Eye-Tracking Experiment. *Basic and Applied Social Psychology, 39*(3), 176-182. https://doi.org/10.1080/01973533.2017.1307748
72. Pecka, D. (2020). *Mate-guarding as a theoretical framework to the study of wedding ritual practices* [Unpublished bachelor's thesis]. Masaryk University.
73. Puth, M.-T., Neuhäuser, M., & Ruxton, G. D. (2014). Effective use of Pearson's product–moment correlation coefficient. *Animal Behaviour, 93*, 183–189. https://doi.org/10.1016/j.anbehav.2014.05.003
74. RStudio Team (2020). *RStudio: Integrated Development Environment for R*. RStudio, PBC. http://www.rstudio.com/

75. R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/
76. Rappaport, R. A. (1999). *Ritual and Religion in the Making of Humanity*. Cambridge University Press.
77. Reddish, P., Fischer, R., & Bulbulia, J. (2013). Let's Dance Together: Synchrony, Shared Intentionality and Cooperation. *PLoS ONE 8*(8): e71182. https://doi.org/10.1371/journal.pone.0071182
78. Röder, S., Carbon, C., Shackelford, T. K., Pisanski, K., Weege, B., & Fink, B. (2016). Men's visual attention to and perceptions of women's dance movements. *Personality and Individual Differences, 101*, 1-3. https://doi.org/10.1016/j.paid.2016.05.025
79. Schaeffer, M. S., & Levitt, E. E. (1956). Concerning Kendall's tau, a nonparametric correlation coefficient. *Psychological Bulletin, 53*(4), 338–346. https://doi.org/10.1037/h0045013
80. Slingerland, E., Atkinson, Q., Ember, C., Sheehan, O., Muthukrishna, M., Bulbulia, J., & Gray, R. (2020). Coding culture: Challenges and recommendations for comparative cultural databases. *Evolutionary Human Sciences, 2*: E29. http://doi.org/10.1017/ehs.2020.30
81. Smith, J. Z. (1971). *History of Religions*. The University of Chicago Press
82. Tarr, B., Launay, J., Cohen, E., & Dunbar, R. (2015). Synchrony and exertion during dance independently raise pain threshold and encourage social bonding. *Biology Letters, 11*: 2015076720150767. http://doi.org/10.1098/rsbl.2015.0767
83. Tay, R. (2017). Correlation, Variance Inflation and Multicollinearity in Regression Model. *Journal of the Eastern Asia for Transportation Studies, 12*, 2006-2015. https://doi.org/10.11175/easts.12.2006
84. Weege, B., Pham, M. N., Shackelford, T. K., & Fink, B. (2015). Physical Strength and Dance Attractiveness: Further Evidence for an Association in Men, But Not in Women. *American Journal of Human Biology, 27*(5), 728-730. https://doi.org/10.1002/ajhb.22703
85. Wade, T. J., Weinstein, E., Dalal, N., & Salerno, K. J. (2015). I can dance: Further investigations of the effect of dancing ability on mate value. *Human Ethology Bulletin, 30*(2), 10-20.
86. Wallner, B., Windhager, S., Schaschl, H, Nemeth, M., Pflüger, L. S., Fieder, M., Domjanić, J., Millesi, E., & Seidler, H. (2019). Sexual Attractiveness: a Comparative Approach to Morphological, Behavioral and Neurophysiological Aspects of Sexual Signaling in Women and Nonhuman Primate Females. *Adaptive Human Behavior and Physiology, 5*, 164–186. https://doi.org/10.1007/s40750-019-00111-6
87. Watts, J., Jackson, J. C., Arnison, C., Hamerslag, E. M., Shaver, J. H., & Purzycki, B. G. (2022). Building Quantitative Cross-Cultural Databases From Ethnographic

Records: Promise, Problems and Principles. *Cross-Cultural Research, 56*(1), 62–94. https://doi.org/10.1177/10693971211065720

88. White, D. R., Burton, M. L. , & Whiting, J. W. M. (1986). Forms and frequencies of polygyny: Standard Sample codes. *World Cultures, 2*(2).
89. Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* [R package version 3.4.0]. Springer-Verlag New York. https://ggplot2.tidyverse.org
90. Wickham, H., François, R., Henry, L., Müller, K. (2022). *dplyr: A Grammar of Data Manipulation* [R package version 1.0.8]. https://CRAN.R-project.org/package=dplyr
91. Wiltermuth, S. S., & Heath, C. (2009). Synchrony and Cooperation. *Psychological Science, 20*(1), 1–5. https://doi.org/10.1111/j.1467-9280.2008.02253.x
92. Zweerus, N. L., van Wijk, M., Smallegange, I. M., & Groot, A. T. (2022). Mating status affects female choice when females are signalers. *Ecology and Evolution, 12*(4), e8864. https://doi.org/10.1002/ece3.8864

Appendix A Decision code (text)

As the ethnographic material was coded three times by three different coders there had to be a decision made about which coded values will be used for analysis in case there is a disagreement between the coders.

- The method chosen to do this was to use the value selected by the majority of coders (2/3).
- In case there was no overlap, the decision was made by the researcher based on the coder's comments.
- In case there were no comments to make the decision, the paragraph was revisited and the final decision was made by the researcher.
- Decisions were not done on the "Dance" column as this data is not relevant for hypothesis testing. This data was coded to calculate IRR and correlation between the degrees of coders' agreement and document evaluation, which was done before the data transformation. The column is therefore not present in *supplements 6-8* and *13*.

Appendix B Recode and enrich (text)

In *supplement 8*, *supplement 7* is recoded into binary values and enriched with variables region and subregion of cultures (eHRAF classification) and transformed data from D-PLACE datasets. There are two tables, on the left is the resulting merge table, with variables marked for clearance, on the right is the recoded and enriched table that is made into the *supplement 13* after being finished. Details regarding D-PLACE data transformation can be found in the thesis text and a*ppendices C*, *D*, *E* and *supplements 9-12*.

Variables from the *supplement 7* are marked with asterisks, and recoded variables are marked as recoded asterisks (such as r*) to make it clear which recoded variable originated in which code.

These are codes for the original three variables:

* 0 = only men dance; 1 = only women dance; 2 = both dance explicit; 3 = both dance implicit; N/A = unstated and unclear from context

** 0 = men; 1 = women; 2 = both explicit; 3 = both implicit; N/A = unstated and unclear from context

*** 0 = parties dance together; 1 = parties dance apart; N/A = unstated and unclear from context

The original variables were recoded into five new variables as follows:

r* Men dance: 0 = men do not dance (originally value: 1), 1 = men dance (originally values: 0, 2, 3), N/A = unstated and unclear from context

Women dance: 0 = women do not dance (originally value: 0), 1 = women dance (originally values: 1, 2, 3), N/A = unstated and unclear from context

r** Men observe: 0 = men do not observe (originally value: 1), 1 = men observe (originally values: 0, 2, 3), N/A = unstated and unclear from context

Women observe: 0 = women do not observe (originally value: 0); 1 = women observe (originally values: 1, 2, 3), N/A = unstated and unclear from context

r*** Location: 0 = parties dance apart; 1 = parties dance together; N/A = unstated and unclear from context

Appendix C Polygyny recoding

Original code:

0 Polyandry
1 Monogamy prescribed
2 Monogamy preferred, but exceptional cases of polygyny
3 Limited polygyny <20% of married males
4 Full polygyny 20% or more of married males

Recoded as follows:

0 Non-polygynous: includes cultures coded as 0 and 1 in the original code
1 Minimal polygyny: includes cultures coded as 2 in the original code
2 Polygynous: includes cultures coded as 3 and 4 in the original code

eHRAF lists all Ojibwa regions under the same OWC, however, in my data none of the Ojibwa records is localized in the same region as Northern Saulteaux (Northern Ojibwa) - not to be confused with North Albany Ojibwa.
Therefore, data for Ojibwa from this dataset could not be paired, so records of polygyny variable were replaced with missing data in the dataset for hypotheses testing (*supplement 13*).

Appendix D Social stratification recoding

Original code:

1 no differences in access to economic resources, political power, and/or status
2 differences in access to economic resources, political power, and/or status, not resulting in class formation
3 two classes
4 complex stratification into more than two classes

Recoded as follows:

0 Not stratified: includes cultures coded as 1 in the original code
1 Partly stratified: includes cultures coded as 2 and 3 in the original code
2 Highly stratified: includes cultures coded as 4 in the original code

Appendix E Marriage organization recoding

Original code:

1 Demes, i.e., communities revealing a marked tendency toward local endogamy but not segmented into clan-barrios

2 Segmented communities, i.e., those divided into barrios, wards, or hamlets, each of which is essentially a localized kin group, a clan or ramage, in the absence of any indication of local exogamy. Large extended families (see "Domestic organization"), are treated as clan-barrios if they are integrated by a rule of ambilineal, matrilineal, or patrilineal descent.

3 Agamous communities without localized clans or any marked tendency toward either local exogamy or local endogamy

4 Exogamous communities, i.e., those revealing a marked tendency toward local exogamy without having the specific structure of clans

5 Segmented communities where a marked tendency toward local exogamy is also specifically reported

6 Clan-communities, each consisting essentially of a single localized exogamous kin group or clan (see "Organization of clan communities" for more detail)

Recoded as follows:

0 Tendency towards endogamy or absence of an indication or tendency towards exogamy: includes cultures coded as 1, 2 in the original code

1 Agamous communities without localized clans or any marked tendency toward either local exogamy or local endogamy

2 Exogamous, or tendency towards exogamy: includes cultures coded as 4, 5 and 6 in the original code

Appendix F Agreement recoding

Agreement recoding in *supplement 21* was done as follows:

Document evaluation:
perfect = 2 (orig. value "5")
good = 1 (orig. value "4.5")
average (orig. values "3", "3.4", "4")
missing = N/A (orig. value "none")

Agreement between coders:
full (all coders agreed) = 2
partial (2/3 agreed) = 1
none (all 3 coded differently) = 0