



Universiteit Utrecht

A Game of Cultures:

Affinity-based Communities in Online Interaction

A.C. Vredevoogd, B. Teunissen



July 2015

MA Thesis Intercultural Communication

Supervisor: Dr. R.J.L. Supheert

Second Reader: Dr. M.E.H. Schouten

Preface

This thesis is the result of a collaborative effort between Alexandra Vredevoogd and Bas Teunissen; written as a part of their MA Intercultural Communication at Utrecht University.

The division of labour was as follows:

The concept, construction of corpus and model were established through a joint effort between the authors. The corpus analysis was performed simultaneously by both authors, with each author analysing half of each corpus set. Statistical analysis of the resulting datasets was performed by Alexandra Vredevoogd.

The Introduction, Literature Review and Discussion were drafted by Bas Teunissen; the Method, Results and Conclusion were drafted by Alexandra Vredevoogd. The final version was edited by both authors.

We would like to express our thanks to the following people for making this project possible; your support and advice has been invaluable to us.

Dr. Roselinde Supheert

For her boundless patience, feedback and enthusiasm for our project;

Gerda Blees, MA

For taking the time to share her statistical expertise with us;

Dr. Bert Schouten

For providing valuable feedback allowing us to improve upon our pre-final version;

ALMG B.V.

For their hospitality by providing us with a productive and stimulating environment to work in;

Our parents:

Willem & Cara Vredevoogd;

Meta vd Vliert, Alex Teunissen & Luuk Tepe.

Thank you for your unwavering love, support and hospitality.
We couldn't have done it without you

Table of Contents

1. Introduction.....	1
2. Literature Review.....	7
2.1 Cultural Dimensions	7
2.2 Online Communication.....	15
2.3 Culture and Community.....	19
2.4 Linguistic Features and Cultural Dimensions.....	23
2.5 Cultural Dimensions Online	28
2.6 Cultural Dimensions Framework.....	31
3. Method	34
3.1 Online Data Collection	35
3.2 Framework for Cultural Analysis	41
3.2.1 Directness/indirectness	41
3.2.2 Orientation to Content/person.....	43
3.2.3 Orientation to self/other	43
3.2.4 High context/low context.....	45
3.2.5 Idiomatic Context.....	46
3.3 Reliability Testing.....	48
3.4 Data Analysis	55
4. Results.....	59
4.1 Results Data Set: Game of Thrones	60
4.2 Results Data Set: Bioshock Infinite	63
5. Discussion	66
6. Conclusion	72
6.1 Limitations	74
6.2 Suggestions for Further Research	76
Bibliography	79
Appendices.....	82
Appendix 1: Forum URLs	82

1. Introduction

The Internet is a phenomenon that has been rapidly expanding over the last twenty years. In just over two decades the global Internet-user population has grown to 1.6 billion in 2009, and up to 2.4 billion by 2012 (World Internet Users Statistics and World Population Stats, 2012). This means that well over a quarter of the world's population is now making use of the Internet and, with ever advancing technology, it is likely that the Internet will continue to grow further. Crystal (2001) states that "the linguistic consequences of evolving a medium in which the whole world participate [...] are bound to be far-reaching" (2001, p. 5). The rapid increase of Internet users from around the globe has turned the Internet into a catalyst for intercultural interaction on a scale never seen before. In recent years, social media in particular fulfil an emancipatory function in allowing people to express themselves, voicing their identity, opinion and ideas through the Internet. The Internet has rapidly become the way to reach target audiences from across the globe, making it impossible to for marketing, PR, and public diplomacy to ignore. In fact, it has become a key element in the core-strategy of corporations and governments alike. With the Internet now beginning to play a major role in communication and events that change the direction of entire nations (e.g. most recently the situation in Ukraine) it is becoming more and more important to explore and research the role and influence of the Internet on culture and interaction.

The bulk of Internet related research has taken place in recent years (2000s and onward) but it is important to note that even though much of the research is only close a decade old, it might already be outdated due to the rapidly developing trends and technologies. The field of research most relevant to the present study has surfaced in the past five years concerns Computer-mediated Communication (CMC). CMC research looks at any communication that occurs by way of electronic devices (e.g. Instant messaging, e-mail, chat rooms, etc.).

Additionally, research has begun to emerge relating to the field of intercultural communication in CMC environments, including: language on the Internet (Crystal, 2001), second language use and learning (Thorne, Black, & Sykes, 2009; Hewling, 2006) and (genre specific) cultural adaptations (Snelders, Morel, & Havermans, 2011). However, researchers have yet to pay adequate attention to the Internet as a social space where communities emerge, develop, change and engage in interaction, regardless of their users' national identity. This interaction on an intercultural level in communities and spheres where anonymity blurs the lines that would otherwise set people apart (i.e. national culture, social class, gender, politics, etc.) has yet to be fully explored.

In the current research project culture is a pivotal concept that must be addressed to convey its significance for this project. Culture has been notoriously difficult to define. The very nature of culture is complex in a way that it seems to defy any single definition imposed upon it. Traditionally, scholars like Hall and Hofstede (1966, 1976; 2001) have tried to capture and categorise culture by equating it to national identity and national culture. This essentialist point of view presents culture as a static concept. By reducing cultures and their members to an almost stereotypical set of values, these definitions do not acknowledge the internal diversity and change that exist in the form of ethnic and linguistic variations. In essence it assumes that individuals primarily identify themselves as members of a particular nation state (Hewling, 2006, p. 339). The dimensions of Hall and Hofstede are an attempt to predict and analyse cultural aspects that will only work to a certain degree and with varying degrees of success (Ess & Sudweeks, 2006, p. 183), and a new approach for analysing ICC online that goes beyond simple polarities is needed (Ess & Sudweeks, 2006, p. 186). This new approach, which will be explored in the next chapter, is in tandem with a different definition of culture.

Unlike the essentialist view, where culture is static, the non-essentialist view of culture regards culture as something dynamic and multi-faceted, relying on the idea that culture is not

bound to a geographical place (Holliday, Hyde, & Kullman, 2010, p. 73) but a social force that flows between people in different ways. Holliday states that “It both binds and separates us” (2010, p. 74), meaning that a non-essentialist view does not deny cultural differences, but it makes clear that they cannot be boxed in a limited essentialist framework. In this paper we view culture from a broader perspective, regarding culture as something that is dynamic and interactive, a never ending process. Culture is also layered, as described by the small cultures approach (2010). The small cultures approach considers all socially cohesive behaviour as a manifestation of culture (2010, p. 74). In this sense, “the world is made up of a vast complex of shifting, overlapping, swirling, combining and splitting cultures” (2010, p. 74), where people engage in interaction; so “cultures do not talk to each other; individuals do” (Scollon & Wong-Scollon, 2001, p. 138). As a result, language plays a pivotal role in intercultural interaction. This ties in with Agar’s concept of *languacultures*, the idea that language and culture are so intricately related that they are, in essence, two side of the same coin (Agar, 1994). From this perspective, language is both indicative and product of culture and it follows that that a difference in language use is indicative of a difference in culture.

In this study online interaction will be analysed from an intercultural perspective, so it is essential to define the concept of intercultural communication (ICC) as it is used in this study. Ten Thije & Maier (2012) provide an overview of the field of intercultural communication, describing the core approaches used in intercultural research. A number of key definitions of intercultural communication are introduced: ICC as interaction taking place between individuals of different national backgrounds (i.e. Passports); ICC as any interaction where cultural differences are of influence on the interaction and ICC as only occurring when actors readjust their perspectives on life (knowledge, values, beliefs, etc.) as a consequence of the interaction. Intercultural communication can be interpreted in a broad sense or a narrow sense. The individual definitions by ten Thije do not specifically take new and emerging

developments into account, especially regarding individuals whose intercultural communication represents: “([I]) a multitude of ‘cultures,’ and ([II]) ‘culture’ as a series of practices and habits that are fluid, dynamic, and changing, especially as generated by intercultural communication online” (Ess & Sudweeks, 2006, p. 188). As none of the individual definitions fully encompass ICC in online environments, a broad perspective on ICC is used in this study.

Anonymity is both problematic and interesting for research in ICC and CMC. Williams and Copes explain that anonymity allows individuals to adopt a playful approach to the way they present themselves, allowing them to circumvent or ignore cultural and geographical boundaries posed by the real world (2011, p. 72). Anonymity seems to make users feel safer from the rejection of their identity by others, possibly causing users to adhere less strictly to the social or cultural expectations that they conform to in everyday life (e.g. being more honest or increasingly rude to others; or assuming the role of the opposite gender. This is all possible because identity is hidden and there is no face-to-face interaction) (Androutsopoulos, 2006; Williams & Copes, 2011; Turkle, 1995).

Community is a term that takes on a new meaning when it is applied to a virtual context. According to Denzin, an online community is defined as having its own “norms, its rules (netiquette), its own emotional vocabulary—guidelines for posting, acceptable subjects, regular users, leaders, old-timers, and a constant circulation of newcomers” (Denzin, 1998, p. 99–100). These communities then grow, shrink or change and “over time, members’ interactions develop into a discourse that structures the generation, activation, and diffusion [of] ideas, objects, and practices” (2011, p. 70), meaning that at some point new members step into, and adapt to this new discourse environment, rather than depending on their existing notions of interaction in a community. It is also important to note that users are active across various platforms and communities, switching between new discourse environments and

communities (e.g. different forums or different social media). A different definition, *Affinity-based Communities*, that specifically accommodates online communities in public, anonymous environments is proposed in chapter two.

For the present research, the corpus consists of forum threads. A thread is an asynchronous textual conversation, chronologically arranged by topic on the Forum. Users can start these topics and other users can in turn read and respond to this thread. According to Williams and Copes, these threads and the posts inside them “are cultural artefacts that are amenable to empirical content analysis” (2011, p.73). These threads are also the key location of interest for this study, because the online discussions amongst participants in online communities “are the manifestation of intercultural discourse and thus ‘the nexus of cultural production’” (Hewling, 2010, p.342). Based on the fact that interaction on the Internet is primarily textual, language becomes the primary catalyst through which interaction is constructed (Androutsopoulos, 2006). This makes the idea that language use equates to culture even more essential and it is the language use in these threads that will be analysed.

The aim of this paper is to gain insight in the way groups of users and communities achieve successful interaction in online contexts. This paper will discuss in what way online interaction is different from real-life interactional contexts and explore the idea that online communities seem to operate around affinity with a certain genre, topic, phenomenon or even individual (e.g. Korean drama, video games, celebrities, etc.), rather than aligning along nationality and socio-economic factors that play a major role in real-life interaction. To achieve this, this paper will perform quantitative corpus analysis on threads from discussion boards discussing the same topic. A new model of cultural dimensions is created based on the cultural dimensions by House (2007) and Hall (1966; 1976) and consists of the dimensions: *Directness/Indirectness, Orientation to Self/Other, Orientation to Content/Person, high*

context/low context and *Idiomatic Context*. The main question that this paper seeks to answer is:

Is there a difference in communicative style between Dutch and English threads on online discussion forums when discussing the same topic?

In chapter two this study will explore literature on the subjects of culture, cultural analysis, computer mediated communication and cultural analysis in online contexts. This will be followed by a description of the analysis in chapter three, the results of the analysis in chapter four and a discussion of the results in chapter five. Lastly, the conclusions drawn from this paper can be found in chapter six.

2. Literature Review

To answer our research question, it will be important to look at previous research on culture first, visiting the framework of dimensions as described by Hofstede for cultural analysis (2001), Hall for interpersonal communication (1966; 1976) and House for discourse analysis (2007). These models are established tools that enable insight into the differences and preferences between national cultures, but their effectiveness in the context of online interaction has not been fully explored nor operationalised. By looking at research from the fields of translation studies, marketing research and by exploring the newly emerging field of Computer Mediated Communication (CMC), this study attempts to lay the foundation for a model that facilitates contrastive linguistic research in online communication. In this model various cultural dimensions will be employed to analyse online communities.

2.1 Cultural Dimensions

The research by Geert Hofstede is one of the pioneering studies in establishing cross-cultural research. Hofstede conducted surveys among IBM employees globally, to establish a database of responses to examine differences in values, the tendency to prefer certain states of affairs over others (2001, p. 5), and to analyse these at a national, rather than individual level. During his time at IBM (1967-1973) Hofstede collected data from over 70 countries and regions. From the analysis that followed, Hofstede eventually formulated five cultural dimensions. Hofstede classified the results of his analysis along the following dimensions: Power Distance, Individualism/Collectivism, Masculinity/Femininity, Uncertainty Avoidance, and Long/Short-Term Orientation.

These dimensions have been commonly utilised in research on culture, lending itself particularly well to contrastive, cross-cultural research¹. However, there is criticism surrounding cultural dimensions, against Hofstede's dimensions in particular, and how successfully they can be used for cultural analysis. Minchov (2011) indicates that though it has spurred developments in the development of cross-cultural analysis, it has also resulted in a significant amount of work based on "misunderstandings, misrepresentations and misuse of some of its main elements" (2011, p. 11). This criticism is most commonly pointed at the methodological or epistemological aspects of Hofstede's research, about which McSweeney (2002) and Kim (2007) were critical.

The most relevant criticism, which was actively taken into consideration while evaluating possible operationalisation of Hofstede's framework for research include: the use of a single organisation as place of research and claiming that the results are representative of cultures in general, as well as measuring culture through surveys only, the ethno-centric (or Eurocentric) qualities of the formulated dimensions, and the view of culture as a static and nation-based phenomenon without consideration for cultural diversity within nations. The scores for Germany are based solely on West Germany, for instance (Ten Thije & Pinto, 2011). Limiting culture to five binary oppositions, be they dimensions with sliding scales, is also hard to reconcile with the view on culture expressed earlier. However, a quantifiable measure for culture would be a useful tool for contrastive research. Hofstede's model might be too limited in its scope and does not take language into account. Due to the limitations discussed above Hofstede's model will not be included in the analysis model for this paper. However, the interrelation of the core concepts in the dimensional models and the central role Hofstede plays in other research discussed in this thesis has prompted us to provide an overview of his

¹ See Kirkman, Lowe and Gibson (2006) for an overview of research that made use of Hofstede's framework.

dimensional framework. Hofstede has found differences between national cultures, as measured within IBM through questionnaires. The next step is to find something more quantifiable and less interpretable to measure culture with. House (2007) has formulated cultural dimensions based on linguistic evidence, which makes culture a far more tangible concept without constraining it to binary oppositions.

House hypothesises a set of dimensions to function as “a yardstick for explaining language- and culture-conditioned behaviour” (2007, p. 249-264). Having established these dimensions, she proceeds to employ them in order to investigate and link culture clashes to language-specific norms of interaction. The converging patterns that surfaced in analyses are categorised along five dimensions. It is very important to keep in mind that House’s dimensions are a representation of different *preferences* or *tendencies* in communicative behaviour related to, but different from more polarised dichotomies like Hofstede’s score model or Hall’s dimensions (1976; 1966), and with a distinct linguistic focus (2007, p. 251-254). A brief explanation of each of the dimensions relevant to this study can be found below:

Directness vs. Indirectness: The Directness/Indirectness dimension is most clearly illustrated by looking at the performance of and expectations surrounding speech acts. Requests and complaints in particular are a prime example of where differences along this dimension occur. Plainly put, it is the difference between clearly stating intents and wishes, and suggesting them. House illustrates this dimension with examples of customer-employee interactions in stores. The differences along the lines of this dimension between German and English speakers is quite apparent; Anglophone speakers expect a speech act to be indirect, but receiving a more direct response instead. For example, a bare infinitive used by the German speaker, “Gehen Sie da drüben hin” (House, *Communicative styles in English and German*, 2007, p. 255) feels

like a direct order. This strikes a chord with the Anglophone perception of politeness, even though the proper address has been used, namely the polite second person singular “Sie”.

Orientation towards Self vs. Orientation towards Other: Along this dimension the degree of orientation towards the Self or towards the Other is measured. In essence, this manifests itself in the degree of (non-)reciprocation of concern or acknowledgement of the other interlocutor. By not asking after the general well-being of the other through simple statements like ‘How are you’ or ‘Good morning,’ interlocutors are perceived as rude and self-centred.

Orientation towards Content vs. Orientation towards Addressees: This dimension, in essence, measures the degree to which interlocutors focus on the content of the message, rather than the interaction between them. If there is more focus on content, there is less concern for social niceties. Depending on the expectations of the interlocutor, the speaker can be considered polite or rude when less focused on the other person in the interaction.

Explicitness vs. Implicitness: Orientation towards Content/Addressees and Explicit/Implicitness are dimensions that are measured separately, yet House argues they are so intricately related that they can be analysed together. House states that “giving priority to the expression of the content of a message in their communicative behaviour, speakers tend to be more explicit in conveying their message” (p.257). Ergo, a request for someone to “move up a bit” in the bus is made more explicit by adding “there are more people who want to get on the bus” (p.257). Depending on the expectation of the interlocutor, the speaker could be perceived as rude when the message is both explicit and content oriented (p.257).

House sets out to establish the value and explanatory power of her dimensions by utilising them in her analysis of cultural clashes. In order to do this, she compiled a dataset consisting of data from various contexts and subjects, involving observations, interactions and self-reflections across various levels of interaction obtained from textual sources, observation

and audio recordings. House's approach is, by nature of her data and methodology, a qualitative one with subjective elements. House admits that naturalistic data from everyday interactions would be best suited to analyse culture, but while the heterogeneity in her data is not the optimal circumstance for internal consistency, it does illustrate that the dimensions can be successfully be applied to a wide variety of discourse contexts and environments. The context and environment in online interaction, as found in online discussion boards, is one such discourse environment that has not been analysed through House's framework. It does, however, provide a large amount of naturalistic data and should thus provide valuable new insights in the field of contrastive interactional analysis.

In light of the subjective nature of the analyses, House's interpretations of the data regarding the five dimensions were tested by a panel of expert bicultural and monocultural judges. As a result of these tests, the experts supported four out of five dimensions. Only the Self/Other dimension was not fully supported. Instead, elements of this dimension were attributed to the Directness/Indirectness dimension. This illustrates that the dimensions are interrelated, showing overlap that could prove problematic when attempting to quantify them.

House's findings are confirmed by Byrnes (1986) and Agar (1994), with both researchers finding similar anecdotal evidence as House has encountered between American and German speakers. House's dimensions have also shown to function and yield interesting findings when applied to language pairs other than German and English (2007, p. 253). In these projects a number of different language pairs were researched and they were confirmed to be compatible in various contexts, including: face-to-face or telephone talk, interviews and, most interestingly, written discourse (2007, p. 253). It is interesting that House's findings correlate with other researchers' findings, as this opens up new venues of relevance and application of House's dimensions. These additional venues also include similarities between House's dimensions and those presented by Hofstede (2001) and Hall (1976). For instance, Hofstede's

underlying values in the Uncertainty Avoidance dimension relate to House's Implicitness/Explicitness dimension; in order to fulfil the need reduce uncertainty, speakers will express themselves more explicitly (2007, p. 252). Hall's high context cultures rely on non-verbal channels. These non-verbals in interaction are contextualisation cues. This coincides partly with what House calls implicitness, although her examples deal with expectations of speech act patterns and politeness instead of contextualisation cues. The fact that overlap is found between psychological, interpersonal and linguistic dimensions makes for an interesting case regarding this study's attempt to combine these perspectives and utilise them to supplement each other.

House presents the different underlying expectations and preferences related to politeness as one of the prime culturally conditioned undertones that have surfaced as a result of her work. While this study does not analyse the corpus at the speech act level, where politeness-strategies are fully realised, it does take note of linguistic actors that convey politeness at word level. This can be best illustrated by the multimodal model of politeness devised by House, which breaks down the phenomenon of politeness into four levels: biological, philosophical, cultural and linguistic (2007, p. 261). The focus of this study is solely on the fourth level, as this is where socio-cultural phenomena like this manifest themselves on linguistic level, and "it is at [this level] where the hypothesized five dimensions of communicative style are likely to operate" (2007, p. 260-262).

The most important part of House's dimensions is that they suggest "linguistic differences in realising discourse may be taken to reflect deeper differences in cultural preference and expectation patterns" (2007, p. 264). This makes House's dimensions a welcome complementary viewpoint in measuring cultural difference. House's dimensions present us with a set of tools that allow us to gain insight into the dynamics of culture in discursive interaction. This is reminiscent of Agar's (1994) concept of languacultures, where

language and culture are viewed as inseparable concepts. In this sense language is both indicative and product of culture. House's analysis of the speech acts that represent the overarching communication styles are an important window into the complex relationship between language and culture. The next step would be to employ House's dimensions in a quantitative and non-interpretative analysis, rather than qualitative and interpretative.

Along with Hofstede (1980; 2011), Edward T. Hall proposed another one of the more known frameworks that are frequently used in the intercultural field. Hall describes dimensions of cultural difference, but unlike other frameworks Hall is more interested in behavioural difference rather than fundamental values. Hall's dimensions concentrate on three different aspects of cultural behavioural difference: *Monochromatic time versus Polychromatic time*, *High versus Low context communication* and *Use of Personal Space* (1976). The aspect of cultural difference that is discussed in this study pertains to the high context/low context communication. Low-context communication is explicit in meaning and uses words to describe the meaning of the message, while high-context communication draws heavily on context, which includes social backgrounds, shared knowledge and non-verbal communication (Spencer-Oatey & Franklin, 2009, p. 23). Hall explains that while linguistic code can be analysed independently on some levels, meaning and context are so intimately connected to each other that "in real life, the code, context and the meaning can only be seen as different aspects of a single event" and that to measure only one side of the coin is not viable (1976, p. 91). The high context/low context dimension is an important dimension, because the online communication examined in this study lacks the physical nature of the *personal space* dimension, and the concept of Polychromatic-time and Monochromatic-time does not apply as strictly in online environments and on discussion boards in particular. Posts are chronologically

ordered, but neither shun nor adhere to pre-set linear schedules since users are free to enter or leave threads (conversations on a predetermined topic) as they wish.

House and other studies have proven that her dimensions are reliable; but like any single dimensional framework House's dimensions are not all-encompassing, nor perhaps as distinct from each other as they initially appeared to be. However, by using a combination of House and Hall dimensions, a new framework can be compiled that can also be utilised in analysing culture online. The dimensions used in this corpus analysis are *Directness/Indirectness*, *Orientation Content/Person* and *Orientation to Self/Other* from House and *high context/low context* from Hall. The dimensions created by House and Hall are qualitative in nature and their descriptions are interpretative and vague. In order to use them for quantitative corpus analysis this study utilises elements from other related studies to quantify the dimensions. Sections 2.4 and 2.5 highlight the studies from which the elements for analysis were adopted.

2.2 Online Communication

Online interaction differs from real world interaction in some ways and is the same in others. In online environments, a significant portion of the elements that are pivotal to real-life communication are different or no longer available to interactants in their original form. Differences that distinguish online from real-life communication include: a) interaction taking place across various different online media; b) these media functioning through different modes of interaction that c) are essentially limited to textual and visual forms that d) take place across spheres of both public and private interaction which result in varying degrees of anonymity (Androutsopoulos, 2006; 2007; 2013). Additionally, the global nature of the Internet provides individuals with new means to interact with people from around the world; effectively eliminating limitations and diminishing distinctions otherwise created by geographical distance and national culture. This might imply that in online communication national culture is not such a distinguishing factor as it is in real life.

First of all, online interaction takes place using various different media. The size, impact and popularity of these media constantly shift as new platforms develop and distinguish themselves. Unlike face-to-face interaction, there is an important distinction between synchronous and asynchronous modes of interaction on these media (Androutsopoulos, 2006, p. 420). These modes refer to the manifestation of turn-taking in online discourse; these turns take the shape of posts or messages on their respective media. Essentially, posts and messages are the “base-level unit” of online discourse (Androutsopoulos, 2013, p. 495), and it is here that the shape of the interaction is defined. Synchronous modes of interaction are defined by real-time interaction between those involved. This is characterised by speech-like turn-taking: the rapid exchanges of generally brief messages as occurring in real life interaction are also featured in direct messaging and online chat rooms. Asynchronous modes are characterised by posts that are larger in terms of content per turn and typically show variation in the amount of

time between turns – minutes, hours, days and sometimes even weeks – and are more reminiscent of the exchange of letters in real life. Media that utilise the asynchronous mode include discussion boards, email and Social Media. The asynchronous media tend to feature persistent and detailed records of the interaction, making them an excellent source for data-collection. For this reason the corpus used for analysis consists of posts from asynchronous discussion boards discussing *Game of Thrones* and *Bioshock: Infinite*.

It is essential to consider that the Internet, as an interactional medium, is for the most part confined to textual and visual code to express message and context. Auditory information is present in embedded videos but otherwise non-existent on discussion boards. This calls for creative new ways to link contextualisation cues that are taken for granted in conventional face-to-face communication systems, as they cannot be conveyed in the same way. Hall only defines explicit verbal messages as low context, but is more comprehensive when it comes to the high context part of the dimension. The core of the high context/low context dimension is that high context communication relies on factors outside of the explicit meaning of the words in a message. In real life, high context elements translate to body language and other physical and non-verbal contextual cues, such as pauses in interaction, silence and tone of voice, to convey meaning. In online interaction, implicit information must be communicated in other ways. For instance, facial expressions are replaced by emoticons or images. Some of these strategies, such as emoticons and onomatopoeia, will be discussed in more detail later.

Another distinction that can be made is that of the public and private sphere. In online interaction, users are often in control of how much personal information they disclose. Interaction taking place in the private sphere generally involves at least a certain level of disclosed personal information. For instance, social media like Facebook or personal and corporate email communication generally depend on accurate personal information. Interaction in the public sphere, however, does not depend on this and leaves the disclosure of personal

information to the individual. It is important to note that in the public sphere, individuals have very little control over who exactly is monitoring their online exchanges, as there are many individuals, known as *lurkers*, who choose to observe rather than participate in the interaction. This results in interaction in public domains reaching far beyond the interactants themselves. Examples of public sphere platforms include discussion boards and chat rooms that, though they generally require basic registration to participate, are visible to anyone. Users on such public platforms use usernames and nicknames to distinguish themselves from others and build an alternative identity for themselves (Androutsopoulos, 2006; Williams & Copes, 2005; Baym, 1998). The construction and evolution of the identities people construct online are research-worthy in their own right, but in the context of this paper the level of anonymity is most important. Reliance on possibly inaccurate self-description introduces complications if researchers take these descriptions at face value and should thus be either avoided or explicitly investigated. The present research focuses specifically on data collected from discussions boards in the public sphere.

The above factors are ways in which the Internet differs from face-to-face interaction. However, they also contribute to the ongoing discussion surrounding the topic of whether to see the Internet as either a textual medium or a place for interaction. According to Androutsopoulos (2013), defining the Internet as *text* means the Internet is essentially used as a container for written language, whereas defining the Internet as a *place* approaches digital communication as a social process (2013, p. 499-500). However, where Androutsopoulos presents the Internet as place *or* text in a binary opposition, this paper rejects the idea that the Internet is either one thing or another. Seeing the Internet as both a place for interaction as well

as a source of text for analysis is not doublethink, but it echoes the idea stated earlier that language (text) is an aspect and gauge of culture (social interaction).

As a result of the interplay of global reach, public and private space and the influence of anonymity, the real-world distinctions between national culture and other socio-demographic elements are blurred. This is a pivotal element in this research paper, as its goal is to ascertain in how this blurring of distinctions affects the various ways in which online interaction is constructed.

2.3 Culture and Community

Views on culture and their means of analysis are well established in the context of the real world, and even there it is apparent that the essentialist view is too narrow in scope. The necessity of a new approach and a different frame of reference is also felt by Hewling. Hewling researched online interaction in an educational context, analysing online discussion boards as a cultural nexus for the interaction between classmates, tutors and teachers (2006). Hewling challenges the idea that culture is brought into the online classroom through the backgrounds of the interactants' offline identities. Instead, Hewling introduces the concept of *Third Culture*; the idea that the interaction between all the varying frames of reference will result in the emergence of a *Third Culture*. The core of this idea is that "a shared context of understanding in which the members of the group will be both participants and co-owners" emerges in online interaction (Hewling, 2006, p. 340). However, the term third culture as a whole does pose a problem; the idea is intended to step away from an essentialist framework, but the term could imply that the emerging third culture then also exists as a completed, static, and thus essentialist product (2006, p. 341). Hewling attempts to address this issue by stressing that the online classroom is an "evolving site of cultural creation" (2006, p. 343), acknowledging culture as an ever-evolving concept. The online classroom is only one of countless contexts in which online intercultural communication takes place and Hewling's concept also resonates in other places, such as Wenger's *Communities of Practise*, where people form communities in the workplace based on mutual engagement, shared repertoire and a jointly negotiated enterprise (Wenger, 1998); such as employees working together in small groups within departments or students forming smaller workgroups within the context of a course-group. Hewling, however has succeeded in taking more steps towards defining the concept in a manner that could offer insight beyond any one specific context.

Considering Hewling's interesting notion of *Third Culture* in a broader context, the next question to be answered is where *Third Cultures* could occur outside a controlled, familiar and homogenous environments like the online classroom or Wenger's *Communities of Practice* in the workplace. For this, we look towards the concept of online communities. The concept of online communities is, like culture and community in a less specific sense, notoriously difficult to define and this has made it a contested term. The vast range of definitions can be subdivided into inclusive and exclusive definitions (Androutsopoulos, 2006, p. 423). Inclusive definitions look at communities in a broad sense: "a group of people who interact in a virtual environment" (Preece, Maloney-Krichmar, & Abras, 2003, p. 1023), whereas exclusive definitions include a set of prerequisites that are to be met before a group of users is considered a community. In the following paragraph this study will introduce its own inclusive definition of online community.

Large sections of where people interact in online communication consist of public, anonymous environments where people congregate and interact despite that fact that anonymity prevents inhibits their ability to gather according to national or socio-economic backgrounds. This implies that other factors are important in interaction. Online communities in the public, anonymous sphere are best described as a "constantly evolving *collective intelligence* or *collective consciousness*" (Macfayden, Roche, & Doff, 2004, p. 34) that manifest in environments actualised by online technology. They revolve around "regular interaction around a shared interest or purpose" (Androutsopoulos, 2006, p. 422). These communities of mutual affinity, *Affinity-based Communities* (ABCs), are the nexus of online interaction. They are *places* where users gather to seek other like-minded individuals who share their affinity for certain topics. These ABCs grow and evolve, developing shared norms and values, humour and social reciprocation, establish social roles, hierarchies and close

friendships (Androutsopoulos, 2006; Baym, 1998; Danet, 1998; Preece, Maloney-Krichmar, & Abras, 2003; Williams & Copes, 2005).

However, an important focal point of this study is the idea that ABCs rely almost exclusively on linguistic code available to them to construct the above features. As shown by Agar (1994) and House (2007), language and culture are two sides of the same coin. When text is the core of interaction, like on discussion boards, language use and linguistic features become the core of the interaction. Unlike the distinction presented by Androutsopoulos (2006), the internet forum is both the carrier for text and a place for interaction. The language used becomes a cultural construct, the posts and messages that carry it a cultural artefact (Williams & Copes, 2005, p. 73). In this sense, linguistic features do not simply constitute their own framework, but should be seen as resources that users or communities might “draw on in the construction of discourse styles in particular contexts” (Androutsopoulos, 2006, p. 421-422); this implies that linguistic features are indicative of ABCs and the context they operate in. In other words, a difference in language use could be indicative of a difference in culture, which could be brought to light through contrastive interactional analysis. This could express itself in the use of particular jargon or a preference for certain language features like action verbs, personal pronouns, etc.

Since many online communities have members from many different parts of the world, they involve speakers of many different languages, not all of which are native speakers of English. Despite this, English often serves as the common language used in online interaction, playing an important role as the dominant *lingua franca* (Androutsopoulos, 2006, p. 428-429). It is, however, under threat by other super central languages, such as Spanish (Androutsopoulos, 2006, p. 430). This makes the notion of code-switching, here defined as using more than one language within the same post or message (Androutsopoulos, 2007; 2013), an interesting one, as even in communities that thrive on languages other than English, the

dominance of English as an integral part of the topics discussed could certainly play a role. However, no matter the language used, the constraints of textual and visual elements have interesting implications for the way people adjust their contextualisation strategies (e.g. emotion, emphasis, body language, etc.) to online environments. The constraints of online environments have resulted in creative ways for non-verbal contextualisation to persist in a textual environment. Emoticons mimic facial expressions, laughter acronyms and onomatopoeia are used to visualise sounds along with a myriad of other written signs that substitute the contextualisation cues in face-to-face communication, essentially introducing Hall's high and low context dimensions to a textual environment through linguistic elements (Androutsopoulos, 2006; 2007; Georgakopoulou, 2003; Danet, 1998; Crystal, 2001).

2.4 Linguistic Features and Cultural Dimensions

This paper attempts to establish a new approach to analysing language and cultural dynamics in online environments; for this purpose it is essential to build on previous research on language and communication on the Internet. The field of Computer-mediated Communication (CMC) has been growing steadily in recent years, offering valuable insights regarding language and culture online. CMC is defined as any communication between humans through the use of one or more electronic devices (McQuail, 2005). Much of the intercultural research performed under the wings of CMC has focused on the unilateral way the Internet translates cultural elements and preferences in the graphical and organisational design of websites. The Internet could be perceived as the apex of global intercultural interaction, yet the linguistic implications of this have largely gone unexplored from an intercultural perspective within CMC.

Despite the lack of focus on language in the context of intercultural communication, research has looked at and considered the language produced by the Internet. Androutsopoulos indicates that initially sweeping, homogenous categorisations like *netspeak* and e-mail language were popular (Androutsopoulos, 2006; Georgakopoulou, 2003). These generalizations represent a medium-oriented approach rather than a user-oriented approach. A medium-oriented approach assumes medium x uses communication feature set y (Androutsopoulos, 2006, p. 420-422). Androutsopoulos aptly notes that the vast diversity of settings and contextual purposes of language use outweighs such a narrow deductive perspective. This is reminiscent of the dichotomy of culture as either a set of pre-set values versus culture as a dynamic social process. Androutsopoulos points to the lack of attention for the “socially situated discourse in which these features are embedded” that exists in most CMC research to date (2006, p. 420), a gap the present research seeks to fill.

Linguistic features that have been quantified and compared in CMC studies include unconventional spellings, representations of spoken language features, emoticons and code-

switching; these studies have found that features can be “patterned by age, gender and location” (Androutsopoulos, 2006, p. 425). For instance, emoticons and laughter have been found to be used more frequently by women in Witmer and Katzman (1997). Similarly, Herring attributes language use and discourse style in online blogs to age and gender; finding that men write more and longer messages and use less politeness strategies in contrast to women. Women, in turn, tend to write shorter messages, express feelings and laughter as well as use emoticons more frequently (2003). However, the findings by Witmer-Katzman and Herring cannot be fully explained by age or gender alone, as Huffaker and Calvert conclude that the use of emoticons was found to be used more frequently by teenage males and that “[language and discourse styles on] blogs operated by young males and females are more alike than different” (2005, p. 19). This implies that there are other factors that are of importance, which is further supported by Androutsopoulos who states that Herring’s findings suggest that these linguistic features depend on the genre of weblog writing rather than gender (2006, p. 428). This is similar to this paper’s perspective that language use is linked to affinity for a topic or genre. This means that differences, or more importantly, the lack thereof are indicative of the cultural context in which the interaction is embedded.

Unlike the dimensions of Hall (1976) and Hofstede (2001), the dimensions created by House (2007) have, at time of writing, not yet been employed to analyse online interaction. However, as online interaction is mostly textual, this paper looks to the fields of translation studies to gain insight into which lexical elements can be used to operationalise House’s dimensions. The viability of House’s dimensions for use in contrastive cultural analysis of written text is exemplified by Pinto and ten Thije (2011). Ten Thije & Pinto (2011) analyse speech acts from a functional pragmatic perspective by focussing on illocutionary acts. As a result of their comparison of Dutch and Italian IKEA catalogues two relevant differences surfaced when analysing the same context, namely a much higher frequency of action verbs in

the Dutch catalogue juxtaposed by descriptions in the Italian catalogue; and a notable increase in the frequency with which the Dutch catalogue directly addresses the reader (Thije & Pinto, 2011). While couched in different terms, these differences could reflect aspects from the dimensional models of House and Hall. The increased number of action verbs in the descriptions of rooms in the Dutch catalogue could, for example, be indicative of the *orientation to content/person* dimension in the sense that the action verbs increase the orientation towards the *content* of the message by adding additional context through the expression of specific actions. This, in turn, makes the message more *explicit* rather than *implicit*, which relates to the low context dimension (Hall, 1976).

Similarly, ten Thije & Tempel (2012) also find cultural differences along House's dimensions when analysing the texts of Dutch, English and German audio tours. The results of this analysis indicated that most linguistic differences between the original texts were found in the *Orientation to Content/Person* and *High context/Low context* dimensions. For instance, the Dutch and German texts were more oriented to content by focusing on the paintings rather than the audience. The reported lack of cases in the *Orientation towards Self/Other* dimension that this thesis also includes in its analysis can be explained by the non-interactive nature of the audio tour texts (2012, p. 651). The adaptations made to the new texts produced by the researchers were tested on the museum audience, but showed that the cultural adaptations did not result in increased appreciation of the text compared to the original versions.

Like House, Ogiermann (2009) also analyses cross-cultural speech act realisation. Ogiermann analyses English, German, Polish and Russian requests with discourse completion tests. Ogiermann's most notable finding concerns the use of syntactic and lexical downgraders. Lexical and Syntactic downgraders are employed as a hedging strategy, modifying the illocutionary impact of the request, essentially reducing the threat to the interlocutor's face (2009). Ogiermann has found that lexical and syntactic downgraders are indicative of culture-

specific preferences regarding politeness. For this study, this is an essential addition to House, as Ogiemann's downgraders provide us with quantifiable elements to include in the hypothesised framework.

Nickerson (2000) investigates language strategies in corporate settings by analysing email interaction between Dutch and English native speakers. Nickerson employs quantitative analysis based on a wide variety of language features that relate to the underlying concepts of the cultural dimensions by Hall and House. The most relevant of these features focus, like the present research, on (interpersonal) communication styles. Some of the features analysed by Nickerson are also analysed by other researchers mentioned above, where they were found to be indicative of differences in communication style. These include: *contractions and abbreviations, first person pronouns, hedging, downgraders and politeness strategies*. Their successful employment in quantitative corpus analysis further solidifies their selection for quantification of the dimensions in this study. Nickerson's findings, unlike most other studies, indicate that there are more similarities than differences between Dutch and English communication styles in a corporate setting. For instance, similarities are found between the frequency of *first person pronouns, hedging and politeness strategies*. These findings lend further credibility to the concepts of *Third Culture* and *Communities of Practise* identified by Hewling and Wenger respectively, and by extension support the concept of *Affinity-Based Communities* because "a typified corporate discourse may exist regardless of the national culture of individual employees" (p. 176). However, a possible mark against Nickerson's data is that it was gathered from a different context, it involved data from a single company where both language groups consist of native speakers. This paper's dataset is gathered from a heterogeneous environment involving both native and non-native speakers. As the discussion boards are anonymous, they also leave us without a clear distinction between the two. The fact that Nickerson analysed e-mail communication could also account for different findings, as

Androutsopoulos (2006; 2007; 2013) indicates, private sphere corporate e-mail communication is very different from public sphere discussion board interaction and different preferences may not be out of place.

Nickerson is of great value to this study due to its successful employment of quantitative corpus analysis to research communication styles, as well as its bridging and underlining of key concepts and interactional media in this paper. These studies across various disciplines indicate that House and Hall's cultural dimensions can indeed be employed in tandem with a wide array of methodologies to look at a wide variety of subjects and linguistic phenomena.

2.5 Cultural Dimensions Online

Though this paper analyses culture in online communication and employs House's dimensions in a manner that has not been attempted before, other frameworks of cultural dimensions have already been employed for online cultural research. Despite the difficulty of marrying Hofstede's dimensions to current definitions of culture, Hofstede's work has been an essential milestone in attempts to grasp otherwise unquantifiable cultural concepts and has been the basis for many researchers related to this paper's field of research. Hall's model of cultural categories has a following in communications and marketing research, and has proven useful in those fields (Hermeking, 2006, p. 199). A number of studies from these fields that proved relevant to the quantification of the dimensions used in the present analysis will now be discussed.

Singh et al. show that websites can be differentiated quantitatively and reliably based on Hofstede's framework (2006). Singh et al. also include Hall's high context/low context dimension in their analysis; though in doing so, they include three parameters as indicators of high context cultures: indirectness/politeness, soft sell approach and esthetics. However, the present study employs House's dimensions to look at indirectness in more detail while reserving Hall's high context dimension for what Singh Zhao and Hu refer to as *Esthetics*, which includes colours, images, and contextualisation cues (2006, p. 146). Moreover, both politeness and a soft sell approach are considered indirect in this study.

Research in various fields within Computer-Mediated Communication (CMC) has also indicated that both Hall and Hofstede can be used to analyse culture on the Internet. Hermeking (2006) has created a framework based on Hall's dimension of high and low context communication and proceeds to use it in an in-depth analysis of randomly selected brand websites from Europe, Japan and the US. Hermeking found that websites advertising durable products are more adapted to High context preferences or Low context preferences than those advertising industrial goods (2006). This indicates that there are differences between the styles

of particular industries, regardless of national culture. However, the fact that Hermeking found differences between US, Japanese and European websites on the one hand is significant because it does indicate that there are also differences based on national culture. On the other hand, the small size of the corpus analysed can be a mark against Hermeking's findings, even though they are based on content analysis.

Callahan similarly looks at the influence of national cultures on websites, but focuses on organisation and visual design on university websites (2006). Callahan shows that the findings from the websites do correlate to a certain degree with Hofstede's dimensions and his index values per country, although these are statistically weak (2006, p. 269). Moreover, the characteristics of national culture are not the only source of variation and other factors can have influenced design decisions, such as genre, the available technology and institutional guidelines (Callahan, 2006, p. 270). What is most interesting in Callahan's study is that she has identified not only images and what they portray as culturally relevant, but image frequency as well. Image frequency could then be considered a part of the high context dimension as high context communication relies on factors outside of the explicit meaning of the words in a message.

Snelders, Morel and Havermans looked at the relationship between industry style and local culture in particular and also found that both industry style and local culture influence web design, although industry style seemed to be more important than cultural adaptation. This correlates with our own hypothesis, where the overarching interest or topic is more important in the communication style than local cultural backgrounds and adaptation (2011). Snelders et al. identified cultural markers and used them to culturally adapt websites; these cultural markers were mostly visual: fonts, shapes, icons, colour and use of animation (2011, p. 466).

Würz, in her research on Macdonald's websites, turns to Hall's High and low context dimension to analyse and compare high and low context communication strategies (2006).

After first categorising her target nations along on a scale of high and low context cultures, Würz then analyses them according to the relative placement along the scale. Würz concludes that high context websites employ five strategies in particular to accommodate High Context values in online environments; the most interesting of these findings is the use of animation (2006). A mark against Würz is that her research dates from 2006, a cursory glance at the websites she analysed clearly shows they are very different after almost a decade later.

The above studies all use images and other visual elements to look at national culture and explain the above differences by referring to politeness and contextual information. This indicates that visual elements are an important addition to the information provided by textual elements. For this reason, this study includes visual elements like images, video and animation in the analysis as a part of the high context dimension.

2.6 Cultural Dimensions Framework

The successful adaptations of Hofstede (2001), House (2007) and Hall (1976) show that differences in culture are indeed measurable for websites, though these differences are hard to measure using any one single model. To remedy this, the present study will introduce a new framework of dimensions based on the models of House (2007) and Hall (1976) by utilising findings from the studies outlined above to attempt to operationalise the dimensions. As such, the proposed framework of dimensions in this paper is a composite of different theories from the fields of translation, discourse analysis, marketing research, intercultural communication and computer mediated communication; it focuses on both visual and linguistic communication in order to compare the communication styles of different internet forums discussing the same topic. This framework is based on the following dimensions, as discussed earlier:

- *Directness/indirectness*
- *Orientation to content/person*
- *Orientation to self/other*
- *High context/low context*
- *Idiomatic Context*

The first three dimensions are originally taken from House (1996; 2007), who utilises them in her contrastive analysis of English and German language use. The fourth dimension, *high context/low context*, is based on the original dimension as described by Hall (1966; 1976), which is also very similar to House's *Implicitness/Explicitness* dimension.

The last dimension emerged from the data in this study and ties in with the concept of online communities, as discussed by Androutsopoulos (2006) and as a part of the emerging definition of ABCs. The *Idiomatic Context* dimension is constructed from two types of occurrences found in the data. The first is a form of online slang that includes words like *LOL*, and the other component is words that are part of the jargon used in a specific ABC. Both

jargon and abbreviations are indicative of a community-specific idiom, which is why this dimension was coined Idiomatic Context (Androutsopoulos, 2006).

The model of analysis is adapted from these (House, 2007; Hall, 1976) sources, as well as other later studies (ten Thije & Pinto 2011; ten Thije & Tempel, 2012; Ogiermann, 2009, Snelders et al., 2011; Würtz, 2006; Callahan, 2006; Hermeking, 2006) that use the same dimensions or base their model of analysis on these concepts. Generally speaking, these dimensions were measured by analysing speech acts, as exemplified by House (2007), ten Thije and Pinto (2011) and Ogiermann (2009). This study, however, relies on particular lexical and grammatical occurrences and their frequencies; instead of analysing linguistic structures as they relate to the intentions of the speakers, words were counted. Some of the lexical and grammatical occurrences that were counted include action verbs (Ten Thije & Pinto, 2011), downgraders (Ogiermann, 2009) and images (Callahan, 2006; Würtz, 2006). This allows for a quantitative analysis and exploration of the concepts embodied by the dimensions. The exact quantification of each dimension will be described in the next chapter.

As discussed above, much of the interaction in online environments takes place in public, anonymous spheres. This challenges traditional concepts of (online) community as revolving around age, gender, socio-economic background as dominant factors severely reduces the influence of aspects pivotal to national culture. This means that people gather to interact based on reasons not necessarily related to socio-economic factors like age, gender and geographical location. The emerging definition of *Affinity-based Community* (ABC) shows that people interact online by gathering in communities; where they do so primarily according to *affinity*. They then interact and shape their ABC through the *language* that they use. Emphasis is placed on the idea that language and culture are inseparable concepts and that a difference in language use represents a difference in culture. The assumption that ABCs talk about the same topic in the same way regardless of language and national culture forms the basis for the

hypothesis. Based on the literature discussed above, it could be said that if national culture does not play a central role in online Affinity-based Communities, it follows that:

H1: There is no difference in language use between Dutch and English groups in online interaction when discussing the same topic on discussion boards.

The next chapter outlines method of the study, it contains the exact definitions of the lexical and grammatical elements counted in the dimensional framework, as well as a description of the data collection and analysis.

3. Method

This study was conducted as an intercultural, quantitative analysis of Internet forum posts. The analysis uses a framework of dimensions, relying on the links between language and culture, for contrastive analysis of Dutch and English forum posts. The dimensions used are modified from the dimensions mentioned in the theoretical framework. The focus is on language as an indicator of culture. The dimensions are used as indicators of culture in ABCs on a linguistic level; the analysis is done on lexical and grammatical level, rather than functional and pragmatic. The literature indicates that it is very likely that online communication styles do not differ based on language, but rather on genre or topic, which lead to the following hypothesis:

H1: There is no difference in language use between Dutch and English groups in online interaction when discussing the same topic on discussion boards.

The data source used is the Internet, and forum posts specifically. The analysis of the data has been adapted to the mixture of characteristics that the corpus exhibits, namely of verbal face-to-face and written discourse online. These characteristics are divided into a framework of dimensions that can be scored and compared. A pilot was conducted to determine whether the linguistic characteristics used were applicable to online forum posts. The topic chosen for the pilot was the card game *Hearthstone*, and had a data set of 50 English and 50 Dutch forum posts taken from two different forums. This resulted in the addition of several items to the model of analysis. The results from the pilot were not used in analysis.

The data collection process is described below, followed by the operationalisation of the dimensions used for analysis, as well as reliability testing of the proposed dimensions. Examples used to illustrate the dimensions and the items counted are from the analysed data. After testing the reliability of the dimensions, the dimensions used for analysis are established and the statistical analysis is described.

3.1 Online Data Collection

The data used for analysis was taken from eight different Internet forums. Internet forums arrange similar topics together; an Internet forum can be categorised into sections, which contain topic groups that have discussion threads inside. For example, a forum on beauty can have sections on hair, makeup and nails. These sections then have threads inside about topics that can include hair products reviews, nail polish pictures and make-up swatches per brand. The discussion threads display the total number of posts inside as well as the most recent post (see figures 1 and 2). All of the forums used in this study are structured as described, only with different topics. All forum users need to be registered to be able to post messages and have a profile page that shows their usernames and what personal information they want to share. Forums have core posters, casual posters and readers, while administrators have total control over the content of the threads (Androutsopoulos, 2013, p. 345). Administrators can delete posts, close and delete threads and ban users. Threads can be active, dead, or locked and archived, depending on whether they are still regularly being posted to and whether the administrators have blocked access to it.

Topic (= Jump to last post)	Last Post	Posts	Views
Official "Can't remember the title of a game" thread 2.0 [1 2 3 4 5 ... 66]	nuges01 53 minutes ago	3,316	233,469
Games Discussion Lounge - New Year Edition! [1 2 3]	celiajd 8 hours ago	122	9,278
What Should Win Game of the Month for March?	WannabeSbb 9 hours ago	20	286
New releases: What Are You Buying/Post your buys [1 2 3 4 5 ... 8]	Warford_Irochi 1 day ago	375	21,878
Games Discussion Board Rules & FAQs	CrimsonBrute 1 year ago	0	11,012
Choose three games (or franchises) and just rant about them.	Gamerno6666 56 seconds ago	8	83
Anyone else getting tired of the same ips?	Lulu_Lulu 6 minutes ago	31	275
Gargate discussion thread (one and only, KEEP IT HERE) [1 2 3 4 5 ... 33]	dakan45 8 minutes ago	1,651	62,951
Zero Escape 3?	srebmunshclub 40 minutes ago	1	18
What Can You Do In A Game (in 60 seconds or less)?	EL_ZoT2I2o 46 minutes ago	8	93
Why are gamers so upset about BLOPS3?	MarcRecon 1 hour ago	4	51

Figure 1 Gamespot Forum Topic overview

#2 Posted by (8340 posts) - 2 years, 20 days ago
I am half and half about the game. I'm relying on the reviews for this one.

#3 Posted by (4479 posts) - 2 years, 20 days ago
While we wait for IGN's exclusive "tough" (paid off) "tough" review here is an interesting article on how Infinite is one of the most expensive games of all time.
<http://www.destructoid.com/bioshock-infinite-among-most-expensive-games-of-all-time-249310.html>
And people wonder why this industry is unstable
Developers, like Rockstar, have bragged in the past that games make more than films or the box office, so why not fund games like Hollywood does? 2K Games is doing just that by throwing nearly \$200 million at BioShock Infinite, according to The New York Times which cited unnamed analysts. Not since Grand Theft Auto IV – allegedly \$100M and produced by 2K's owners Take-Two – has so much money been thrown at a project
Around \$100M was spent on Infinite's production and as much as another \$100 will be spent on marketing. Remember when we were all gobsmacked that Gears of War cost \$10 million? While game budgets have gone up, BioShock Infinite is pushing to the top of the list of most expensive productions. Shameless (\$94M at today's economy) says hi
Seems they are throwing a ton of money into this. They NEED reviews to be high. Estimate is that it needs to sell about 5 million to break even. I think it can do that but it will be close.

#4 Posted by (8340 posts) - 2 years, 20 days ago
"Marketing to cost another \$100"..... wut

#5 Posted by (4479 posts) - 2 years, 20 days ago
"Marketing to cost another \$100"..... wut
-Earsless
Wep. That is nuts.

Figure 2 Gamespot Bioshock Infinite Thread

The forums were selected on the basis of language, activity, topic and post frequency. The languages selected are Dutch and English, since the researchers are fluent in these languages. The forums were selected with forum activity in mind; the forums had to have at

least a few hundred active users and the selected threads also had a minimum of 100 posts, to ensure that a large number of users have contributed to the topic. Each thread in the corpus contains posts from at least 20 different users per thread used for analysis. This minimised the chance that forum threads, despite having many posts, only contained posts by the same users. The average length of the posts in the *Game of Thrones* data set is 48.8 words on the English forums ($SD=136$) and 31.5 words on the Dutch forums ($SD=33$). The average length of posts in the *Bioshock* dataset is 31.1 words on the English forums ($SD=47.9$) and 27.8 words on the Dutch forums with ($SD=28.5$). The standard deviations are quite high, although this is not very surprising since forum posts, by their nature, can be very different in length. Much like in spoken conversations, turns taken can be either short or long, depending on the conversation.

The data sampling was done by theme, choosing one topic to compare between different forums in Dutch and English. Generally, a random sample is considered the best for generalisation. However, this study focuses on whether there is a difference between nation-based cultures when they discuss the same topic. Moreover, there is the matter of anonymity on the web. Social contexts and personal information of the speakers or users are mostly unavailable; this includes age and gender. Retrieving information would take a great amount of time and the methods required would be ethically questionable. Consequently, nothing is known about the forum users apart from their username and the language used on the forum. Out of ethical considerations, the usernames are not included in the examples shown.

Only screen-based data collection took place; the forum posts were collected and analysed. There was no participant observation or any virtual fieldwork. Two data sets are used, with 50 posts from four forums each to improve the validity of the results. This resulted in 200 posts per data set, with 100 posts in each language. Two forums were selected for each language to rule out the possibility of encountering differences based on the specific forums that were selected rather than differences in language or affinity-based community. This

decreased the chance that differences between groups could be attributed to the forum itself. To improve the validity of the results, the analysis consists of two second data sets, each with a different topic.

The Internet addresses for the forums and threads can be found in appendix 1. The first set of data has the topic of *Game of Thrones*, season 3. This topic was chosen because it is a popular television show, which results in many posts as a source for analysis. It also has widespread popularity globally, meaning that there are many forums in different languages that have threads on the topic. The same reasons apply to *Bioshock: Infinite* as a topic for the second data set; the popularity of the game makes for a plentiful and widespread source for analysis. An additional reason for the choice of topics was that the researchers involved are familiar with the topics on the forums, which gave them insider knowledge about the subjects involved.

The data collection took place in the spring of 2014. The Dutch and English forums used for the first data set can be found below. The tables indicate which 50 posts were used for analysis and which threads were selected on the forums. The forum threads have been labelled according to topic and language; the first English-language *Game of Thrones* forum is labelled GOTEN1, the first Dutch one GOTNL1, etc. The English *Game of Thrones* threads are on *Movieforums* and *Fanforum*. The Dutch *Game of Thrones* threads are on *Moviemeter* and *Filmtotaal*.

Table 1
Game of Thrones Forums Table

	GOTEN1	GOTEN2	GOTNL1	GOTNL1
Forum	Movieforums.com	Fanforum.com	Moviemeter.nl	Filmtotaal.nl
Thread	Game of Thrones Season 3	Season 3 Discussion #1	Game of Thrones	A Game of Thrones
Post 1	04-01-2013, 09:22	02-01-2013, 13:16	05-01-2013, 22:13	15-01-2013, 14:39
Post 50	04-08-2013, 20:13	02-08-2013, 07:25	27-02-2013, 11:24	07-05-2013, 16:15

Note. The posts in either of the forums are not numbered, but each one has a timestamp; these timestamps were used to identify the posts used in analysis. The timestamps display the following information: day, month, year and time of posting.

Movieforums is a website that includes essays and reviews of movies and television shows; general movie discussions, movie reviews, upcoming movies and movie questions make up the bulk of the forum. *Fanforum* has a few thousand users online on average and includes topics on celebrities, music artists, televisions shows and fan extras. Both forums are very active and contain many posts on the topic in a short time-span. *Moviemeter* and *Filmtotaal* are both websites about films and film reviews that also include forums for members. As can be seen in table 1, both of the Dutch *Game of Thrones* threads were active, with a fair number of posts posted within a short timeframe. The forums for the second data set, with *Bioshock: Infinite* as topic, are presented next in table 2. The English forums are *Gamespot* and *Irrational Games*. The Dutch forums are *Gamer* and *XBW* (formerly xboxworld).

Table 2
Bioshock Forums Table

	BIOEN1	BIOEN2	BIONL1	BIONL2
Forum	Gamespot.com	Irrational Games Forums (2k.com)	Gamer.nl	XBW.nl
Thread	Bioshock Infinite Official Thread. One of the highest rated games EVER.	Bioshock Infinite Discussion (Spoiler?)	[Multi] Bioshock Infinite	Bioshock: Infinite
Post 1	04-2013, post #1	08-12-2010, #1	12-08-2010, 21:48	12-08-2010, #1
Post 50	04-2013, post #50	08-12-2010, #50	02-07-2011, 12:41	14-08-2010, #50

Note. Forum posts used for analysis can be identified via post number # or timestamps ordered day, month, year and time of posting. Day and time are not displayed on Gamespot posts.

Gamespot is one of the largest game review websites, which also includes a forum where member can discuss topics generally related to games. The thread was created after the release of the game in 2013, rather than the announcement of the game in 2010. All the other forum threads were created in 2010, after the announcement of *Bioshock: Infinite* in August 2010. *Irrational Games* is the producer of the game *Bioshock: Infinite* and it is the forum on the company website that was used for analysis. *Gamer.nl* is one of the largest Dutch games news websites with a forum for members. Although many users contributed within the dataset, the 50 posts from *Gamer.nl* were posted over a longer period of time, indicating less activity early on. The thread itself still contained hundreds of posts; therefore the lack of posts in the early period did not result in excluding the forum from the analysis. *XBW* is a game news website dedicated to Xbox games and is generally very active.

3.2 Framework for Cultural Analysis

The framework of cultural dimensions devised for this study was outlined in the theory. It contains five dimensions, originally from House (1996; 2007) and Hall (1996; 2007), as well as a final dimension that was put together from the data itself. The dimensions are: *directness/indirectness*, *orientations to content/person*, *orientation to self/other*, *high context/low context* and *Idiomatic Context*. Though these dimensions are, in the literature, inherently qualitative in nature, this study attempts to operationalise these dimensions for use in a quantitative analysis. Nickerson (2000) has used a similar strategy, which proved successful, although this paper is smaller in scope and concerns itself with online communication and ABCs. The most important aspect in operationalising the dimensions was determining that they were indeed quantifiable and not interpretable. Each dimension is listed below, along with the corresponding definitions of the lexical and grammatical phenomena that were scored.

3.2.1 Directness/indirectness

For this dimension, to be able to quantify the directness or indirectness, three lexical occurrences were counted: imperatives, passives and politeness routines. Examples from the Dutch and English data respectively are:

Denk eraan het eerste boek kwam uit 1996

Keep your eyes out.

There is no difference between Dutch and English imperatives and they require no additional analysis. When it comes to passives, both Dutch and English use a periphrastic passive voice rather than verb inflection. These sentences from the data exemplify the passives found in the data:

Het schip van Littlefinger was al vaker in eerdere afleveringen te zien.

They're described as tall and gaunt with pale white skin and blue eyes, and they wear reflective armor.

Politeness routines are words and constructions used to be polite or mitigate utterances. Nickerson (2000) also used hedging, downgraders and politeness strategies in her analysis of corporate e-mail, although we count hedging and downgraders as politeness routines. In Dutch, the personal pronoun *u* is the polite form of addressing someone else. There is no specific form of address in English that is more polite than the second person singular *you*. General apologies (e.g. *Sorry*) in both Dutch and English are counted as a politeness routine, The Dutch pronoun *u* is as well. The following examples are from the analysed Dutch and English posts respectively:

Sorry 🙄

**puts on flame-shield* (sorry guys* 😞)

Other forms of apology like *excuse me* were also counted in both Dutch and English posts. The last type of politeness routine is hedging, a grammatical or lexical means of decreasing the impact or certainty of a statement. A grammatical means of hedging, by using a syntactic downgrader (Ogiermann, 2009, p. 199), is shown in the next example:

*Before the explosion of threads I thought I **would** start a discussion thread about the new Bioshock announced moments ago.*

Hedging is accomplished by inserting a modal verb after the regular verb. Modal verbs in this configuration are part of politeness routines and were counted as such in analysis. This form of hedging is present in both the Dutch and English language, although less common in Dutch. Other ways of syntactic downgrading are through tense and negation, although these are far

less frequently found in either Dutch or English. A lexical downgrader (Ogiermann, 2009, p. 202) as a form of hedging is found in the next Dutch sentence:

Misschien dat de boekenlezers hier een definitieve uitspraak over kunnen doen.

Lexical downgrading is generally achieved by inserting a word or a phrase into the head act of a sentence to mitigate the impact. In the example above, the word “misschien” softens the impact of the sentence. Other lexical downgraders would be words like “please” and phrasal downgraders such as “I believe.” These lexical downgraders, especially adverbs like “maybe” and “perhaps,” are found in Dutch as well as English.

3.2.2 Orientation to Content/person

The *orientation to content/person* dimension is also part of House’s (2007) framework of dimensions. This dimension is based on whether sentences focus on the message or content, rather than the person involved in the interaction. However, instead of analysing speech acts or intentions, the frequency of particular lexical categories was measured. These categories are: personal pronouns, demonstrative pronouns and action verbs. Personal pronouns are mostly the same for Dutch and English, although Dutch personal pronouns also change based on social distance. Both demonstrative pronouns and demonstrative determiners have been counted in the analysis. Demonstratives counted as orientation to content on the scale of the dimension. Action verbs are words that express a specific action (ten Thije & Pinto, 2011). Examples of action verbs are game, play, talk, etc. Modal verbs and auxiliary verbs are not action verbs, since they do not carry a specific meaning.

3.2.3 Orientation to self/other

Another dimension originally taken from House (2007) is *orientation to self/other*. The dimension encompasses the degree of acknowledgement of the other interlocutor. Instead of analysing speech acts that either directly or indirectly refer to the other interlocutor, personal

pronouns are tallied. Because the source material for analysis is from Internet forums, a few other categories were also added to this dimension. First-person and second-person pronouns were included in analysis as well as usernames and implied first person pronouns. Only first-person singular pronouns were scored, since they exclusively refer to the speaker, whereas first-person plural pronouns refer to a group that includes the speaker. During the pilot, a surprising phenomenon was observed; the subject of a sentence was frequently omitted. This circumstance generally manifested itself when the subject was the first-person singular pronoun. An example of this occurrence is shown below:

***Don't** mind that the dragons seem a bit bigger than **I** remember them being in book 3, either.*

The verb seems to express the imperative mood, but the first-person pronoun in the subordinate clause suggests that there is a first person subject missing in the main clause. Since these missing first-person subjects were relatively frequent, they were included in the analysis. All second-person pronouns were scored as well, which effectively meant all singular, plural, subject and object forms of *you* were counted. The use of nicknames or usernames is standard on Internet forums, giving users an alternate identity to use online. There is some occurrence of speakers referring to themselves by their username, as evidenced by the pilot. This is essentially a reference to self and counted as such in the study. Users tend to refer to others by a personal pronoun or a username, since they do not have access to the personal information of their interlocutors. A username or nickname, as discussed earlier, is always displayed next to a message on a forum. It is the online equivalent of using nametags. This is demonstrated in the next examples:

*If it chaps **hapax** ' hide, tough tittays.*

*Thanks to **Neogaf** for the following:*

All of the speakers in the examples above refer to the other interlocutor by username or nickname. This reference to others via their usernames generally occurs when speakers react to a certain post specifically, rather than the previous post in the thread.

3.2.4 High context/low context

The high context/low context dimension is originally taken from Hall (1966; 1989) and modified by House's hypothesised explicitness/implicitness dimension (2007). High context communication relies on context outside of the actual message. Online communication excludes contextualisation cues that are found in face-to-face interaction, but these are replaced by other elements; images, animations, emoticons and onomatopoeia are all visual representations of emotions or messages. Fonts and colours are not included in the data analysis, since the forum users almost exclusively use the standard font, font size and colour for their messages.

Images are seen as part of high context cultures since they do not literally depict a message as words do, they suggest a meaning instead. The analysis indicated that images on Internet forums can range from Internet memes to screenshots of the game or show that is being discussed. Emoticons are not included as pictures here, since they have their own category. Animations are an extension of images. Examples of animations found in the data are Youtube videos, reaction GIFs² and animated memes. Animated emoticons are still counted under the emoticons category, instead of animations. Emoticons are, in actuality, either images or animation. However, the phenomenon of emoticons is so pervasive in online communication that they were placed in a separate category. Androutsopoulos (2006) describes several studies that used emoticons as part of online communication analysis; these studies showed different

² GIF is short for Graphics Interchange Format and is a file format that can easily be used to create short animations.

results based on the age, gender and nationality of the users. The pilot also demonstrated regular use of emoticons across all forums. Emoticons are, like images and animations, counted as high context. Lastly, onomatopoeia were counted. Onomatopoeia are another a way of visualising certain contextualisation cues, which include laughter, sighs, groans and other auditory actions. Since these words are visual representations of sounds unrelated to languages, this item is also considered as part of high context communication.

3.2.5 Idiomatic Context

The Idiomatic Context dimension is not part of the set of dimensions that either House (2007) or Hall (1966; 1989) constructed, but it emerged from the data. The Idiomatic Context dimension consists of abbreviations and jargon, both of which were frequently and consistently found in the data sample. Abbreviations or alternate spellings are often used in online communication. A few examples of these are “lol,” “wut,” “r” and “u.” All of these examples are probably familiar to people who communicate online regularly; they stand for “laughing out loud,” “what,” “are” and “you” respectively. Jargon is defined as a characteristic idiom or slang belonging to a group (Androutsopoulos, 2006, p. 422). For the *Bioshock* data set, these words would include “Colombia,” “Rapture,” “Final Fantasy,” “Bioshock,” “demo,” “CGI” and “gameplay.” These words were character names, places, objects or anything that directly refers to the specific topic, as well as gaming terms in general; they can also be found in both the Dutch and English data. For the Game of Thrones database, other words and phrases are considered jargon: “You know nothing,” “Mance,” “wights,” “the Others,” “LF.” The *Game of Thrones* idiom is related to the television series or the *A Song of Ice and Fire* book series on which it is based. Classifying a word or phrase as jargon in this context does require some insider knowledge of the group, or a list of previously agreed on words and terms. On a separate

note, the examples used to illustrate abbreviations and jargon were from both the Dutch and English data sets. The quote below is from a Dutch forum:

*“damn you **Kaj!** 😊”*

Code-switching is consistently found on the Dutch forums, although generally the code-switching occurs on a lexical level. On those occasions, the use of jargon or abbreviations coincides with manifestations of code-switching. Only in instances like the example above does the code-switching not occur alongside jargon or abbreviations. This type of code-switching is very infrequent and is also only found on the Dutch forums. Consequently, code-switching has not been counted as an item for testing.

3.3 Reliability Testing

The dimensions as formulated above are extrapolated from research in different fields and have not been tested on real data. The dimensions need to be internally consistent measures of cultural difference in order to use them for means comparison. The dimensions that are proposed in the framework for analysis are shown in table 3. To test whether the underlying constructs correlate with the theorised model, Crohnbach's alpha (CA) was used.

Table 3
Dimensions Tested with Crohnbach's Alpha

Dimension	Items
<i>Directness/indirectness</i>	Imperatives, passives, politeness routines.
<i>Orientation to content/person</i>	Personal pronouns, demonstrative pronouns, action verbs.
<i>Orientation to self/other</i>	First-person pronouns, second-person pronouns, own username, other username, implied first person pronouns
<i>High context/low context</i>	Images, animations, emoticons, onomatopoeia
<i>Idiomatic Context</i>	Abbreviations, jargon

Note. The items are ordered by the dimensions they represent.

However, some things need to be taken into consideration for reliability testing. Crohnbach's alpha is not robust to non-normal data, so all non-normal data is transformed before calculating the coefficient alpha (Sheng & Sheng, 2012). Normality is assessed through histograms and the Shapiro-Wilk test ($p > 0.5$). Since all items displayed a large amount of zero values in the forum posts analysed, which resulted in a very irregular testing distribution, the forum posts with no occurrences of an item were removed from the analysis of that item in order to improve statistical analysis³. Depending on the item used for analysis, between 30 and

³ The zero values across the data sets did not differ between the Dutch and English groups based on t-tests comparing the frequency and distribution of zero values.

98 percent of the forum posts did not contain any occurrences of the item. After removal all items were transformed by using log10 in order to get a normal distribution.

The sample size of the corpus analysis is 200 forum posts per data set; however, the removal of zero values has decreased the sample size of both data sets. The amount of forum posts used for analysis is referred to as N , which are only forum posts that contain occurrences of the item that is being analysed. Outliers were checked for via box plots and removed only if they severely affected normality. The variables were tested for an alpha of 0.68 or higher to see if they are measuring the same underlying construct. Aside from the coefficient alpha, the squared multiple correlation r^2 was determined to see if the different variables are measuring the same single construct, rather than multiple constructs⁴. The r^2 has values between 0 and 1, with a value of 0.5 meaning that 50% of the item values are explained by the other items in the sum. If the Crohnbach's alpha exceeded 0.68 and r^2 was around 0.4 or higher (at least 40% of item variance is explained by the other items), the variables are considered as consistently measuring the same construct and compiled into a new variable.

Table 4 shows the CA values for different combinations of items, the number of cases used (N) and how highly correlated the items were, this information was used to determine whether the dimensions can be used for the data set.

Table 4
Crohnbach's Alpha Table for Dimensions

Dimension	<i>Game of Thrones</i>			<i>Bioshock</i>		
	CA	N	r^2	CA	N	r^2
Directness/Indirectness (Politeness, passives, imperatives)	-	0	-	-	1	-
Orientation to self/other	-	0	-	-	0	-

⁴ The squared multiple correlation r^2 is a value that indicates to what degree one item explains (correlates) the other items in the same dimension. The Crohnbach's alpha does not differentiate between one or multiple underlying constructs, so the r^2 value is as a control for the alpha.

(Second person, First person, Implied first person, Other username)						
Orientation to content/person (Demonstrative pronouns, action verbs, personal pronouns)	0.264	28	0.306	0.596	44	0.351
High context/Low context (Images, Animations, Emoticons, Onomatopoeia)	-	0	-	-	0	-

Note. N is the total number of forum posts used for measuring internal consistency between items. CA is Crohnbach’s alpha and r^2 is the squared multiple correlation. The items are log10 transformed. The following items were reverse coded in order for them to measure the same dimension as the other items: *imperatives, second person, other username.*

The dimensions as extrapolated from the theory and as used in the model for analysis do not have enough cases for comparison for the dimensions *directness/indirectness, orientation to self/other* and *high context/low context* in both data sets. Looking at the column CA, for the dimensions *directness/indirectness, orientation to self/other* and *high context/low context* there are no cases for comparison and there is no proof of internal consistency; these dimensions cannot be used in further analysis. The dimension *orientation to content/person* has enough cases for comparison in both data sets ($N=28$ and $N=44$), but with a CA of 0.264 or 0.596, neither the *Game of Thrones* or the *Bioshock* data set have a sufficient correlation in the dimension. The CA and squared multiple correlation are both too low for the item to reliably measure the same underlying construct and explain item variance. This means that none of the dimensions in the model of analysis are valid for this data set.

Although none of the dimensions as outlined originally in the method are applicable to either data set, table 5 below shows some combinations of items that do appear to measure the same construct. Instead of using all categories from the different dimensions, sets of two items from the original dimensions are used to see if these combinations do have enough valid cases to identify internal consistency and if they can be compiled into one new item.

Table 5
Crohnbach's Alpha Table for Item Combinations

Items	<i>Game of Thrones</i>			<i>Bioshock</i>		
	CA	N	r ²	CA	N	r ²
Politeness, passives	1.00*	4	-	-	-	-
Emoticons, Onomatopoeia	0.902*	11	0.705	0.975*	13	0.904
Second person, Other username	0.400	5	0.981	-	-	-
Demonstrative pronouns, Action verbs	0.653	38	0.235	0.830*	54	0.504
Jargon, Abbreviations	0.586	33	0.211	0.808*	40	0.459

Note. N is the total number of forum posts used for measuring internal consistency between items. CA is Crohnbach's alpha and r² is the squared multiple correlation. The items are log10 transformed. *Crohnbach's alpha higher than 0.68.

The coefficient alpha for the items in table 5 are higher than in table 4, and the corresponding N is also larger, making the CA calculation more reliable in measuring how internally consistent items are. Only the alphas for *politeness/passives* and *emoticons/onomatopoeia* are high enough for them to be reliably measuring the same underlying construct in the *Game of Thrones* data set; both the CAs are higher than 0.68, as can be seen in the first column of the table. The CA for the items *politeness* and *passives* is 1.00⁵ and only one item away from the same combination as proposed in the *(in)directness* dimension. No r² can be calculated since group size is too small. The items *politeness* and *passives* do appear to measure the same underlying construct and those items are part of the *indirectness/directness* dimension, they are compiled into a new dimension called *indirectness* and used for further analysis. Emoticons and onomatopoeia are visual representations of feelings or actions and both belong to the *high context/low context* dimension according to the theorised model. A CA of 0.902 is very high and points to a very high internal consistency

⁵ Theoretically, a CA of 1.00 means there is perfect internal consistency and the items used are definitely measuring the same construct. However, a CA of 1.00 is very rare, although the high value can be explained by the small number of cases, which makes it possible that all 4 cases correlate perfectly.

between the variables. The squared multiple correlation value of 0.705 indicates that the tested items explain roughly 71% of the variance in the data sample. Both of these values suggest that it is very likely that these items are measuring the same underlying construct. The other CA values indicate poor internal consistency between items, meaning that these are kept separate in the t-test. However, it should be noted that the second person/other username combination (or *orientation to other* as a dimension) has a very high r^2 ; 98% of item variance is explained by the items themselves. Because the r^2 is high, it is likely that a larger N would have improved the CA value, which means a larger data set might have validated this dimension.

In the *Bioshock* data set, the CA of *emoticons/onomatopoeia* is particularly high, with a matching high r^2 , meaning that they are almost certainly measuring the same construct. The combination of *demonstrative pronouns* and *actions verbs* also results in a high CA, while the r^2 is moderate; it is still likely that the same underlying construct is measured. The same is true for *jargon* and *abbreviations*; they also have a high CA and a moderate r^2 . The r^2 values for these last two combinations indicate that about 50% of the variance in the cases is explained by the items themselves. Since the CA for all the above combinations in table 11 is higher than 0.68 and the r^2 is at least moderate, they are compiled into new variables. The CA for *passives/politeness* cannot be calculated, since there are not enough cases. This is fairly remarkable, since *passives* have a group size (N) of 27, although *politeness* only has a total group size of 7. A larger data sample might have confirmed the combination as a dimension for indirectness.

For the *Game of Thrones* data set, the values for *politeness* and *passives* are compiled into one new variable (*indirectness*), as well as those for emoticons and onomatopoeia (*high context*). The combinations are very similar to the groups of items used to test the internal consistency for *directness/indirectness* and *high context/low context* dimensions, which does make it more likely that these dimensions would work in other experiments. In the *Bioshock*

data set, *Emoticons* and *onomatopoeia* are combined into the variable *high context*. *Demonstrative pronouns* and *actions verbs* are turned into the variable *orientation to content*. The combination of *jargon* and *abbreviations* is called *Idiomatic Context*. The new dimensions used per data set are as follows:

Table 6
New Dimensions Used for Analysis Table

Data set			
<i>Game of Thrones</i>		<i>Bioshock</i>	
Dimension	Items	Dimension	Items
Indirectness	Politeness, passives	High context	emoticons, onomatopoeia
High context	Emoticons, onomatopoeia	Orientation to Content	demonstrative pronouns, action verbs
		Idiomatic Context	Jargon, abbreviations

3.4 Data Analysis

The testing procedure included scoring for the categories outlined above. Each occurrence was counted as one. The definitions of the items scored were as precise as possible, to decrease room for interpretation and improve the internal validity of the study. Although forum posts function like turn-taking, they do not constitute turns since more than one conversation move can be inserted by addressing multiple other users in one message (Androutsopoulos, Online Data Collection, 2013). However, a single post was counted as one message, regardless of length or number of conversation moves and the number of forum posts used for analysis is referred to as *N*. The researchers each scored half of each data set, checking the numbers independently to improve the reliability of the analysis. To control scoring errors, random checks were performed on the results by the other researcher. One hypothesis was tested:

(H1) There is no difference in language use between Dutch and English groups in online interaction when discussing the same topic on discussion boards

The experiment used a between-subjects design. The independent variable is language, which is either Dutch or English, and the dependent variables are detailed on the next page.

Table 7
Dependent Variables Used for Analysis

Data set	
<i>Game of Thrones</i>	<i>Bioshock</i>
Indirectness	High context
High context	Orientation to content
Imperatives	Idiomatic Context
Images	Imperatives
Animations	Passives
First person	Politeness
Second person	Images
Other username	Animations
Implied first person	First person
Demonstrative pronouns	Second person
Action verbs	Other username
Personal pronouns	Implied first person
Jargon	Personal pronouns
Abbreviations	

The variables were coded into ratio variables, dividing the scores by the word count per message, and used to analyse the mean differences between the English and Dutch forum posts.

To test differences between groups with a large number of continuous dependent variables, a series of t-tests with a correction for multiple testing was executed. The newly compiled variables as well as variables that did not qualify as measuring the same underlying construct were used in the series of t-tests. Assumptions for a t-test include assumptions for normality and a lack of outliers. The new variables were checked for normality and outliers and the variable were log₁₀ transformed. Univariate outliers were only removed if it severely

affected normality. A t-test further requires homogeneity of variances, which can be checked for by looking at the significance of Levene's test of equality of variances. The t-test had a significance level α , which was set at the standard 0.05. However, the large number of dependent variables requires a series of t-tests. However, if there is no correction of the alpha, which is set at 0.05, the chance of a type I error increases exponentially⁶. In this study the false discovery rate is consequently controlled by correcting the significance levels to q^* , which replaces the alpha (Benjamini & Hochberg, 1995). When the assumption of homogeneity of variance was violated, the alternative t-test result was used. The significance of the t-tests was reported as p .

The M is the mean occurrence of an item within the forum posts used for analysis. Since posts with zero values of a particular item were removed, the M is calculated only over posts that do include occurrences of that item. The items were log10 transformed, which means that the M is abstract.

The strength of the differences between two groups in a t-test was determined by calculating Cohen's d for effect size. Generally, an effect size of 0.8 or higher is considered large, values between 0.5 and 0.3 are moderate and anything below 0.3 is small. The sign of Cohen's d tells the direction of the effect. The effect size is unrelated to the statistical significance of the performed test; it only measures the size of the effect. The effect size or power of the test indicates the probability of correctly rejecting the null hypothesis when it is false. The effect size of a t-test cannot be calculated if there is no homogeneity of variances, since it needs the pooled standard deviation. Cohen's d is used to measure how practically significant a difference is.

⁶ For a single t-test with an alpha (α) of 0.05, the chance of finding false positive is 5%. If the number of tests is increased by N , the chance for false positives is much larger. The formula for the alpha in percentages over multiple tests is $100(1.00 - 0.95^N)$. For instance, at 20 t-tests the equation reads $100(1.00 - 0.95^{20})$, which amounts to a 64 percent chance of at least one false positive.

The next formula is used for correcting the significance levels per test. The formula requires rank i and number of tests N . The rank is the order of the results of the t-tests from small to large.

$$q^* = \frac{i}{N} \alpha$$

For a p ranked third in the list of fifteen tests, the corrected significance level q^* is 0.01. By correcting for multiple testing, the alpha is controlled at 0.05 over all the t-tests. The constant $c(N)$ is not included in the formula since $c(N)=1$ for most cases (Benjamini & Yekutieli, 2001). The statistical significance of the series of t-tests can be determined by comparing p to q^* with the test being significant when $p < q^*$.

4. Results

The literature study indicated that online interaction could be determined by affinity or topic, much more so than the language used for interaction. This resulted in the concept of an Affinity-based Community. This concept underscores the relationship between language and culture. The objective of the corpus analysis was to see whether there is a difference in communication style between the Dutch and English forums on the same topic. It is based on the premise that ABCs communicate in the same way, which led to the hypothesis used in this study:

(H1) there is no difference in language use between Dutch and English groups in online interaction when discussing the same topic on discussion boards

The differences between the language used on Dutch and English forums were determined by comparing the dimensions and items as found in the literature and validated by the reliability analysis. The analysis was conducted twice: first with the *Game of Thrones* dataset and second with the *Bioshock: Infinite* dataset. The analysis of more than one genre allows us to rule out the possibility that our results are not specific to any one topic or genre. This chapter will outline the results of the analysis.

4.1 Results Data Set: Game of Thrones

Data set 1 contains 200 forum posts from four different websites, two of which were English language forums and the other two were Dutch. From these forums, 100 English language posts and 100 Dutch language posts were analysed according to items and dimensions in the framework. The sources of the posts are displayed in table 8; the selected forums topics are based on the *Game of Thrones* books and series

Table 8

Game of Thrones Forums Table

	GOTEN1	GOTEN2	GOTNL1	GOTNL1
Forum	Movieforums.com	Fanforum.com	Moviemeter.nl	Filmtotaal.nl
Thread	Game of Thrones Season 3	Season 3 Discussion #1	Game of Thrones	A Game of Thrones
Post 1	04-01-2013, 09:22	02-01-2013, 13:16	05-01-2013, 22:13	15-01-2013, 14:39
Post 50	04-08-2013, 20:13	02-08-2013, 07:25	27-02-2013, 11:24	07-05-2013, 16:15

Note. The posts in either of the forums are not numbered, but each one has a timestamp; these timestamps were used to identify the posts used in analysis. The timestamps display the following information: day, month, year and time of posting.

The Dutch and English language groups are compared based on the different dimensions and variables as described in the previous chapter. The hypothesis states that there should be no differences between the Dutch and English groups. The means comparison was done with repeated independent-samples t-tests, followed by an FDR (*false discovery rate*) correction.

Table 9 shows t-test results of the *Game of Thrones* data set.

Table 9

Game of Thrones T-Test and Effect Size Table

Item	<u>English</u>			<u>Dutch</u>			<i>p</i>	<i>q*</i>	<i>d</i>
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>			
Indirectness	18	-1.59	0.47	43	-1.35	0.45	0.70	0.0429	-0.522
High context	45	-1.18	0.51	49	-1.26	0.39	0.404	0.0286	-0.176
Imperatives	2	-2.43	0.53	2	-1.74	0.54	0.324*	0.0214	-
Images	5	0.10	0.11	2	0.29	0.13	0.231*	0.0179	-
Animations	2	-1.20	0.35	12	-1.03	0.49	0.652	0.0357	-0.399
First person	60	-1.27	0.34	50	-1.29	0.34	0.760	0.0464	0.059
Second person	23	-1.44	0.54	21	-1.44	0.32	0.989*	0.05	-
Other username	6	0.24	0.16	4	0.21	0.09	0.705	0.0429	0.231
Implied first person	11	7.60	3.23	24	5.42	1.67	0.055*	0.0143	-
Demonstrative pronouns	27	-1.47	0.36	55	-1.23	0.36	0.005	0.0036	-0.667
Action verbs	52	-1.25	0.34	41	-1.41	0.26	0.011	0.0071	0.523
Personal pronouns	67	-1.13	0.27	51	-1.26	0.36	0.043*	0.0107	-
Jargon	56	-1.07	0.20	50	-1.11	0.35	0.473*	0.0321	-
Abbreviations	24	-1.53	0.44	19	-1.40	0.43	0.347	0.025	-0.299

Note. $\alpha = 0.05$. *N* is the total number of forum messages that were used for analysis, across the English and Dutch groups. All items are log₁₀ transformed. The *q** values are the corrected significance levels. The value *d* is Cohen's *d* for effect size.

*Levene's test for equality of variances was not significant.

High context, *first person*, *other username* and *abbreviation* have small *d* values, meaning that the observed difference was small⁷. The variables *indirectness*, *animations*, *demonstrative pronoun* and *action verb* have moderate effect sizes, which indicates that the differences measured were of moderate size. All negative *d* values indicate that the Dutch language group had a higher mean than the English language group.

⁷ There are no large differences between the *M* of the Dutch and English groups. However, the *M* values are the mean of the transformed variables, which might skew perception. As stated in the method, the values for *d* indicate the size of the difference and the direction of the difference between groups.

Looking at the significance of the differences between the groups, only *demonstrative pronoun*, *action verb* and *personal pronoun* are significantly different. However, after correcting for multiple comparisons and comparing p to the corrected significance levels (q^*), no significant differences remain. For example, the *demonstrative pronoun* variable has a p value of 0.005, which would be significant if the normal significance level ($\alpha = 0.05$) was kept. However, after correcting for multiple testing for multiple comparisons, the corrected significance level q^* is 0.0036 for *demonstrative pronouns*. The value for p should be smaller than the (corrected) significance level in order for the difference to be statistically significant. However, the q^* value is still smaller than p for the *demonstrative pronoun* variable, which means that it is not significant and that there is no difference between groups. This is also true for the *action verb* and *personal pronoun* variables, which are not significant after correcting the significance level. Without any significant differences in the variables, it means that this data set confirms the hypothesis; there is no difference between English and Dutch language posts on threads with the same topic.

4.2 Results Data Set: Bioshock Infinite

Data set 2 also contains 200 forum posts from four different websites, two of which were English language forums and the other two were Dutch. From these forums, 100 English language posts and 100 Dutch language posts were analysed according to the devised model. These forums are used to discuss games, rather than film and television. The analysed forum threads are on the topic of the game *Bioshock*. The same method was employed to analyse the second data set. The hypothesis is tested again with this second analysis; whether there are differences in language style between the Dutch and English forums. The sources for the data from the four forums can be found below (table 10).

Table 10
Bioshock Forums Table

	BIOEN1	BIOEN2	BIONL1	BIONL2
Forum	Gamespot.com	Irrational Games Forums (2k.com)	Gamer.nl	XBW.nl
Thread	Bioshock Infinite Official Thread. One of the highest rated games EVER.	Bioshock Infinite Discussion (Spoiler?)	[Multi] Bioshock Infinite	Bioshock: Infinite
Post 1	04-2013, post #1	08-12-2010, #1	12-08-2010, 21:48	12-08-2010, #1
Post 50	04-2013, post #50	08-12-2010, #50	02-07-2011, 12:41	14-08-2010, #50

Note. Forum posts used for analysis can be identified via post number # or timestamps ordered day, month, year and time of posting. Day and time are not displayed on Gamespot posts.

The hypothesis states that there should be no differences between the Dutch and English forums. The newly compiled variables, as well as the other variables measured, are compared through independent-samples t-tests. The t-test results, corrected significance levels and Cohen's *d* are found in table 11.

Table 11
Bioshock t-test and Effect Size Table

Item	<u>English</u>			<u>Dutch</u>			<i>p</i>	<i>q*</i>	<i>d</i>
	<i>N</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>			
High context	28	-1.24	0.39	54	-1.11	0.45	0.221	0.021	-0.309
Orientation content to	71	0.56	0.15	70	0.61	0.16	0.052	0.013	-0.322
Idiomatic Context	29	-1.31	0.34	38	-1.31	0.26	0.969	0.050	0
Imperatives	13	-1.37	0.53	8	-1.16	0.43	0.313	0.033	-0.435
Passives	15	-1.52	0.48	12	-1.38	0.44	0.440	0.038	-0.304
Politeness	4	-1.82	0.58	3	-1.57	0.29	0.524	0.046	-0.545
Images	2	-1.79	0.36	2	-0.73	0.39	0.106*	0.017	-
Animations	1	-2.34	-	7	-1.02	0.39	-	-	-
First person	54	-1.19	0.27	47	-1.44	0.20	1E-06	0.004	1.052
Second person	13	-1.50	0.42	14	-1.40	0.30	0.496	0.042	-0,274
Other username	4	0.33	0.17	2	0.14	0.05	0.230	0.025	1.516
Implied first person	13	-1.35	0.32	23	-1.46	0.18	0.278*	0.029	-
Personal pronouns	65	-1.16	0.27	53	-1.32	0.21	3.62E-04*	0.008	-

Note. $\alpha = 0.05$. *N* is the total number of forum messages that were used for analysis, across the English and Dutch groups. All items are log10 transformed. The *q** values are the corrected significance levels. The value *d* is Cohen's *d* for effect size.

*Levene's test for equality of variances was not significant.

The mean differences between the groups appear small; no large differences are found between the *M* values of Dutch and English variables. Looking at the effect sizes and the practical differences in mean, the effect size for *second person* is small; the effect sizes for *high context*, *orientation to content* and *imperatives* are moderate. Both the effect sizes for first person (*d*= 1.052) and other username (*d*= 1.516) are very large, indicating very large differences in mean. When *d* is positive, the English forums have a higher mean.

Looking at the *p*-values, *first person* and *personal pronoun* are the only two items that are significantly different. Even with the corrected significance levels *q**, the differences in

first person and *personal pronoun* remain significant. The p for *first person* is incredibly small, it is 0.00001, which is much smaller than the corrected significance level q^* of 0.004. Looking at the absolute values measured during the analysis of the forums, there are far more instances of first person pronouns on the English websites (115 occurrences in the English forums, versus 68 in the Dutch forums). This corresponds with the *first person* variable having a very large effect size, implying that not only is the difference significant, it is also large in size.

The *personal pronoun* variable has a p of 0.000362 and is also smaller than its corrected significance level 0.008. For *personal pronoun*, there is no homogeneity of variances, so estimating the practical size of the difference is difficult. First person pronouns are also included in the *personal pronoun* variable, which means it is unsurprising that it is significantly different, or that there is a lack of homogeneity of variances. If the first person pronouns were not included in this variable, there would not be a significant difference between the usage of personal pronouns on the Dutch and English forums. Without the first person pronouns, the occurrences of other pronouns on English forums would be 34 and 30 on the Dutch forums. For the variable *personal pronouns*, the difference as perceived in the t-test is false and is not taken into account when considering the validity of the hypothesis.

Aside from the *first person* variable, there are no differences between the Dutch and English groups. However, for that variable, there is a difference between Dutch and English posts on the same topic on a forum. The difference is also large enough to have influenced the *personal pronoun* variable results of the t-test. This would lead to the conclusion that for the *Bioshock* dataset, the hypothesis is not entirely confirmed.

5. Discussion

This study was designed to gain insight into the manifestation of culture and *linguacultural* differences in an online context. The premise that Affinity-based Communities do not differ in communication style is put to the test by comparing two different language groups discussing the same topic. In this experiment the language groups that are compared are Dutch-language groups and English-language groups discussing *Game of Thrones* and *Bioshock Infinite*, respectively. The results show that there is no significant difference overall between the Dutch and English language groups in both the *Game of Thrones* and the *Bioshock Infinite* datasets, which confirms the hypothesis that *there is no difference in language use between Dutch and English groups in online interaction when discussing the same topic on discussion boards*. The only exception is that there is a significant difference in first person pronoun usage in the *Bioshock Infinite* dataset between the Dutch and English language groups.

During analysis it became clear that the hypothesised dimensions as described in the method were not internally consistent and could not be applied to the datasets in this study. As an alternative, some variables were compiled into new combinations, namely *indirectness*, *high context*, *orientation to content* and *idiomatic context* (see table 6), which did correctly measure the underlying constructs. The variables that did not belong together in a dimension were compared as individual variables. Some of the successful new combinations remained very similar to the original dimensions, but represent one side of the sliding scale, rather than the full spectrum of the dimension, including *indirectness*, *high context* and *orientation to content*. This could indicate that perhaps the opposing sides of the sliding scale should be operationalised separately to accurately represent these dimensions within a quantitative model of analysis, or that sliding scales should not be included in such a model at all.

The lack of overall significant differences within the examined ABCs is in line with earlier research discussed in this study. The findings are in agreement with conclusions by

Hewling and Wenger, and their concepts of *Third Culture* and *Community of practise* (Hewling, 2006; Wenger, 1998), which is very similar to the concept of ABCs. Nickerson (2000) also finds that communication styles between Dutch and English speakers share more similarities than they display differences. The findings in this study confirm and expand upon their ideas, indicating that groups construct their own ABCs online, but push them beyond the homogenous and familiar spheres of the classrooms and workplaces analysed by Hewling, Wenger and Nickerson.

The implications glimpsed from the findings by Witzmer and Katzman, Herring and Androutsopoulos (1997; 2003; 2006) are also confirmed. These studies found that national culture seem only to partly influence communication styles; this study confirms the suggestion that communicational styles are more related to genre and affinity than gender, socio-economic background and/or national culture. Interaction in communities flourishes on the Internet, even within the public spheres where national culture is partly (Herring, 2003; Witmer & Katzman, 1997; Androutsopoulos J., 2006) or fully concealed (this study) by anonymity. The notion of equating culture to national cultures is thus regarded as an outdated yet widely employed concept; this is visible in the many publications that employ Hofstede despite the various critiques of Hofstede's research (McSweeney, 2002; Kim, 2007). Instead, the concept of Affinity-based Community grants insight in the way culture is socially constructed in online contexts and could serve as a valid perspective for analysing the dynamics of culture in communities on the Internet. The analysis of ABCs through the operationalisation of the dimensions by House and Hall has been mostly preliminary. No other study has attempted to combine, quantify and employ these dimensions to analyse contexts as new and dynamic as the internet. As expected, this study ran into the methodological issues that can be expected from any study testing a new model. Despite the necessity for modifications to the initial

operationalisation, the dimensions by Hall (1976) and House (2007) show great potential for the analysis of culture in online interactional spheres.

The creative ways in which emoticons, onomatopoeia and imagery (Androutsopoulos, 2006; Herring, 2003; Callahan, 2006; Singh et al, 2006) are embraced within both ABCs instil great promise for the quantification and analysis of *high context* elements in online (textual) environments and interaction (Androutsopoulos, 2006). Even when quantified, the cultural dimensions by House (2007) and Hall (1976) represent the sliding scales that have successfully served other studies by revealing differences in preferences and tendencies in communication styles; the results of this study indicate that, within each ABC, both language groups display no significant differences along the constructed dimensional framework and this allows us to conclude that, in line with our expectations, the language groups within both ABCs share the same tendencies and preferences regarding communicational style. This preliminary study seeks to help future researchers further explore the tools to measure culture by quantitative means without the need to resort to predetermined sets of binary oppositions like Hofstede's framework (2001).

Unexpectedly, however, there were two variables in the *Bioshock Infinite* dataset where analysis did show a significantly different result between groups. The first difference concerns the *first person* pronoun variable and the second concerns the *personal pronoun* variable. The difference between the Dutch and English groups is both statistically significant and large, as indicated by the large effect size ($d= 1.052$). This implies that the English *Bioshock* group used first person pronouns much more frequently than the Dutch group did. The second significantly different result concerns the *personal pronouns* variable. However, since the *personal pronouns* variable also included first person pronouns, this difference is not considered significant, seeing that the absolute occurrences of pronouns other than first person pronouns was small and similar in both groups. It is unsurprising that the high frequency occurrence of

first person pronouns influences the amount of personal pronouns as a whole. This results in the conclusion that the difference encountered in the *personal pronouns* variable can be directly attributed to the difference in the *first person* variable. The difference in first person pronoun usage, however, cannot be ascribed to mistakes in testing procedure or analysis, since human error cannot account for such a large difference. The high frequency of first person pronouns in the English group might be explained by the posts used for analysis; these could contain a disproportionately large number of first person pronouns. On the other hand, the posts were sourced from multiple English language forums, rendering it more unlikely that this phenomenon is a coincidence. Further research should indicate if this result is an anomaly or a pattern, since the difference was only found in one data set.

The difference in first person pronoun frequency in this dataset is interesting in that it is limited to a single variable within the *orientation to Self/Other* dimension. The fact that a significant difference is found within this dimension could be explained by House's own research. As part of her research, House tested the validity of each of the dimensions she had formulated, which resulted in the orientation to self/other dimension as the only dimension that was not externally verified. Our data from the *Bioshock* data set might partially support this claim. However, in House's findings, the judges attributed the factors House identified as pertaining to the *orientation to Self/Other* dimension to the *Directness/Indirectness* dimension instead. Since first person pronouns are only a part of what this study originally attributed to the *orientation to Self/Other* dimension, this finding cannot be confirmed nor denied by the present study. If we were to assume that the dimensions as originally devised are correct, other differences along this dimension should be found and would not be limited to a single variable. However, since the originally proposed dimensions are not internally consistent, this does not necessarily have to be true.

It is also possible that the high frequency of first person pronouns is normal in English communities when discussing *Bioshock Infinite*, while the rest of the linguistic and cultural factors show no significant difference. This is reminiscent of Snelders, Morel and Havermans (2011), since they conclude that small cultural adaptations can be more appealing to different communities, while the overarching industry style remains intact as the dominant factor in the interaction. The difference between first person pronoun frequencies could signify an adaptation favoured by the English Bioshock community, while the lack of significant differences along the other dimensions and their variables indicate the presence of an overarching ABC style. In both cases the dominant style remains intact despite the introduction of minor adaptations. Despite the fact that, unlike this paper, Snelders et al. do rely on the concept of national culture in their adaptation, they have uncovered an interesting concept that could also be of some influence when further exploring the emerging definition of ABCs and the concept of community as culture. ABCs are defined as *places* where users who share the same affinity interact, where they construct their culture and, the language and communication style. However, the possibility remains that other factors, like national culture and specific language features, could have an influence in the ABC.

The data and results in this study suggest that English speakers on *Bioshock* threads use first person pronouns more frequently than Dutch speakers. Further testing should indicate whether the difference remains significant if a larger dataset is analysed and if this result is an anomaly or a trend in the *Bioshock* ABC. Since there was only one variable that was found to be different in one data set, the hypothesis is tentatively confirmed. As discussed previously, the concept of ABCs is very similar to Snelders et al.'s industry styles, Hewling's Third Culture and Wenger's Communities of Practice. Although different in setting and context, an ABC also has its own communicative style and culture. Future research should reveal whether minor

cultural adaptations that differ from the overarching communicative style occur depending on the nationality or native language of its members.

6. Conclusion

The subject of this study was the underexplored topic of online culture, with emphasis on exploring the virtual world as a place rather than a source of text. This study focuses on whether there is any difference in culture on Internet forums within same-topic threads and does so by employing dimensions found in contrastive culture research. However, these dimensions cannot be directly applied to an online context, as factors like anonymity and medium constraints challenge assumptions made by these frameworks. Challenging these assumptions opens up new avenues and possibilities for application in online environments; it broadens the perspective on concepts of community, place and identity. These factors indicate that traditional boundaries and differences along the concept of national culture may have become obsolete within large areas of public and anonymous online interaction that have not yet been fully explored in literature.

The emerging definition of Affinity-based Communities implies that communication style is more dependent on affinity for certain topics, rather than on gender, age or geographical location. The medium restraints of computer-mediated communication limit most online interaction to on-screen, visual means of communication. In most online communication, lacking the extra information that face-to-face interaction provides, the language and communicative style used has become far more significant, in the sense that they make up the very core of online interaction. In this way, language becomes indicative of culture in these online communities. Online forums are an example of fast-paced communication in a public, anonymous sphere where users are free to discuss any topic with other users from all over the world. This makes online forums an ideal subject for contrastive cultural analysis. With this in mind, this paper set out to answer the question of whether there is a difference in language use, and by extension culture, between language groups when discussing the same topic on the

Internet. In order to see whether differences in communicative style occur, this study compared Dutch- and English-language groups. The hypothesis used in this study was:

There is no difference in language use between Dutch and English groups in online interaction when discussing the same topic on discussion boards.

To answer this question, discussion board posts on the topic of *Bioshock: Infinite* and *Game of Thrones* were selected. These two data sets were analysed by using a hypothesised framework of cultural dimensions based on House (2007) and Hall (1987). The most important finding is that there was generally no difference between the English and Dutch groups in the analysis. A single variable in one of the data sets, *first person pronoun*, was the only variable to show significant difference between groups. The question remains whether this is an anomaly or if it is a cultural adjustment accepted within the ABC. Except for this one irregularity, the hypothesis is confirmed for both data sets. The confirmation of the hypothesis also supports the concept of ABCs, which are places of online interaction based on a shared interest or purpose. ABCs structure their communication style based on this common ground, rather than by language or nationality. This can influence the definition of communities online, as well as the manner in which online cultural research should be conducted. This is already beginning to manifest itself, as companies like Crimson Hexagon are already offering their services for analysing online social media with a tool based on consumer affinities, rather than traditional demographics. The concept of affinities and ABCs are perhaps the start of a new way of looking at online communication.

6.1 Limitations

There are some limitations to this study. After removing all cases with zero results from the data set, the sample was smaller than expected. However, the different statistical tests used could have produced valid significant results based on smaller sample sizes; both independent t-tests as well as the Cronbach's alpha for internal consistency could produce valid results as long as the testing assumptions are adhered to. Testing for normality was also done with multiple testing instruments so as to mitigate the influence of the sample size, as well as a repetition of the experiment to further improve validity. However, a larger sample size would have enabled the researchers to draw stronger conclusions from the data. A larger data sample could also have confirmed the internal consistency of the dimensions proposed in the framework, rather than only confirming internal consistency between a limited number of variables. With the available data, the confirmed dimensions do have a high internal consistency. This at least partly confirms the postulated dimensions and these do contribute to content validity of the model.

Scoring was performed by two researchers, which introduces human error to the study. The framework used for analysis is mostly based on research from Western Europe and the researchers are also based in Western Europe, possibly making the framework Eurocentric and limiting the generalisability of the framework. The items in the dimensions were, however, defined in such a way that lexical and grammatical function was employed to eliminate bias and interpretability as much as possible. Both researchers performed random checks on the other researcher's scoring results in order to minimalise human error. The jargon category in the *Idiomatic Context* dimension was added based on observations from the data, but it required insider knowledge from the researchers to count the phenomena. This means the item is also more susceptible to interpretation, since it depends on interpretation by the researchers. While defining the item as narrowly as possible, it remains more sensitive to bias than the other items

scored. Character names, place names, specific events and quotes from *Game of Thrones* and *Bioshock: Infinite* found in the data sets were easily identified as jargon. More general game- or television-related language was harder to identify as jargon. On the whole, the items used for scoring were not prone to interpretation and scoring errors can be limited to human error.

The statistics used for analysing the results has its own limitations. As mentioned before, the Crohnbach's alpha and t-test are more reliable when the test assumptions are complied with. The reliability of the Crohnbach's alpha as a test for internal consistency is good, however, it is sensitive to non-normal error scores when the data sample is small ($N < 100$). This was resolved by transforming non-normal variables in the data set and removing extreme outliers. Since the coefficient alpha provides an overall reliability coefficient, more than one underlying construct could be measured. With a smaller data sample, the squared multiple correlation was used to review the construct being measured. A very high Crohnbach's alpha can suggest redundancy, although this is unlikely to have happened since the scored categories are all different. An independent t-test is a solid measure of means difference, but it does require fairly normal distribution, no outliers and homogeneity of variances for standard t-tests. Moreover, since the study utilises a series of t-tests the chance for a false positive, instead of being at $\alpha = 0.05$, is increased exponentially per repeated test. For this reason, the Benjamini-Hochberg procedure was used to control the false discovery rate (FDR). Power analysis was conducted post-hoc, resulting in the Cohen's d for all t-tests that have homogeneity of variances. The d shows the size of the difference, rather than only the significance of the difference as a t-test does. Together, these extra procedures make the t-test more reliable and give additional information about the differences between the Dutch and English forums; the chance for type I errors is controlled and power of the difference is included with the significance of the differences between groups.

6.2 Suggestions for Further Research

To improve upon the research design in this study, future research can employ several methods. To get a larger data sample, more researchers can collaborate for data collection and scoring. To limit the chance of human error in scoring, these researchers require training in the scoring model. Initially, they should all start scoring with a small data set, after which the agreement between rates can be determined with the calculation of Cohen's kappa (k). This procedure will improve the reliability of the scoring of the actual data set as well as increase the data sample in a shorter amount of time. If human error is to be removed entirely from the scoring, a computer programme should be created or appropriated to perform the scoring. The advantage of computerised scoring is that a large amount of data can be scored with no chance of error in a short amount of time. There is one condition for scoring to be done accurately, which is that all the scored items need to be defined very specifically. This means that a significant amount of time should be invested in setting up and programming the scoring definitions. Items from the model, like jargon, might be split up into even more specific categories. These separate categories can contribute to a more detailed map of what lexical and grammatical occurrences are found in the context of jargon on online forums and communities, even across other affinity-based communities and languages.

With a larger data sample, the framework of dimensions can be tested more easily and accurately. An increased data sample means that a factor analysis can be executed to remove any variables from the data that do not fit with the model of analysis and create a principle component that combines scores of variables that correlate highly. This means a factor analysis is similar to the use of Cronbach's alpha to check for internal reliability, but on a larger scale with automatically generated new variables to represent the dimensions. This also removes any high multicollinearity from the test, which is an assumption for a one-way multivariate analysis of variance (MANOVA). Instead of using a series of t-tests on the different variables, a one-

way MANOVA can test all the dependent variables at the same time. A one-way MANOVA can be utilised, provided that there is an adequate (large) sample size, lack of multicollinearity and a linear relationship between the dependent variables as well as multivariate normality, lack of outliers and homogeneity of variance-covariance matrices. These assumptions are similar to the assumptions for a t-test and are tested for in a very similar way, while other assumptions are tested for in the MANOVA procedure. What is being tested does not differ much from the original design, although a larger data sample does enable an easier and more reliable way to test for differences and verify the framework's dimensions.

It would also be useful for new research to investigate differences between other languages, and to test the framework against a different linguistic background. The two languages used for this study are both Indo-European and West-Germanic as well as from countries that are geographically close. Using languages from a different branch entirely, or non-Indo-European languages for instance, would further test the validity of the framework and give insight in how to refine it for further research; it may be that the framework only works when the languages are fairly closely related. Additionally, *Bioshock: Infinite* and *Game of Thrones* are both originally English-language media. It would be very likely that the language background of the selected topic influences the communicative style of the ABC in question. Further research should contrast languages that do not share the same language background as the topic, to give insight into how much influence the language background of the topic used for analysis exerts on the communication style of the ABC.

The material for analysis could also be taken from other public, asynchronous online sources. Instead of scoring items on forums, social media could be utilised. Facebook has a similar structure to forum posts, although these online interactions might differ from forum posts since Facebook tends to be used for more private communication as well as public. The topics are also not as concentrated or grouped together as they are on a forum thread. Analysing

Tweets with certain hashtags could be an alternative which does give the researcher the ability to collect data based on the topic being discussed. However, Twitter messages are limited to a maximum of 140 characters and hashtags tend to have a very narrow topic scope. A microblogging website like Tumblr also groups posts by hashtag, with people reblogging and commenting on the main post. Tumblr is also a breeding ground for animated GIFs and memes, making it an interesting medium for analysis. Selecting a different online medium for analysis brings new challenges and opportunities for exploration of the framework depending on the medium.

Bibliography

- Agar, M. (1994). *Language Shock. Understanding the Culture of Conversation*. New York: Morrow.
- Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics*, 10(4), 419-438.
- Androutsopoulos, J. (2007). Language Choice and Code Switching in German-Based Diasporic Web Forums. In B. Dannet, & S. Herring. *The Multilingual Internet* (To appear). Oxford: Oxford University Press.
- Androutsopoulos, J. (2013). Online Data Collection. In C. Mallinson, B. Childs, & G. van Herk, *Data Collection in Sociolinguistics: Methods and Applications* (pp. 492-520). New York: Routledge.
- Baym, N. (1998). The emergence of on-line community. In S. Jones, *Cybersociety.2.0. Revisiting Computer-Mediated Communication and community*. (pp. 35-68). London: Sage.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57(1), 289-300.
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics*, 29(4), 1165-1188.
- Byrnes, H. (1986). *Interactional Style in German and American Conversation*.
- Callahan, E. (2006). Cultural Similarities and Differences in the Design of University. Web sites. *Journal of Computer-Mediated Communication*, 11, 239-273.
- Crystal, D. (2001). *Language and the Internet*. Cambridge: Cambridge University Press.
- Dannet, B. (1998). Text as mask: Gender, play and performance on the internet. In S. Jones, *Cybersociety 2.0. Revisiting Computer-Mediated Communication and community*. (pp. 129-158). London: Sage.
- Denzin, N. (1998). In Search of the Inner Child: Co-dependency and Gender in a Cyberspace. In G. Bendelow, & S. J. Williams, *Emotions in Social Life: Critical Themes and Contemporary* (pp. 97-119). London: Routledge.
- Ess, C., & Sudweeks, F. (2006). Culture and Computer-Mediated Communication: Toward New Understandings. *Journal of Computer-Mediated Communication*, 11, 179-191.
- Georgakopoulou, A. (2003). Computer-mediated Communication. In J. Verschuren, J. Blommaert, & C. Bulcaen, *Handbook of Pragmatics* (pp. 1-20). Pennsylvania: Benjamins.
- Hall, E. T. (1966). *The Hidden Dimension*. Toronto: The Bodley Head.
- Hall, E. T. (1976). *Beyond Culture*. New York: Anchor Books / Doubleday.
- Hermeking, M. (2006). Culture and Internet Consumption: Contributions from Cross-Cultural Marketing and Advertising Research. *Journal of Computer-Mediated Communication*, 11, 192-216.
- Herring, S. (2003). Gender and Power in online communication. In H. & Meyehoff, *the Handbook of Language and Gender* (pp. 202-228). Oxford: Blackwell.
- Hewling, A. (2006). Culture in the Online Class: Using Message Analysis to Look Beyond Nationality-Based Frames of Reference. *Journal of Computer-Mediated*

- Communication*, 11, 337-356.
- Hofstede, G. (2001). *Culture's Consequences: Comparing Values, Behaviours, Institutions and Organizations across Nations*. Thousand Oaks: Sage Publications.
- Holliday, A., Hyde, M., & Kullman, J. (2010). *Intercultural Communication: An Advanced Resource Book for Students*. New York: Routledge.
- House, J. (2007). Communicative styles in English and German. *European Journal of English Studies*, 249-267.
- Huffaker, D., & Calvert, S. (2005). Gender, identity and language use in teenage blogs. *Journal of Computer-Mediated Communication*, 10(2), n.a.
- Kim, M.-S. (2007). Our Culture, Their Culture and Beyond: Further Thoughts on Ethnocentrism in Hofstede's Discourse. *Journal of Multicultural Discourses*, 2(1), 26-31.
- Macfayden, L., Roche, J., & Doff, S. (2004). *Communicating across Cultures in Cyberspace*. München: University of Munich.
- McQuail, D. (2005). *McQuail's Mass Communication Theory*. London: Sage.
- McSweeney, B. (2002). Hofstede's Model of National Cultural Differences and Their Consequences: A triumph of faith - a failure of analysis. *Human Relations*, 55(1), 89-118.
- Minchov, M., & Hofstede, G. (2011). The Evolution of Hofstede's Doctrine. *Cross-Cultural Management: An international Journal*, 10-20.
- Nickerson, C. (2000). *Playing the Corporate Language Game*. Amsterdam: Rodopi.
- Ogiermann, E. (2009). Politeness and in-directness across cultures: A comparison of English, German, Polish and Russian requests. *Journal of Politeness Research* (5), 189-216.
- Preece, J., Maloney-Krichmar, D., & Abras, C. (2003). History of Online Communities. In K. Christiansen, & D. Levinson, *Encyclopedia of Community* (Vol. 3, pp. 1023-1027). Thousand Oaks, California: Sage.
- Scollon, R., & Wong-Scollon, S. (2001). *Intercultural Communication*. Oxford: Blackwell Publishers Ltd.
- Sheng, Y., & Sheng, Z. (2012). Is Coefficient Alpha Robust to Non-Normal Data? *Frontiers in Psychology*, 3(34), 1-13.
- Singh, N., Zhao, H., & Hu, X. (2006). Analyzing the cultural content of web sites: a cross-national comparison of China, India, Japan, and US. *International Marketing Review*, 22(2), 129-146.
- Snelders, D., Morel, K. P., & Havermans, P. (2011). The cultural adaptation of web design to local industry styles: A comparative study. *Design Studies*, 32(5), 457-481.
- Tempel, M., & ten Thije, J. (2012). The Appreciation of Cultural and Linguistic adjustments in Multilingual Museum Audio Tours by International Tourists. *Journal of Multilingual and Multicultural Development*, 33(7), 643-664.
- Thije, J. D. ten, & Pinto, M. (2011). Twee IKEA catalogi, een Italiaanse en een Nederlandse Lezerswaardering van een 'covert translation'. *Filter*, 18(4), 67-80.
- Thije, J. D ten, & Maier, R. (2012). Managing cultural and linguistic diversity in multiple organisational settings: editorial. *Journal of Multilingual and Multicultural Development*, 33(7), 629-643.
- Thorne, S., Black, R., & Sykes, J. (2009). Second Language Use, Socialization and Learning

- in Internet Interest Communities and Online Gaming. *The Modern Language Journal*, 802-821.
- Turkle, S. (1995). *Life on the Screen*. New York: Simon & Schuster.
- Wenger, E. (1998). *Communities of Practise*. Cambridge: Cambridge University Press.
- Williams, J., & Copes, H. (2005). "How Edge are You?" Constructing Authentic Identities and Subcultural Boundaries in a Straightedge Internet Forum. *Symbolic Interaction*, 28(1), 67-89.
- Witmer, D., & Katzman, S. (1997). On-line smiles: Does gender make a difference in the use of graphic accents? *Journal of Computer-Mediated Communication*, 2(4), n.a. *World Internet Users Statistics and World Population Stats*. (2012, January 18). Retrieved from internetworldstats: <http://www.worldinternetstats.com/stats.htm>
- Würtz, E. (2006). Intercultural Communication on Web sites: A Cross-Cultural Analysis of Web sites from High-Context Cultures and Low-Context Cultures. *Journal of Computer Mediated Communication*, 11, 274-299.

Appendices

Appendix 1: Forum URLs

Table 1

Forum Internet Links Table

Forum	URL
Movieforums (EN)	http://www.movieforums.com/community/showthread.php?t=31511&highlight=game+thrones
Fanforum (EN)	http://www.fanforum.com/f428/got-season-3-discussion-1-a-63114957/
Moviemeter (NL)	http://www.moviemeter.nl/forum/12/11648/750
Filmtotaal (NL)	http://www.filmtotaal.nl/forum/viewtopic.php?f=32&t=5281&start=240
Gamespot (EN)	http://www.gamespot.com/forums/games-discussion-1000000/bioshock-infinite-official-threadone-of-the-highes-29369866/?page=1
Irrational Games (EN)	http://forums.2k.com/showthread.php?79887-Bioshock-Infinite-Discussion-(Spoilers-
Gamer (NL)	http://forum.gamer.nl/showthread.php?t=79303
XBW (NL)	http://forum.xboxworld.nl/showthread.php?threadid=146083

Note. The URLs are for the specific threads used in analysis.